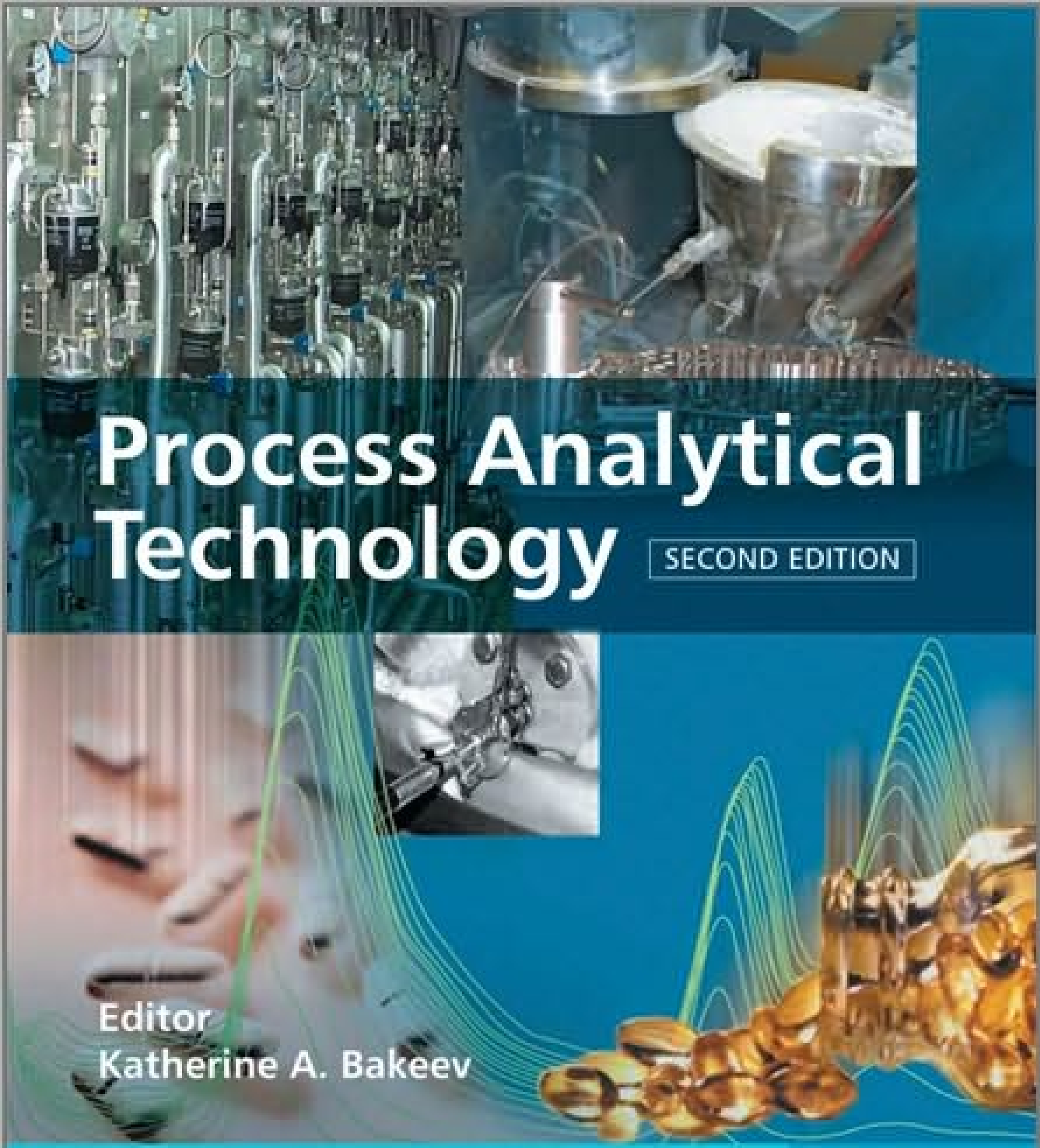


Process Analytical Technology

SECOND EDITION

Editor
Katherine A. Bakeev

WILEY

Process Analytical Technology – Second Edition

Spectroscopic Tools and Implementation Strategies for the Chemical and Pharmaceutical Industries

Edited by

Katherine A. Bakeev
CAMO Software, Inc, NJ, USA

WILEY

A John Wiley & Sons, Ltd., Publication

Process Analytical Technology

Second Edition

Process Analytical Technology – Second Edition

Spectroscopic Tools and Implementation Strategies for the Chemical and Pharmaceutical Industries

Edited by

Katherine A. Bakeev
CAMO Software, Inc, NJ, USA

WILEY

A John Wiley & Sons, Ltd., Publication

This edition first published 2010
© 2010 John Wiley & Sons, Ltd

Registered office
John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, United Kingdom

For details of our global editorial offices, for customer services and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com.

The right of the author to be identified as the author of this work has been asserted in accordance with the Copyright, Designs and Patents Act 1988.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, except as permitted by the UK Copyright, Designs and Patents Act 1988, without the prior permission of the publisher.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Designations used by companies to distinguish their products are often claimed as trademarks. All brand names and product names used in this book are trade names, service marks, trademarks or registered trademarks of their respective owners. The publisher is not associated with any product or vendor mentioned in this book. This publication is designed to provide accurate and authoritative information in regard to the subject matter covered. It is sold on the understanding that the publisher is not engaged in rendering professional services. If professional advice or other expert assistance is required, the services of a competent professional should be sought.

The publisher and the author make no representations or warranties with respect to the accuracy or completeness of the contents of this work and specifically disclaim all warranties, including without limitation any implied warranties of fitness for a particular purpose. This work is sold with the understanding that the publisher is not engaged in rendering professional services. The advice and strategies contained herein may not be suitable for every situation. In view of ongoing research, equipment modifications, changes in governmental regulations, and the constant flow of information relating to the use of experimental reagents, equipment, and devices, the reader is urged to review and evaluate the information provided in the package insert or instructions for each chemical, piece of equipment, reagent, or device for, among other things, any changes in the instructions or indication of usage and for added warnings and precautions. The fact that an organization or Website is referred to in this work as a citation and/or a potential source of further information does not mean that the author or the publisher endorses the information the organization or Website may provide or recommendations it may make. Further, readers should be aware that Internet Websites listed in this work may have changed or disappeared between when this work was written and when it is read. No warranty may be created or extended by any promotional statements for this work. Neither the publisher nor the author shall be liable for any damages arising herefrom.

Library of Congress Cataloging-in-Publication Data

Process analytical technology / [edited by] Katherine Bakeev. – 2nd ed.
p. ; cm.
Includes bibliographical references and index.
ISBN 978-0-470-72207-7 (hardback : alk. paper) 1. Chemistry, Technical. 2. Chemistry, Analytic–Technological innovations. 3. Chemistry, Analytic–Technique. 4. Spectrum analysis. 5. Pharmaceutical chemistry. I. Bakeev, Katherine A.
[DNLM: 1. Chemistry, Pharmaceutical. 2. Chemistry, Analytic. 3. Spectrum Analysis–methods. QV 744 P9627 2010]
TP155.75.P737 2010
660′.2–dc22

2010004250

A catalogue record for this book is available from the British Library.

Set in 10/12pt Times by Toppan Best-set Premedia Limited
Printed and Bound in Great Britain by CPI Antony Rowe, Chippenham, Wiltshire

Contents

Preface to the Second Edition *xvii*
List of Contributors *xix*
List of Abbreviations *xxi*

1 Overview of Process Analysis and PAT **1**
Jason E. Dickens
1.1 Introduction 1
1.1.1 Historical perspective 3
1.1.2 Business drivers 4
1.2 Execution of Process Analysis Projects 5
1.2.1 Wisdoms 5
1.2.2 Team structure 6
1.2.3 Project life cycle 6
1.2.4 Project scoping 9
1.2.5 Common challenges and pitfalls 10
1.3 Process Instrumentation 12
1.3.1 Process instrumentation types 12
1.3.2 Novel process instrumentation 12
1.4 Conclusions 13
1.5 Glossary of Acronyms and Terms 14
References 14

2 Implementation of Process Analytical Technologies **17**
Robert Guenard and Gert Thurau
2.1 Introduction to Implementation of Process Analytical Technologies (PAT) in the Industrial Setting 17
2.1.1 Definition of process analytics 18
2.1.2 Differences between process analyzers and laboratory analysis 19
2.1.3 General industrial drivers for PA 19
2.1.4 Types of applications (R&D versus manufacturing) 20
2.1.5 Organizational considerations 20
2.2 Generalized Process Analytics Work Process 23
2.2.1 Project identification and definition 24
2.2.2 Analytical application development 26

2.2.3 Design, specify and procure 26
2.2.4 Implementation in production 28
2.2.5 Routine operation 29
2.2.6 Continuous improvement 30
2.3 Considerations for PAT Implementation in the Pharmaceutical Industry 30
2.3.1 Introduction 30
2.3.2 Business model 30
2.3.3 Technical differences 31
2.3.4 Regulatory Aspects of Process Analytics in the Pharmaceutical Industry – the Concept of Quality by Design 33
2.4 Conclusions 36
References 36

3 Process Sampling: Theory of Sampling – the Missing Link in Process Analytical Technologies (PAT) 37
Kim H. Esbensen and Peter Paasch-Mortensen
3.1 Introduction 37
3.2 Theory of Sampling – Introduction 39
3.2.1 Heterogeneity 41
3.2.2 Constitutional heterogeneity 41
3.2.3 Distributional heterogeneity 42
3.2.4 Structurally correct sampling 45
3.2.5 Incorrect sampling error 45
3.2.6 Increment delimitation error 45
3.2.7 Increment extraction error 46
3.2.8 Increment preparation error 46
3.2.9 Increment weighing error 47
3.2.10 Total sampling error 48
3.2.11 Global estimation error 48
3.3 Mass Reduction as a Specific Sampling Procedure 48
3.4 Fundamental Sampling Principle 51
3.5 Sampling – a Very Practical Issue 51
3.5.1 Sampling unit operations 52
3.5.2 Understanding process sampling: 0-D versus 1-D LOTS 52
3.5.3 Grab sampling – 0-D and 1-D 54
3.5.4 Correct process sampling: increment delimitation/extraction 56
3.5.5 PAT versus correct process sampling – what is required? 58
3.6 Reactors and Vessels – Identical Process Sampling Issues 60
3.6.1 Correct process sampling with existing process technology 62
3.6.2 Upward flux – representative colocated PAT sampling 62
3.6.3 Upstream colocated PAT sampler 64
3.7 Heterogeneity Characterization of 1-D lots: Variography 66
3.7.1 Process sampling modes 67
3.7.2 The experimental variogram 67
3.7.3 Sampling plan simulation and estimation of TSE 71
3.7.4 TSE estimation for 0-D lots – batch sampling 72
3.7.5 Corporate QC benefits of variographic analysis 73

3.8 Data Quality – New Insight from the TOS 75
3.9 Validation in Chemometrics and PAT 76
3.10 Summary 78
References 79

4 UV-visible Spectroscopy for On-line Analysis 81
Marcel A. Liauw, Lewis C. Baylor and Patrick E. O'Rourke
4.1 Introduction 81
4.2 Theory 82
4.2.1 Chemical concentration 82
4.2.2 Color 84
4.2.3 Film thickness 85
4.2.4 Turbidity 85
4.2.5 Plasmons/nanoparticles 85
4.3 Instrumentation 85
4.4 Sample Interface 86
4.4.1 Cuvette/vial 87
4.4.2 Flow cells 87
4.4.3 Insertion probe 87
4.4.4 Reflectance probe 89
4.5 Implementation 89
4.5.1 A complete process analyzer 89
4.5.2 Troubleshooting 89
4.6 Applications 91
4.6.1 Gas and vapor analysis 92
4.6.2 Liquid analysis 92
4.6.3 Solid analysis 96
4.6.4 Other applications 99
4.7 Detailed Application Notes 100
4.7.1 Gas and vapor analysis: toluene 100
4.7.2 Liquid analysis: breakthrough curves 101
4.7.3 Solids analysis: extruded plastic color 101
4.7.4 Film thickness determination: polymer 103
4.8 Conclusion 104
References 104

5 Near-infrared Spectroscopy for Process Analytical Technology: Theory, Technology and Implementation 107
Michael B. Simpson
5.1 Introduction 107
5.2 Theory of Near-infrared Spectroscopy 112
5.3 Analyser Technologies in the Near-infrared 114
5.3.1 Light sources and detectors for near-infrared analyzers 114
5.3.2 The scanning grating monochromator and polychromator diode-array 119
5.3.3 The acousto-optic tunable filter (AOTF) analyzer 123
5.3.4 Fourier transform near-infrared analyzers 127
5.3.5 Emerging technologies in process NIR analyzers 134

5.4 The Sampling Interface 136
5.4.1 Introduction 136
5.4.2 Problem samples: liquids, slurries and solids 142
5.4.3 The use of fiber optics 145
5.5 Practical Examples of Near-infrared Analytical Applications 147
5.5.1 Refinery hydrocarbon streams 148
5.5.2 Polyols, ethoxylated derivatives, ethylene oxide/propylene oxide polyether polyols 149
5.5.3 Oleochemicals, fatty acids, fatty amines and biodiesel 151
5.6 Conclusion 152
References 153

6 Infrared Spectroscopy for Process Analytical Applications 157
John P. Coates
6.1 Introduction 157
6.2 Practical Aspects of IR Spectroscopy 161
6.3 Instrumentation Design and Technology 163
6.4 Process IR Instrumentation 166
6.4.1 Commercially available IR instruments 167
6.4.2 Important IR component technologies 172
6.4.3 New technologies for IR components and instruments 176
6.4.4 Requirements for process infrared analyzers 178
6.4.5 Sample handling for IR process analyzers 185
6.4.6 Issues for consideration in the implementation of process IR 187
6.5 Applications of Process IR Analyzers 189
6.6 Process IR Analyzers: a Review 191
6.7 Trends and Directions 192
References 193

7 Raman Spectroscopy 195
Nancy L. Jestel
7.1 Attractive Features of Raman Spectroscopy 195
7.1.1 Quantitative information 195
7.1.2 Flexible sample forms and sizes used as accessed without damage 196
7.1.3 Flexible sample interfaces 196
7.1.4 Attractive spectral properties and advantageous selection rules 197
7.1.5 High sampling rate 197
7.1.6 Stable and robust equipment 198
7.2 Potential Issues with Raman Spectroscopy 198
7.2.1 High background signals 198
7.2.2 Stability 198
7.2.3 Too much and still too little sensitivity 199
7.2.4 Personnel experience 199
7.2.5 Cost 200
7.3 Fundamentals of Raman Spectroscopy 200

7.4 Raman Instrumentation 203
7.4.1 Safety 203
7.4.2 Laser wavelength selection 204
7.4.3 Laser power and stability 204
7.4.4 Spectrometer 205
7.4.5 Sample interface (probes) 206
7.4.6 Communications 208
7.4.7 Maintenance 209
7.5 Quantitative Raman 209
7.6 Applications 212
7.6.1 Acylation, alkylation, catalytic cracking, and transesterification 213
7.6.2 Bioreactors 213
7.6.3 Blending 214
7.6.4 Calcination 214
7.6.5 Catalysis 215
7.6.6 Chlorination 216
7.6.7 Counterfeit pharmaceuticals 217
7.6.8 Extrusion 218
7.6.9 Forensics 218
7.6.10 Hydrogenation 218
7.6.11 Hydrolysis 219
7.6.12 Medical diagnostics 219
7.6.13 Microwave-assisted organic synthesis 219
7.6.14 Mobile or field uses 220
7.6.15 Natural products 220
7.6.16 Orientation, stress, or strain 221
7.6.17 Ozonolysis 222
7.6.18 Polymerization 222
7.6.19 Polymer curing 224
7.6.20 Polymorphs (crystal forms) 225
7.6.21 Product properties 228
7.6.22 Purification: distillation, filtration, drying 229
7.6.23 Thin films or coatings 229
7.7 Current State of Process Raman Spectroscopy 230
References 231

8 Near-infrared Chemical Imaging for Product and Process Understanding 245
E. Neil Lewis, Joseph W. Schoppelrei, Lisa Makein, Linda H. Kidder and Eunah Lee
8.1 The PAT Initiative 245
8.2 The Role of Near-infrared Chemical Imaging (NIR-CI) in the Pharmaceutical Industry 246
8.2.1 Characterization of solid dosage forms 246
8.2.2 'A picture is worth a thousand words' 247
8.3 Evolution of NIR Imaging Instrumentation 247
8.3.1 Spatially resolved spectroscopy – mapping 247
8.3.2 The infrared focal-plane array 247
8.3.3 Wavelength selection 248

8.3.4 The benefits of NIR spectroscopy 248
8.3.5 NIR imaging instrumentation 249
8.4 Chemical Imaging Principles 251
8.4.1 The hypercube 251
8.4.2 Data analysis 251
8.4.3 Spectral correction 252
8.4.4 Spectral preprocessing 253
8.4.5 Classification 253
8.4.6 Image processing – statistical 255
8.4.7 Image processing – morphology 257
8.5 PAT Applications 257
8.5.1 Content uniformity measurements – 'self calibrating' 258
8.5.2 Quality assurance – imaging an intact blister pack 260
8.5.3 Contaminant detection 261
8.5.4 Imaging of coatings – advanced design delivery systems 263
8.6 Processing Case Study: Estimating 'Abundance' of Sample Components 267
8.6.1 Experimental 268
8.6.2 Spectral correction and preprocessing 268
8.6.3 Analysis 268
8.6.4 Conclusions 273
8.7 Processing Case Study: Determining Blend Homogeneity Through Statistical Analysis 273
8.7.1 Experimental 273
8.7.2 Observing visual contrast in the image 274
8.7.3 Statistical analysis of the image 274
8.7.4 Blend uniformity measurement 276
8.7.5 Conclusions 276
8.8 Final Thoughts 277
Acknowledgements 278
References 278

9 Acoustic Chemometric Monitoring of Industrial Production Processes 281
Maths Halstensen and Kim H. Esbensen
9.1 What is Acoustic Chemometrics? 281
9.2 How Acoustic Chemometrics Works 282
9.2.1 Acoustic sensors 282
9.2.2 Mounting acoustic sensors (accelerometers) 283
9.2.3 Signal processing 284
9.2.4 Chemometric data analysis 284
9.2.5 Acoustic chemometrics as a PAT tool 284
9.3 Industrial Production Process Monitoring 285
9.3.1 Fluidized bed granulation monitoring 285
9.3.2 Pilot scale studies 286
9.3.3 Monitoring of a start-up sequence of a continuous fluidized bed granulator 291
9.3.4 Process monitoring as an early warning of critical shutdown situations 295
9.3.5 Acoustic chemometrics for fluid flow quantification 296
9.4 Available On-line Acoustic Chemometric Equipment 299

9.5 Discussion 301
9.5.1 Granulator monitoring 301
9.5.2 Process state monitoring 301
9.5.3 Ammonia concentration monitoring 301
9.6 Conclusions 302
References 302

10 Process NMR Spectroscopy: Technology and On-line Applications 303
John C. Edwards and Paul J. Giammatteo
10.1 Introduction 303
10.2 NMR Spectroscopy Overview 305
10.2.1 The NMR phenomenon 305
10.2.2 Time–domain-NMR: utilization of the FID and spin relaxation 309
10.2.3 High-resolution NMR: obtaining a spectrum with resolved chemical shift information 312
10.3 Process NMR Instrumentation 313
10.3.1 Spectrometer and magnet design 313
10.3.2 Sampling and experimental design 316
10.4 Postprocessing Methodologies for NMR Data 317
10.5 Advantages and Limitations of NMR as a Process Analytical Technology 320
10.5.1 Advantages 320
10.5.2 Limitations 321
10.6 On-line and At-line Applications 321
10.6.1 Time–domain NMR 322
10.6.2 High-resolution NMR: chemometric applications 323
10.7 Current Development and Applications 330
10.8 Conclusions 331
References 332

11 Fluorescent Sensing and Process Analytical Applications 337
Jason E. Dickens
11.1 Introduction 337
11.2 Luminescence Fundamentals 338
11.2.1 Luminescence nomenclature 338
11.2.2 Luminescence processes 338
11.2.3 Fluorophore classification 338
11.3 LIF Sensing Fundamentals 341
11.3.1 LIF sensing classification 341
11.3.2 Luminescence spectroscopy 342
11.3.3 LIF signal response function 343
11.4 LIF Sensing Instrumentation 343
11.4.1 LIF photometric instrument specification 345
11.4.2 LIF Instrument selection 347
11.5 Luminescent Detection Risks 347
11.6 Process Analytical Technology Applications 348
11.6.1 Petrochemical, chemical and nuclear field applications 349
11.6.2 Pharmaceutical PAT applications 349

11.7 Conclusions 350
References 351

12 Chemometrics in Process Analytical Technology (PAT) 353
Charles E. Miller
12.1 Introduction 353
12.1.1 What is chemometrics? 353
12.1.2 Some history 354
12.1.3 Some philosophy 355
12.1.4 Chemometrics in analytical chemistry? 355
12.1.5 Chemometrics in process analytical chemistry? 356
12.2 Foundations of Chemometrics 356
12.2.1 Notation 356
12.2.2 Some basic statistics 358
12.2.3 Linear regression 359
12.2.4 Multiple linear regression 361
12.2.5 Principal components analysis (PCA) 362
12.2.6 Design of experiments (DOE) 366
12.3 Chemometric Methods in PAT 368
12.3.1 Data preprocessing 369
12.3.2 Quantitative model building 377
12.3.3 Qualitative model building 389
12.3.4 Exploratory analysis 397
12.4 Overfitting and Model Validation 407
12.4.1 Overfitting and underfitting 407
12.4.2 Test set validation 408
12.4.3 Cross validation 410
12.5 Outliers 413
12.5.1 Introduction to outliers 413
12.5.2 Outlier detection and remediation 413
12.6 Calibration Strategies in PAT 416
12.6.1 The ‘calibration strategy space’ 417
12.6.2 Strategies for direct versus inverse modeling methods 418
12.6.3 Hybrid strategies 419
12.7 Sample and Variable Selection in Chemometrics 420
12.7.1 Sample selection 420
12.7.2 Variable selection 421
12.8 Troubleshooting/Improving an Existing Method 425
12.8.1 Method assessment 425
12.8.2 Model improvement strategies 425
12.9 Calibration Transfer and Instrument Standardization 426
12.9.1 Slope/intercept adjustment 428
12.9.2 Piecewise direct standardization (PDS) 428
12.9.3 Generalized least squares (GLS) weighting 429
12.9.4 Shenk–Westerhaus method 429
12.9.5 Other transfer/standardization methods 429

12.10 Chemometric Model Deployment Issues in PAT 430
12.10.1 Outliers in prediction 430
12.10.2 Deployment software 432
12.10.3 Data systems, and control system integration 432
12.10.4 Method updating 433
12.11 People Issues 433
12.12 The Final Word 434
References 434

13 On-line PAT Applications of Spectroscopy in the Pharmaceutical Industry 439
Brandye Smith-Goettler
13.1 Background 439
13.2 Reaction Monitoring 441
13.3 Crystallization 442
13.4 API Drying 443
13.5 Nanomilling 444
13.6 Hot-melt Extrusion 445
13.7 Granulation 446
13.7.1 Wet granulation 446
13.7.2 Roller compaction 449
13.8 Powder Blending 450
13.8.1 Lubrication 451
13.8.2 Powder flow 451
13.9 Compression 452
13.10 Coating 452
13.11 Biologics 453
13.11.1 Fermentation 453
13.11.2 Freeze-drying 454
13.12 Cleaning Validation 454
13.13 Conclusions 455
References 455

14 NIR spectroscopy in Pharmaceutical Analysis: Off-line and At-line PAT Applications 463
Marcelo Blanco Romía and Manel Alcalá Bernárdez
14.1 Introduction 463
14.1.1 Operational procedures 464
14.1.2 Instrument qualification 466
14.2 Foundation of Qualitative Method Development 466
14.2.1 Pattern recognition methods 467
14.2.2 Construction of spectral libraries 468
14.2.3 Identification and qualification 470
14.3 Foundation of Quantitative Method Development 471
14.3.1 Selection and preparation of samples 472
14.3.2 Preparation and selection of samples 473
14.3.3 Determination of reference values 474
14.3.4 Acquisition of spectra 474

14.3.5 Construction of the calibration model 475
14.3.6 Model validation 476
14.3.7 Prediction of new samples 476
14.4 Method Validation 476
14.5 Calibration Transfer 476
14.6 Pharmaceutical Applications 478
14.6.1 Identification of raw materials 478
14.6.2 Homogeneity 478
14.6.3 Moisture 480
14.6.4 Determination of physical parameters 481
14.6.5 Determination of chemical composition 483
14.7 Conclusions 485
References 486

15 Near-infrared Spectroscopy (NIR) as a PAT Tool in the Chemical Industry: Added Value and Implementation Challenges 493
Ann M. Brearley and Susan J. Foulk
15.1 Introduction 493
15.2 Successful Process Analyzer Implementation 494
15.2.1 A process for successful process analyzer implementation 494
15.2.2 How NIR process analyzers contribute to business value 497
15.2.3 Issues to consider in setting technical requirements for a process analyzer 498
15.2.4 Capabilities and limitations of NIR 499
15.2.5 General challenges in process analyzer implementation 500
15.2.6 Approaches to calibrating an NIR analyzer on-line 502
15.2.7 Special challenges in NIR monitoring of polymer melts 505
15.3 Example Applications 506
15.3.1 Monitoring monomer conversion during emulsion polymerization 506
15.3.2 Monitoring a diethylbenzene isomer separation process 508
15.3.3 Monitoring the composition of copolymers and polymer blends in an extruder 509
15.3.4 Rapid identification of carpet face fiber 512
15.3.5 Monitoring the composition of spinning solution 514
15.3.6 Monitoring end groups and viscosity in polyester melts 516
15.3.7 In-line monitoring of a copolymerization reaction 518
References 520

16 Future Trends for PAT for Increased Process Understanding and Growing Applications in Biomanufacturing 521
Katherine A. Bakeev and Jose C. Menezes
16.1 Introduction 521
16.2 Regulatory Guidance and its Impact on PAT 522
16.3 Going Beyond Process Analyzers Towards Solutions 524
16.3.1 Design of experiments for risk-based analysis 526
16.3.2 Sample and process fingerprinting with PAT tools 527
16.3.3 Design and Control Spaces 528
16.3.4 Chemometrics and process analysis 528

16.4 Emerging Application Areas of PAT 529
16.4.1 Biofuels 529
16.4.2 Biomanufacturing 530
16.5 New and Emerging Sensor and Control Technologies 531
16.5.1 Terahertz spectroscopy 531
16.5.2 Integrated sensing and processing 532
16.5.3 Dielectric spectroscopy 533
16.5.4 Process chromatography 533
16.5.5 Mass spectrometry 534
16.5.6 Microwave resonance 534
16.5.7 Novel sensors 535
16.5.8 Inferential sensors 536
16.6 Advances in Sampling: NeSSI 537
16.7 Challenges Ahead 537
16.7.1 Continuous process validation 538
16.7.2 Data challenges: data handling and fusion 539
16.7.3 Regulatory challenges 539
16.7.4 Enterprise systems for managing data 539
16.8 Conclusion 540
References 540

Index **545**

Preface to the Second Edition

Process analytical technology (PAT) continues to evolve and develop, with new tools and more areas of implementation constantly emerging. In such a dynamic area, it is difficult to be current on all that is new. It is exciting to be able to present a second edition of this book, in an effort to cover some of the more recent advances in the five short years since the release of the first edition.

PAT has been around for some time now, providing a strong foundation of knowledge and well-defined tools that should serve as a starting point for anyone wishing to work in this area. All practitioners can benefit by learning from examples, keeping in mind that the similarities of the technology and approach make them applicable to numerous problems. One needs to be open to the fact that PAT work done in any industry does provide lessons that can be applied to new problems we may have to tackle.

With such a multidisciplinary topic as PAT, one can look at such work from many perspectives: chemist, engineer, manufacturing engineer, controls, regulatory, QA, production, chemometrics. The important thing, regardless of ones specific niche, is to acknowledge that it is truly multidisciplinary in nature and hence requires people from many areas of expertise working as a team to reach successful implementations that deliver business value.

This book is presented from the viewpoint of a spectroscopist, and as such focuses on spectroscopic tools, while also providing some guidance on important considerations for the successful implementation of an analyzer to monitor and control a process. Regardless of the industry in which PAT is used, there is a need to focus on the science and use these tools in the scientific understanding of processes and in the manufacture of quality product, consistently.

The contents of the book are intended to help a newcomer in the field, as well as to provide current information including developing technologies, for those who have practiced process analytical chemistry and PAT for some time. The main spectroscopic tools used for PAT are presented: NIR, Raman, UV-Vis and FTIR, including not just the hardware, but many application examples, and implementation issues. As chemometrics is central for use of many of these tools, a comprehensive chapter on this, now revised to more specifically address some issues relevant to PAT is included. In this second edition many of the previous chapters have been updated and revised, and additional chapters covering the important topic of sampling, and the additional techniques of NMR, fluorescence, and acoustic chemometrics are included.

I would like to thank all of the people that have helped make this book possible, including the numerous teachers and mentors I have had in my life. Of course the strong support of my family allows me to indulge in the exercise of editing, and for this I am grateful. Thanks also to those who contributed chapters to this and the previous edition, as I have learned from each of them. I also salute them for their dedication in writing, when they already have so many other activities in their lives.

Katherine A. Bakeev
Newark, DE
December 2009

List of Contributors

Katherine A. Bakeev
CAMO Software Inc.
One Woodridge Center, Ste. 319
Woodbridge, NJ, USA

Manel Alcalá Bernárdez
Universitat Autònoma de Barcelona
Cerdanyola del Vallès, Barcelona, Spain

Lewis C. Baylor
Equitech International Corp.
New Ellenton, SC, USA

Ann M. Brearley
Biostatistical Design and Analysis Center
University of Minnesota
Minneapolis, MN, USA

John P. Coates
Coates Consulting
Newtown, CT, USA

Jason E. Dickens
GlaxoSmithKline
Research Triangle Park, NC, USA

John C. Edwards
Process NMR Associates, LLC,
Danbury, CT, USA

Kim H. Esbensen
ACABS research group
Aalborg University, campus Esbjerg (AAUE)
Esbjerg, Denmark

Susan J. Foulk
Guided Wave Inc
Rancho Cordova, CA, USA

Paul J. Giammatteo
Process NMR Associates, LLC
Danbury, CT, USA

Robert Guenard
Merck and Co.
Global Pharmaceutical Commercialization
Merck Manufacturing Division.
West Point, PA, USA

Maths Halstensen
ACRG (Applied Chemometrics Research Group)
Telemark University College
Porsgrunn, Norway

Nancy L. Jestel
SABIC Innovative Plastics
New R&E
Selkirk, NY, USA

Linda H. Kidder
Malvern Instruments
Columbia, MD, USA

Eunah Lee
Horiba Jobin Yvon
Edison, NJ, USA

E. Neil Lewis
Malvern Instruments
Columbia, MD, USA

Marcel A. Liauw
ITMC, RWTH Aachen University
Aachen, Germany

Lisa Makein
Malvern Instruments
Malvern, UK

Jose C. Menezes
Institute for Biotechnology and Bioengineering
IST
Technical University of Lisbon
Lisbon, Portugal

Charles E. Miller
Merck and Company
West Point, PA, USA

Patrick E. O'Rourke
Equitech International Corp.
New Ellenton, SC, USA

Peter Paasch-Mortensen
Novozymes A/S
Kalundborg, Denmark

Marcelo Blanco Romía
Department of Chemistry
Universitat Autònoma de Barcelona
Cerdanyola del Vallès, Barcelona, Spain

Joseph W. Schoppelrei
National Geospatial Intelligence Agency
Reston, VA, USA

Michael B. Simpson
ABB Analytical Measurements
Quebec, Canada

Brandye Smith-Goettler
Merck and Co., Inc.
West Point, PA, USA

Gert Thurau
Merck and Co.
Global Pharmaceutical Commercialization
Merck Manufacturing Division
West Point, PA, USA

List of Abbreviations

ADC	analog-to-digital circuit
AE	acoustic emission
ANN	artificial neural network
ANOVA	analysis of variance
AOTF	acousto-optical tunable filter
API	active pharmaceutical ingredient
AQ	analyzer questionnaire
AR	antireflection (as in antireflection, AR coated optics)
ASO	acid soluble oils
ASTM	American Society for Testing and Materials
ATR	attenuated total reflectance
ATR FTIR	attenuated total reflectance Fourier transform infrared
BEST	bootstrap error-adjusted single-sample technique
BSPC	Batch Statistical Process Control
BTEM	band target entropy minimization
CBZ	carbamazepine
CCD	charge coupled device; chemical composition distribution
CD	compact disk
CE	European certification
CE	capillary electrophoresis
CEC	capillary electochromotography
CF	cash flow
CFE	cyclic fluctuation error
CFR	Code of Federal Regulations
cGMP	current Good Manufacturing Practice
CH	constitutional heterogeneity
CHO	Chinese hamster ovary
CIE	Commission Internationale de L'éclairage
CIP	clean-in-place
CLS	classical least squares
COGM	cost of goods manufactured
COW	correlation optimized warping
CPAC	Center for Process Analytical Chemistry
C_{pk}	process capability
CPMG	Carr-Purcell-Meiboom-Gill
CPP	critical process parameter
CQA	critical quality attribute
CQV	continuous quality verification
CRDS	cavity ring down spectroscopy
CSA	Canadian Standards Association
CSE_1	small-scale fluctuations
CSE_3	*cyclic* fluctuations
CVD	chemical vapor deposition
CVF	circular variable filter
DA	discriminant analysis
DAQ	data acquisition
DCS	distributed control system
DH	distributional heterogeneity
DOE	design of experiments
DP	degree of polymerization
DRIFT-IR	diffuse reflectance Fourier-transform infrared
DS	design space; direct standardization
DSC	differential scanning calorimetry
DTGS	deuterated triglycine sulfate (detector)
EMA	European Medicines Agency (formerly known as EMEA)
EMEA	European Medicines Agency (changed to EMA, December 2009)

EMSC	extended multiplicative signal correction
EPA	Environmental Protection Agency
ESR	electron spin resonance
EU	European Union
FALLS	forward angle laser light scattering
FAT	factory acceptance test
FBRM	focus beam reflectance measurements
FDA	Food and Drug Administration
FFT	fast Fourier transformation
FIA	flow injection analysis
FID	free induction decay
FM	factory mutual
FMEA	failure modes and effects analysis
FOV	field of view
FPA	focal-plane array
FPE	Fabry-Perot etalon
FSE	fundamental sampling error
FSP	fundamental sampling principle
FTIR	Fourier transform infrared
GA	genetic algorithms
GC	gas chromatography
GEE	global estimation error
GLP/GMP	good laboratory practice/good manufacturing practice
GLS	generalized least squares
GSE	grouping and segregation error
HCA	hierarchical cluster analysis
HME	hot melt extrusion
HPLC	high performance liquid chromatography
HR	hurdle rate
HR-NMR	high-resolution NMR
HTS	high-throughput screening
ICH	International Conference on Harmonization
ICS	incorrect sampling errors
IDE	increment delimitation error
IEE	increment extraction error
II	initial investment
InGaAs	indium gallium arsenide
InSb	indium antimonide
IP	index of profitability
IPE	increment preparation error
IQ/OQ/PQ	installation qualification/operational qualification/performance qualification
IR	infrared
IRE	internal reflectance element
IRR	initial rate of return
IRRAS	infrared reflection-absorption spectroscopy
ISE	incorrect sampling errors
ISP	integrated sensing and processing
ISPE	International Society of Pharmaceutical Engineers
IT	information technology
IWE	increment weighing error
J–T	Joule–Thompson
KBr	potassium bromide
KF	Karl Fischer
KNN	K-nearest neighbor
LCO	light cycle oil
LCTF	liquid crystal tunable filter
LD	laser diode
LDA	linear discriminant analysis
LDPE	low-density polyethylene
LED	light-emitting diode
LIBS	laser induced breakdown spectroscopy
LIF	laser-induced fluorescence or light-induced fluorescence
LLDPE	linear low-density polyethylene
LOD	loss on drying
LPG	liquid petroleum gas
LTCO	long-term cost of ownership
LV	latent variable
LVF	linear variable filter
MCR	multivariate curve resolution
MCR-ALS	multivariate curve resolution alternating least-squares
MCT	mercury-cadmium-telluride
MCWPCA	mean-centered window principal component analysis
MEMS	micro-electromechanical systems
mid-IR	mid-infrared
MIM	molecularly imprinted monolayer
MIR	mid-infrared
MLR	multiple linear regression
MMA	methyl methacrylate
MON	motor octane number

MPE	minimum practical error
MS	mass spectrometry
MSA	measurement system analysis
MSC	multiplicative scatter correction, multiplicative signal correction
MSPC	multivariate statistical process control
MST	minimal spanning tree
MTBF	mean time between failure
MTTR	mean time to repair
MVDA	multivariate data analysis
MVI	multivariate identification
MWS	multivariate wavelength standardization
N6	nylon 6
N66	nylon 6,6
NAS	net analyte signal
NCCW	no contact check weighing
NDIR	nondispersive infrared
NEP	noise equivalent power
NeSSI	new sampling/sensor initiative
NFPA	National Fire Prevention Association
NIR	near-infrared
NIR-CI	near-infrared chemical imaging
NIRS	near-infrared spectroscopy
NIST	National Institute of Standards and Technology
NOC	normal operating conditions
NPV	net present value
OD	optical density
OEM	original equipment manufacturer
OPA	orthogonal projection approach
OSC	orthogonal signal correction
OTC	over the counter
P&ID	process and instrument diagram
PA	process analytics
PAC	process analytical chemist/chemistry
PACLS	prediction augmented classical least squares
PAI	process analytical instruments
PAR	proven acceptable range
PASG	Pharmaceutical Analytical Sciences Group
PAT	process analytical technology
PbS	lead sulphide
PbSe	lead selenide
PC	personal computer
PC	principal component
PCA	principal component analysis
PC-MBEST	principal component modified bootstrap error-adjusted single-sample technique
PCR	principal component regression
PCS	process control system
PDA	photodiode array
PDS	piecewise direct standardization
PET	photoinduced electron transfer; polyethylene terphthalate
PFM	potential function method
PID	proportional-integral derivative
PIE	process integration error
PLS	partial least squares or projection to latent structures
PLS-DA	partial least squares-discriminant analysis
PLSR	partial least squares regression
PM	preventive maintenance
PMT	photomultiplier tube
PoC	proof of concept
PP	polypropylene
PRIMA	pattern recognition by independent multicategory analysis
PRM	pattern recognition method
PSD	particle size distribution
PUC	process understanding and control
PVA	parallel vector analysis
PVM	particle vision monitor
PX	p-xylene
Q	number of increments per sample
QA	quality assurance
QbD	quality by design
QC	quality control
QCL	quantum cascade laser
QCM	quartz crystal microbalance
r	sampling rate
ra	random sampling
RAM	random access memory
RBG	red blue green
RET	resonance energy transfer
RF	radio frequency
RFI	radio frequency impedance
RFP	request for proposal
RMSEP	root-mean-squared error of prediction

ROA	return on assets
ROI	return on investment
RON	research octane number
RRS	resonance Raman spectroscopy
RSD	residual standard deviation
RSEP	relative standard error of prediction
RTA	real-time assurance
RTM	real-time monitoring
RTO	release-to-operations
RTR	real-time release
SAXS	small angle X-ray scattering
SBR	signal-to-background ratio
SD	standard deviation
SEC	standard error of calibration
SECS	SEMI equipment communications standard
SEP	standard error of prediction
SERRS	surface enhanced resonance Raman spectroscopy
SERS	surface enhanced Raman spectroscopy
SIMCA	soft independent modeling of class analogy
SimDis	Simulated Distillation
SIPP	semi-industrial pilot plant
SLC	system life cycle
SMCR	self-modeling curve resolution
SMV	spectral match value
SNR	signal-to-noise ratio
SNV	standard normal variate
SOP	standard operating procedure
SORS	spatially offset Raman system
SPE	square prediction error
SR	specular reflectance
St	stratified random sampling
STD	standard deviation relative to the mean
SUO	sampling unit operations
SVD	singular value decomposition
SVM	support vector machine
SWS	simple wavelength standardization
sy	sampling scheme; systematic sampling
Sy	systematic sampling
T_1	spin-lattice or longitudinal relaxation
T_2	spin-spin or transverse relaxation time
TAE	total analytical error
TDL	tunable diode laser
TD-NMR	time-domain NMR
TE/TEC	thermoelectric/thermoelectric cooler
TFA	target factor analysis
TFE	time fluctuation error
TOS	theory of sampling
TPI	terahertz pulsed imaging
TPS	terahertz pulsed spectroscopy
TSE	total sampling errors
UL	Underwriters Laboratories
USP	United States Pharmacopeia
UV	ultraviolet
UV-vis	ultraviolet-visible
VCSEL	vertical cavity surface-emitting laser
VPO	vanadium pyrophosphate
WAI	wide-are illumination
WFA	window factor analysis
XRD	X-ray diffraction
ZnSe	zinc selenide

1
Overview of Process Analysis and PAT

Jason E. Dickens

Preclinical Development, GlaxoSmithKline, Research Triangle Park, NC, USA

1.1 Introduction

Process analysis (PA) continues to be an evolving field across various sectors as is evident by its recent adoption within the pharmaceutical industry as an element of process analytical technology (PAT).[1] PA by definition is the application of 'field'-deployable instrumentation (real-time analytics) and chemometrics for monitoring a chemical or physical attribute(s) (CQA) or detection of events that cannot be derived from conventional physical variables (temperature, pressure, flow, etc.). While PA is most often associated with the application of real-time analytics to production problems, the discipline can be considered to be much broader, encompassing sectors outside industrial manufacturing such as environmental, surveillance (chemical or biological agents, explosives, irritant, etc.) and hazmat consequence management. That is, the skills, techniques and instruments are applicable across a wide spectrum of real-time analytical problems. PAT is a broader field encompassing a set of tools and principles to enhance manufacturing process understanding and control (PUC) which includes process analysis, chemical engineering, chemometrics, knowledge and risk management, and process automation and control. Manufacturing quality by design (QbD) in part involves PAT strategies to reduce identified manufacturing risks that are associated with product quality.[1,2]

Real-time differentiates process analysis from off-line laboratory techniques, where the former is on the timescale of seconds to minutes as opposed to hours or days. Furthermore, off-line approaches are often inadequate for root cause analysis in identify the process events that lead to off specification or poor product quality. The basic application of PA involves relative process trending or real-time monitoring (RTM) via the process instrument data stream (e.g., process spectral data). For example, various chemical reactions across industries (fine chemical, polymer, pharmaceutical, biochemical, etc.) can be monitored by *in situ* Fourier transform infrared–attenuated total reflectance (FTIR-ATR) spectroscopy where relative yet meaningful process 'signatures' are extracted via appropriate chemometric treatment.[3] At a higher level, further process understanding is achieved by deriving relationships among multiple data streams including the process instrument data, engineering variables and analytical laboratory reference data. These deterministic

Process Analytical Technology 2e Edited by Katherine Bakeev
© 2010 John Wiley & Sons, Ltd.

models provide real-time monitoring of product CQAs of interest and relationships to critical process parameters (CPPs). This level of process understanding affords a more specific control space definition, facilitates process control and process variance management or real-time assurance (RTA) of product quality. Product parametric real-time release (RTR) is the highest level of PAT where PA results (determinations, end points, etc.) replace conventional laboratory methods. All three levels, to varying degrees, provide a means to increase product quality (e.g., lower product scrap and rework), facilitate cost avoidance, increase production efficiency, reduce laboratory testing requirements and aid in identifying process improvement opportunities, all of which lead to reduced product manufacturing costs and risks.[4,5]

Figure 1.1 depicts the hard and soft fundamentals of process analysis. First, the process instruments range from simple to sophisticated measurement technologies. The majority of process instruments are based upon electromagnetic radiation attenuation as evident by the techniques described herein. Further discussion on real-time instrumentation is provided in Section 1.3.

The next element is the process implementation approach involving either: (i) an in-line interface such as an *in situ* insertion probe (transmission, transflectance or reflectance), an optical flow cell or noninvasive sensors (e.g., acoustics); (ii) an on-line method where an autonomous sample conditioning system containing process instrument(s) is integrated to the process stream. These systems reduce the processing conditions (pressure, flow, etc.) that are suitable for *in situ* or integrated instruments and for appropriate measurement performance. They can also facilitate other analytical requirements such as sample dilution and reagent addition; or (iii) an at-line method where manually or autonomously acquired grab samples are measured with an offline instrument that is proximate to the process or process area as opposed to a remote quality assurance/control laboratory. Simple loss on drying instruments, FTIR or Raman spectrometers for material identification and sophisticated real-time high performance liquid chromatography (HPLC) or flow injection analysis (FIA) systems for QA determinations are examples of at-line approaches.[6–9]

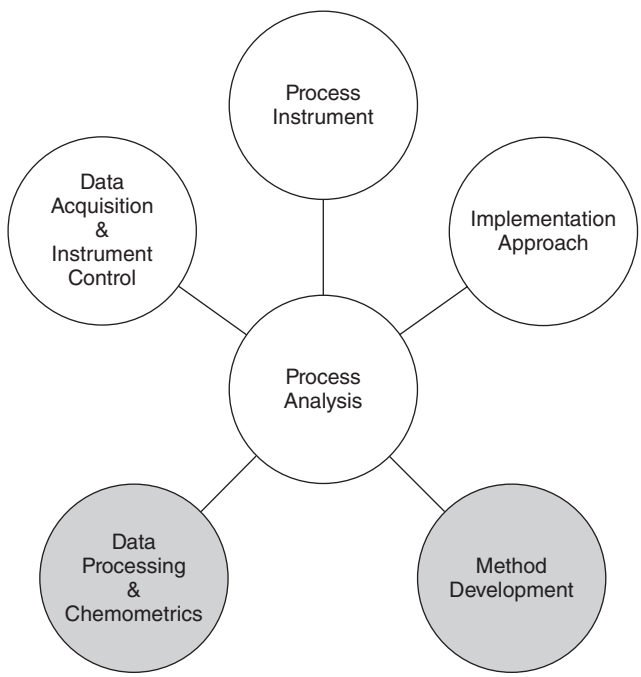


Figure 1.1 *Hard and soft (gray) process analysis elements.*

The data acquisition (DAQ) and instrument control element is an autonomous electronic system that provides several important capabilities including:

- data acquisition and archiving
- remote instrument control (i.e., measurement parameters)
- execution of real-time chemometric models
- instrument diagnostics and real-time measurement quality assurance.

Implementing this level of automation intelligence has been the most difficult to realize within manufacturing industries. That is, while automation controls integration of simple univariate instruments (e.g., a filter photometer) is seamless, it is much more problematic for multivariate or spectral instruments. This is due to the 'tower of babble' problem with various process spectroscopic instruments across process instrument manufactures. That is, the communications' protocols, wavelength units and file formats are far from standardized across spectral instruments, even within a particular class of techniques such as vibrational spectroscopy. Several information technology (IT) and automation companies have recently attempted to develop commercialized solutions to address this complex problem, but the effectiveness of these solutions has yet to be determined and reported.

Data processing and chemometrics are methods for extracting useful information from the complex instrumental and other data stream(s) (see Chapter 12) for process understanding and the development of deterministic models for process control. The final element, the analytical method development life cycle, will be discussed further within this chapter.

PA applications are now found in various manufacturing industries such as chemical, petrochemical, agriculture and food, pharmaceutical and electronics, as well as service industries such as energy and utilities (e.g., water, sewage, etc.). While product quality and production efficiency are most often the objectives of PA, there are several attractive operational-type applications with substantial business benefits. These applications in general are simpler and have shorter development life cycles, often have lower financial risks, tend to require lower capital and staff resource investment, are wider in utility and easier to transfer across groups or sites. Real-time monitoring and control of industrial waste, site environmental monitoring, health and safety area monitoring and equipment cleaning verification are a few examples of operational-type applications.

It is indeed humbling to compose this introductory chapter as there is already a vast array of introductory literature on process analysis.[10–13] It is worthwhile, however, to expand upon these works as the field continues to advance. This chapter is written from a PA experience base spanning three disparate sectors (chemical, pharmaceutical and surveillance) and various real-time analytical problems. The experience includes disparate products (fine chemical, polymer, pharmaceutical materials during product manufacture, etc.) and material physical states as well as PA solutions.

In this chapter I will provide a brief historic perspective, outline the manufacturing drivers for process analysis, provide a high-level overview of process analytical instrumentation, describe the PA method development life cycle prior to implementation and highlight the common pitfalls and challenges within the PA field. I have taken a pragmatic approach herein as the many benefits of PA are realized when a suitable process instrument and method is successfully implemented within a routine manufacturing environment, which is most often a multifaceted endeavor.

1.1.1 Historical perspective

Process analytics began nearly 70 years ago with a rich heritage within the petrochemical and chemical industries.[7] The pilgrimage began in Germany where by the end of World War II their modern plants had

Table 1.1 *History of process analyzers*

Process analyzer technology	Adopted	Origin
Infrared photometers Paramagnetic oxygen sensors Thermal conductivity sensors Distillation-type analyzers	1930s and 1940s	Refining
Real-time gas chromatography[a] On-stream fixed-filter UV-Vis photometry[b]	1950s and 1960s	Chemical/petrochemical
Mass spectroscopy[c]	1970s	Chemical/petrochemical
Near-infrared spectroscopy Multicomponent UV-Vis filter photometry[b] Gas phase FTIR	1980s	Chemical/petrochemical
Mid-IR by ATR[c] On-line HPLC	1990s	Chemical/petrochemical
Process Raman Particle size instruments	2000s	Chemical/petrochemical

[a] Union Carbide, [b] Dupont, [c] Dow Chemical.

been extensively instrumented.[14] In the two decades proceeding World War II, numerous refineries, petrochemical and nuclear plants were applying process analyzers world wide. In more recent decades, sophisticated process analyzers have become more commonplace across several sectors as summarized in Table 1.1. Today process analysis is mainstream within several manufacturing industries and in some cases is an integral component to process control.[7]

An introductory chapter would not be complete without highlighting the alternatives to PA: phenomenological and soft sensing approaches. The former is suitable for processes where the fundamental driving forces can be clearly identified and are generally understood, such as material transport and chemical reactor kinetics. For more complex processes, soft sensing can sometimes be effective in deriving a primary variable of interest from secondary variables. Soft sensing was first introduced by Joseph and Brosillow,[15] and in recent years has been widely studied and applied for industrial process control. Soft sensing is preferred when the critical process variable or attribute is difficult or impossible to determine with a PA solution; product composition in a large distillation column is one such example. The selection of a phenomenological, soft sensing or a process analytical approach is based upon the required performance, implementation risks, routine operational factors and business considerations. For complex higher risk processes, a combination of these approaches (e.g., PA and soft sensing) is attractive as the convergence of independent methods ensures higher PUC assurance.

1.1.2 Business drivers

The goal of lean manufacturing and QbD is robust and efficient manufacturing processes that deliver consistent high quality product at the lowest possible cost where PAT is one route among several in achieving this goal. A fixed and variable costs of goods manufactured (COGM) matrix for a given factory (see Table 1.2) and associated process risk assessments to product quality (i.e., FMEA) provides the framework for identifying PA opportunities. The decision to launch identified PA projects should be carefully weighed

Table 1.2 *Process analysis by COGM*

COGM	Real-time process analysis application
Receiving	Input material QC/QA (e.g.,) identification
Manufacturing	Process end points Process fault detection Quality attributes trending (SPC or MSPC) Process control (feedback or feed forward)
Product release	Determination of final product quality attributes
Plant operations	Real-time water quality Cleaning verification Waste stream monitoring and control
Environmental health and safety	Hazardous area monitoring Environmental monitoring & compliance

QA, quality attributes; SPC/MSPC, statistical process control/multivariate statistical process control; CPPs, critical process parameters.

against the objectives and their probability of success as the workflow is often multifaceted and necessitates suitable resource as the next section describes.

1.2 Execution of Process Analysis Projects

1.2.1 Wisdoms

As we transcend from the information age to the age of 'wisdom and knowledge' as many modern philosophers suggest,[16] it is perhaps fitting to offer the common 'wisdoms' within the field of PA.

- Sponsors often underestimate process analysis projects. Like any other project, PA projects necessitate appropriate staffing and capital resource as well as rigorous project management and leadership.
- While 'turnkey' process instruments are marketed, they (see hard elements in Figure 1.1) often require a degree of customization to realize the desired capabilities.
- Process analysis projects succeed or fail based on the degree of attention to detail and planning.[11]
- A PA team should include experienced plant operators who are involved throughout the development life cycle. That is, plant operators have intimate daily knowledge of the manufacturing process and as such provide 'real-life' production experience that helps aid in developing the current process knowledge base that facilitates developing a suitable PA or PAT strategy. Moreover, they also often become the 'plant champions' and lead trainers and thus promote acceptance of the new process analytical system(s) during development and post commissioning. In addition, process instrument technicians or specialist are also key contributors to a PA team, particularly during the process instrumentation system specification and installation.
- Effective collaboration among the core four PA disciplines: analytical, chemometrics, process engineering and control automation along with other disciplines (e.g., pharmacist, chemist, analyst, product formulators, etc.) is imperative to realize effective PAT solutions that are consistent with the intended lean manufacturing or QbD objectives.

- A common theme among stakeholders is the significant and sometimes surprising process knowledge that is realized upon implementing a PA solution. PA can demonstrate that a given process is not as well understood or robust as previously assumed by phenomenological principles, soft sensing approaches or as determined by conventional means such as off-line laboratory verification testing methods.
- Installation of a real-time instrument and associated method is not always the goal or outcome of a PA project. Most PA projects can be deemed successful on the basis of the significant increased process understanding that is achieved during the development life cycle.
- A real-time process instrument can either be implemented in routine production or be used as a reachback capability for diagnostic purposes during adverse production events.

1.2.2 Team structure

Process analytical projects are most often part of an overriding project such as new product or product line extension development, or continuous improvement initiatives, both of which often involve large multidisciplinary teams. Often a PA or PAT subteam is sanctioned by this wider body to develop the process understanding and control required to deliver consistent product quality and optimal process efficiency. While various organizations construct PA teams differently based on their operating model, Table 1.3 describes the various PA contributors required and the matrix teams that often emerge during the project life cycle. A core group within the PA subteam often drives the project which includes a project manager, a process analytical chemist (PAC), a product/production specialist and a process engineer.

1.2.3 Project life cycle

While Chapter 2 outlines the implementation of PA solutions, it is appropriate to augment this discussion to highlight several prerequisite development steps. Figure 1.2 is a high-level framework for developing a process analytical solution.[17–19] The project initiation phase involves project definition and scoping. The former entails specifying the project problem statement, defining the project goal and outlining the objectives. It also includes a strategic and tactical development plan, a high-level project time line and the required

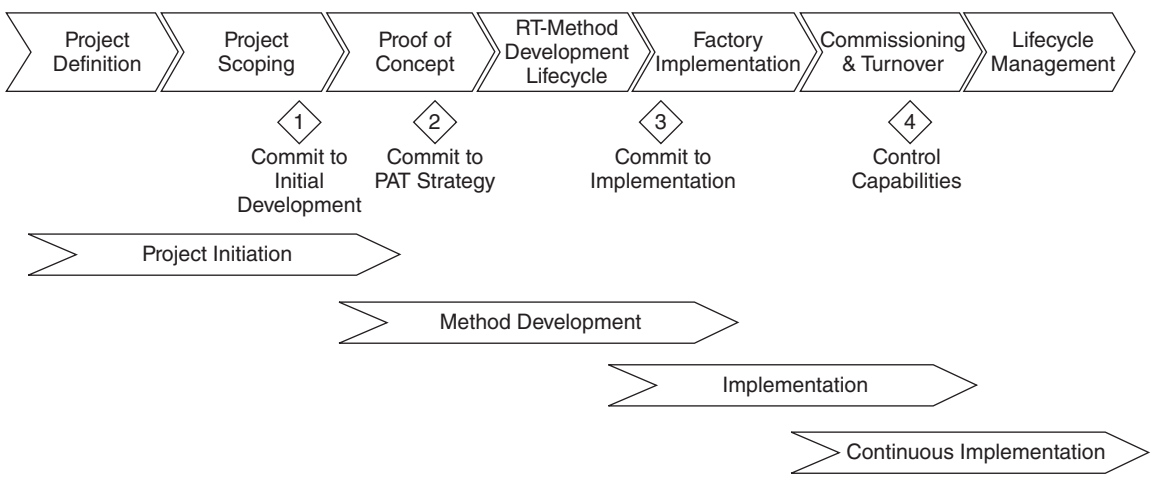


Figure 1.2 *Process analytical method development life cycle.*

Table 1.3 *Typical process analytical contributors*

Matrix team	Contributor	Roles
Project management	Project manager	Project management Capital management Sponsor(s) and stakeholder communications
	Operational excellence expert	Business justification Project metrics Facilitate effective use of lean sigma tools
Analytical matrix team	Process analytical chemist[a]	Process instrumentation selection and specification Process instrument qualification and validation Process analytical method development Analytical Merits and method equivalence Process instrument operation and training Instrument measurement optimization
	Chemometrican	Data management and data fusion Process data analysis Multivariate data analysis Analyzer calibration model development Method equivalence Process models development (e.g., MSPC) Experimental design (e.g., DOE)
	Analytical chemist (analyst)	Reference method selection Reference method capabilities Method equivalence
	Product/production specialist	Product formulation Process steps Input materials and their characteristics
Process engineering	Process engineer	In-line/On-line analyzer integration Process understanding (IPOs, P and IDs, etc.) Facilitate risk assessments Process understanding Process Instrument integration
	Production engineer	Process observations and geology Manufacturing scheduling Coordinates development within production Validation and change control management Operations
	Automation controls	Equipment automation Implementation of control strategies
	Manufacturing representative (e.g., customer)	Production and operational logistics Scheduling and change control Training and SOPs
Production	Plant operations manager	Operator community buy-in Training
	Production engineer	Production experience Operator community buy-in
	Lead operator	
	Process instrument technician/specialist	

resources such as capital costs and staffing. The scoping step is aimed at collecting process details (i.e., engineering walkthroughs), defining analytical requirements (see Section 1.2.4) and developing an understanding of the manufacturing process. The latter includes understanding the process steps, historic trending, mapping the known and potential sources of variance that contribute to the process capability (C_{pk}), accessing existing analytical methods and other knowledge gathering activities such as information about material characteristics and material processing behaviors. This forms the current process knowledge base to develop a suitable PA strategy or PAT strategies. It may also result in realizing a viable simpler solution where existing capabilities could be utilized (e.g., a process engineering solution) rather than developing a new PA approach.

The proof of concept (PoC) step entails accessing various suitable real-time PA techniques which includes instrument capabilities, merits of implementation (i.e., in-line, on-line or at-line) and their viability within a routine manufacturing environment (e.g., robustness, operational complexity, longevity, etc.). It is often useful to determine the maturity of each process instrumental system under consideration, both within the organization and the process instrument marketplace, as it aids in determining the project workflow. That is, a nonstandard, new or novel process instrument is a complex workflow as it involves a significant validation workload, original method development, higher risks and longer time lines. However, the drawbacks of a nonstandard approach should not be the overriding determinant. That is, the common engineering practice of instrumental standardization within large corporations can become an Achilles' heel. While there are sound rationales for standardization such as validation transference, deployment efficiencies (training, procedures, etc.) and regulatory acceptance, limiting the analytical capabilities to specific instrument standards compromises the innovation required to address various process analytical problems. The same applies in over standardization with a particular process analytical technique such as near infrared (NIR).

Initial PoC activities are often at the bench scale where the general process instrument analytical capabilities are determined across several real-time targeted analytical techniques (e.g., NIR, Raman, UV-vis, etc.). If a particular technique has already been determined, it is also worthwhile to evaluate various instrumental variants (e.g., NIR types such as interferometeric, monochromatic, filter photometer, etc.) during this phase. The bench-level investigations are facilitated with either process grab samples or laboratory-prepared surrogate samples. Process samples are preferred provided the process is variable enough to demonstrate measurement capabilities (i.e., accuracy, precision, sensitivity, etc.) across the CQA range of interest (e.g., concentration range). Once the assessment is complete an instrument is selected. The selection is based on the analytical performance and practical considerations such as instrument cost, size, integration and operation complexities, cost of ownership and instrument-to-instrument performance. This information is correlated via a measurement system analysis (MSA) to provide the cohesive knowledge required for the second decision gate.[20,21]

Following the bench studies a project team meeting with key team members and the sponsor(s) is often held to review the analytical data and business case to determine the merits of committing to the piloting stage. Transitioning from the bench to a pilot environment often necessitates capital investment and staff resource to procure the selected process instrument, modify manufacturing equipment and facilitate other change management activities (e.g., compliance, operating procedures, etc.) and thus is an important business decision. The objective of the process analysis piloting step is fourfold:

- determination of the actual process measurement capabilities (sensitivity, accuracy, repeatability and reproducibility, etc.) within the intended real-time application
- optimization of the measurement performance
- developing process understanding to determine the knowledge and control space
- developing required chemometric models and
- evaluating proposed process control strategies

Additional piloting activities include calibration model optimization and verification, determining instrumental robustness (i.e., MTBF) and method and equipment validation. The pilot PoC concludes by revising the MSA as required, completing the business case[22] and reconciling the method performance against the *a priori* acceptance criteria as prerequisites for the pivotal implementation decision gate. Once a PA solution is deemed 'fit for purpose' and the business case ratified, the project moves to the implementation stage as discussed in Chapter 2.

1.2.4 Project scoping

The project scoping provides the framework for selecting a suitable technique, defining the method and its implementation. Generic data sheets are helpful for capturing the information required for formulating the PA development approach.[23] Often these data sheets are working documents throughout the PA project lifecycle. The following list provides the common elements:

Project management

1. Project goal, objectives, timeline and milestones
2. Technical and business rationales: (increase process robustness, process efficiency, improved quality, cost reduction, etc.)
3. Cost considerations

Process engineering

4. Process description (equipment, materials, process steps, etc.)
5. Normal process conditions (temperature, flow, pressure, etc.)
6. Process startup and shutdown effects on product quality
7. Current and desired process capability (C_p)
8. SPC or MSPC review:
   a. process upsets and when they occur
   b. process genealogy
   c. common trends and relationships
9. What is the current process control strategy?
10. Engineering walk through:
    a. process and instrument diagram (P&ID) verification or redlining
    b. determine suitable process analysis monitoring points
11. Define PA constraints
    a. What is the desired PA approach (i.e., in-line, on-line or at-line)?
    b. Define in-line or on-line process or mechanical constraints (e.g., insertion probe size, penetration depth, location, etc.)?
    c. Process instrument limitations (e.g., size, supporting utilities, safety, communications, etc.)

Analytical

12. What is the measurement intent: determining a quality attribute, detecting a process event (end point or fault), etc.?
13. What are the current off-line analytical methods and process analytical approaches?
    a. Define the performance of these methods (accuracy and precision, speed, sensitivity, etc.)
    b. Can they be further optimized?
    c. Can an accepted off-line method be converted to a real-time method?

14. Define the method equivalence criteria or acceptance criteria.
15. What are the analytical requirements of the real-time method? (Accuracy and precision, range, speed, sensitivity, etc.)
16. What are the attributes of interest? How many? What are their ranges?
17. Define other analytical details (i.e., sample matrix, sample state, potential interferants, etc.)?
18. What is the current grab sampling approach?
    a. Frequency, sample size, chain of custody, etc.
    b. Is it a suitable protocol for PA method development (e.g., calibration; see Chapter 3 on sampling)
    c. Is the sampling approach representative of the process?

Operational and automation controls[24]

19. What are the criteria for instrument automation control and data acquisition?
20. Is the PA method used in tandem with other real-time control solutions (e.g., process engineering)?
21. What are the measurement assurance metrics (health, status and diagnostics)?
22. Define the data storage requirements
23. What type of reported real-time result(s) is desired? (e.g., quantitative value, pass/fail, etc.)

Other

24. What is the area classification?
25. What are the validation criteria and when is it required?
26. Define the technology transfer team. Who is the system owner? Who will operate and maintain the PA system? Who are the reachback technical experts post commissioning?

1.2.5 Common challenges and pitfalls

While throughout this chapter various challenges and pitfalls have already been discussed, this section provides the additional common difficulties in realizing a PA solution within routine production. In general PA requirements far exceed laboratory-based analytical methods, even for the simplest of applications. That is, a successfully implemented process analytical solution provides reliable quality data while withstanding the routine factory operational conditions (environmental, work practices, etc.). It also functions with negligible operator and expert intervention. Based on this definition a PA solution necessitates:

- robust process instrumentation with suitable analytical merits
- optimized process integration and operation
- robust chemometric models or data processing algorithms
- representative process data over time and conditions
- autonomous process instrument control and data acquisition
- real-time measurement assurance (i.e., smart sensing)[25]
- suitable information technology and automation controls infrastructure for efficient data fusion and archive management
- sufficient method development, validation and ongoing compliance
- suitable instrument validation and compliance

- a comprehensive onsite process instrument program (metrology, instrument maintenance, training, sufficient on site instrument specialist, etc.)
- identified performance metrics and continuous improvement plans

Inappropriate selection of a suitable process instrument is the most common pitfall hence the focus within this chapter and the literature.[17–19] That is, piloting of various process instruments, including current or accepted PA solutions, is an important PoC activity as it forms the technical basis in selecting an appropriate process instrument. Insufficient process interfacing is also a common pitfall, in particular for in-line approaches. For example, in-line probe location, its angle of insertion and penetration depth requires sufficient investigation to realize superior measurement performance and minimal process impact. Often this is a challenge, as it may require the evaluation of various mechanical interface designs and process equipment modifications. This is more problematic within regulated industries where the resistance to change is much higher, in particular in modifying validated manufacturing equipment. Finally, instituting effective operator training, maintenance and continuous improvement programs during commissioning is yet another common pitfall. This justifies including targeted plant operators and other key plant staff during the PA development life cycle as it greatly aides with these change management activities.

Likewise there are four common challenges. First, capturing sufficient and suitable process understanding development data is often a daunting effort. Many routine processes under normal operating conditions, in particular continuous manufacturing processes, are inefficient in providing the data characteristics required for PA method development. That is, many routine manufacturing processes, once in a state of control, do not vary enough on a routine basis to meet the requirements for PA method development. The criteria for robust empirical chemometric calibration models for example necessitate capturing representative process variability spanning the concentration range of interest and a significant volume of data across particular process conditions. Moreover, deliberate changing manufacturing conditions (e.g., during designed experiments) for process understanding purposes that often produce scrap product are often not a viable option, particularly in the chemical or equivalent industry where the financial margins are often very competitive. Thus, when empirical models are required, the state of the manufacturing process over time determines the rate of progression of the PA method development life cycle. Known seasonal effects further confound this problem. This representative process data problem in particular often necessitates sound technical leadership to manage stakeholder expectations.

Collecting high quality process data and knowledge is also a challenge. While this may be considered a trivial detail for practitioners, orchestrating effective process data collection is a significant team effort as it includes various multifaceted data streams (process instrumental data, engineering variables and reference laboratory data) and disparate groups, often across geographic locations. It also includes data management and fusion, a suitable practical grab sampling protocol and sufficient logging of observed process events that are imperative for developing empirical PAT models. Again, process analysis projects succeed or fail based on the degree of attention to detail and planning.[11] In other words, process manufacturing science necessitates the utmost adherence to the sound scientific practices, effective management and cohesive teams to realize robust PA methods and useful process understanding knowledge to realize effective process quality control.

Unrealistic sponsor expectation is the next common challenge for PA projects. That is, sponsors often treat PA projects analogous to conventional plant engineering upgrades or modifications where rapid return on investment is often expected. Managing expectations throughout a development life cycle is an important element to project success, particularly given the often encountered challenges discussed herein.[11] Finally, PA teams are often comprised of cross-organization members from both research and development (R&D) and manufacturing site(s). While on first inspection this may seem minor, these disparate organizations,

irrespective of the industrial sector, may have widely disparate change management procedures, capital project processes, equipment, engineering standards and data management practices. Moreover, the PA requirements between R&D and at a routine manufacturing site are often not mutually exclusive. That is, R&D necessitates versatile process instrumentation to address the diverse analytical problems encountered within product development. In contrast, factories require robust and automated process instrumental solutions. These differences illustrate the need for effective technology transfer.

1.3 Process Instrumentation

Process instruments are grouped into four categories.[7] Physical property analyzers are the most common, which monitor a physical attribute such as refractive index, thermal conductivity and viscosity. Electrochemical analyzers monitor the voltage or current produced from an electrochemical cell which is related to solute concentration. Examples include conductivity, pH, redox, and trace oxygen analyzers. Combustion analyzers monitor one or more species in a gas or liquid process stream. And spectroscopic, process spectrometers and spectrophotometers, monitor an attribute via electromagnetic interactions (absorbance, emission, scattering, etc.) with the process sample. PA discussed within this treatise fit within this latter category. A thorough treatise of PA instrumentation covering these four categories can be found elsewhere.[7,13,26]

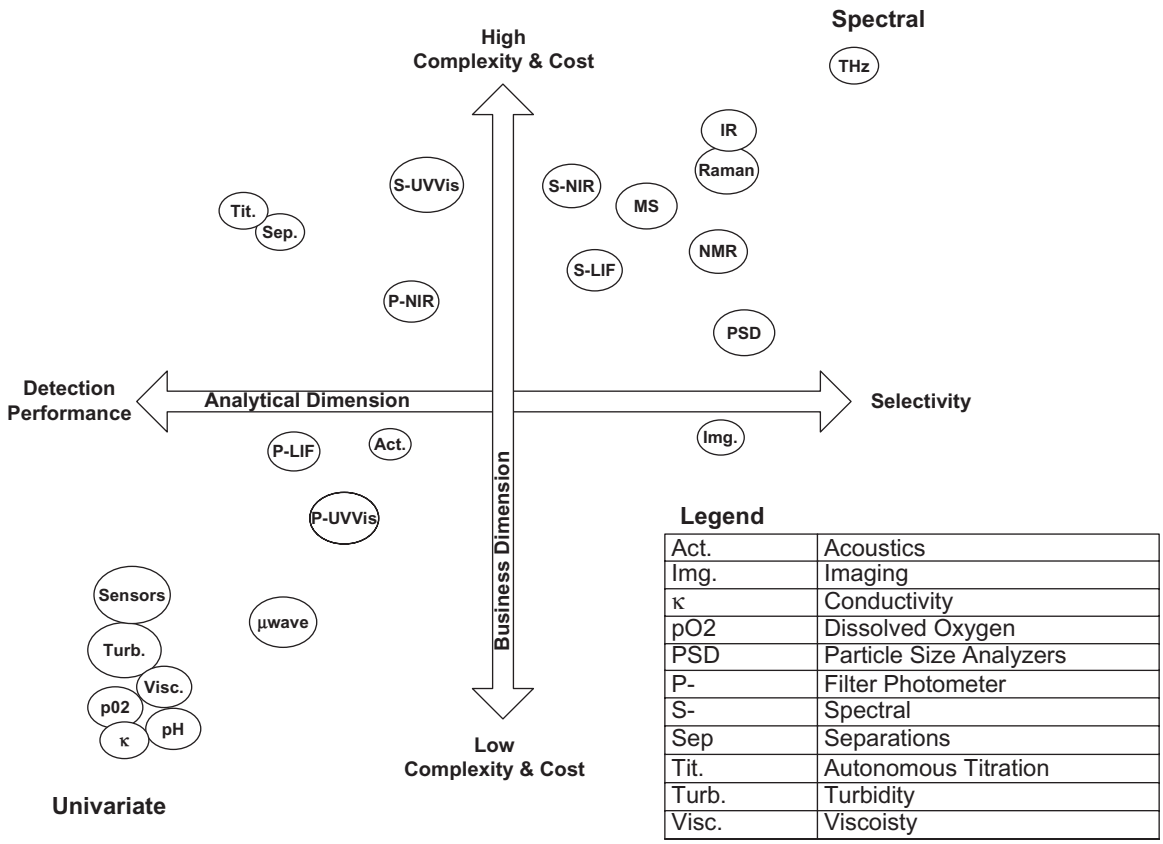
1.3.1 Process instrumentation types

In general terms, process instrumentation fit within two broad categories: analyzers and sensors. Analyzers by definition are large or bulky instrumental 'systems' (e.g., 19-inch rack mountable systems), that necessitate fixed installation, various external utilities (power, plant air, etc.), routing of high cost fiber optics (spectral quality) or cabling from the analyzer to the probe or flow cell and space planning. In contrast, sensors are compact, lightweight and self-contained devices with most of supporting utilities onboard. Small 'handheld' photometers and spectrometers are examples of sensors for PA.[27,28] From a cost perspective analyzers range from $50k to 200k whereas sensors are between $20 to 100k. Sensors tend to be more attractive solutions as they are often much more robust, lower cost, simpler to integrate, easier to maintain and operate, mobile and are suitable for distributed deployment across a manufacturing line or plant. Most current commercial process spectroscopic instruments reside within the analyzer category.

Figure 1.3 depicts a high-level comparison among common process instruments based on 'analytical' and 'business' criteria. The former is a continuum between instrument detection performance (i.e., sensitivity, quantification limits, speed, precision, etc.) and selectivity. The business dimension attempts to quantify the implementation complexities and cost which includes: capital cost, cost of ownership, training, maintenance and implementation and routine operational requirements. While the composite score of each process instrument along these two dimensions may be argued among expert practitioners, the resultant plot would nevertheless highlight the wide disparity among process instruments in terms of their analytical performance and operational-implementation complexities.

1.3.2 Novel process instrumentation

While spectral 'analyzers systems' as described above have served their purpose for several decades as process instrumental solutions, today's complex problems and environments necessitate practical and sophisticated sensors or similar compact process instruments. The electronic revolution over the last several decades has resulted in numerous microelectro-optical devices such as cell phones, PDAs and iPods. These



Act.	Acoustics
Img.	Imaging
κ	Conductivity
pO2	Dissolved Oxygen
PSD	Particle Size Analyzers
P-	Filter Photometer
S-	Spectral
Sep	Separations
Tit.	Autonomous Titration
Turb.	Turbidity
Visc.	Viscoisty

Figure 1.3 *Process instruments classification.*

advances along with other microfabricated capabilities afford the OEM underpinnings for spectral sensors that are beginning to appear on the marketplace.[29–32] Micro-electromechanical systems (MEMS)-based NIR spectrometers and LED sensors [27] are prime examples of modern in-line or on-line process optical sensor technologies.[30,31]

1.4 Conclusions

A fit for purpose process analytical solution provides the analytical and automation capabilities for reliable monitoring of product quality attribute(s) in real-time or process events detection while withstanding the routine factory operational conditions. It also affords smart sensing features to ensure a high degree of measurement assurance for defendable performance that is required for real-time manufacturing control. This introductory chapter has attempted to introduce the multifaceted aspects of process analysis and the complexities involved in realizing a real-time process analytical solution within routine production. The diversity of analytical problems, the depth of knowledge and rigor required is the essence of this rewarding field to process scientists and engineers.

1.5 Glossary of Acronyms and Terms

Analyzer	A large instrumental system (e.g., 19-inch rack mountable system) that necessitate a fixed installation, various external utilities (power, plant air, etc.), routing of high cost fiber optics or cabling and space planning.
COGM	costs of goods manufactured
CQA	Critical quality attribute
DAQ	data acquisition
DCS	distributed control system
FMEA	failure mode and effects analysis
IT	information technology
MSA	measurement system analysis
MTBF	mean-time between failures
PA	process analytics
PAT	process analytical technology
PoC	proof of concept
PUC	process understanding and control
QA/QC	quality assurance/quality control
QbD	quality by design
RTA	real-time assurance
RTM	real-time monitoring
Sensor	Compact, lightweight and self-contained instrument with most of supporting utilities onboard

Acknowledgments

This chapter was made possible in part by several key colleagues who offered their industrial experiences and know-how at pivotal moments during my career. First, Doctors Charles Kettler, Steve Wright, Stephen Jacobs, and Greg Oestreich at Eastman Chemical Company (Kingsport, TN) who introduced me to industrial chemistry and associated analytical problems, and provided pivotal mentorship upon entering my industrial career. Equally my several colleagues at GlaxoSmithKline, including Doctors Luigi Martini, Gordon Muirhead, Dwayne Campbell, and Mr. Anthony Behe also deserve acknowledgment for similar contributions relevant to the pharmaceutical industry.

References

1. FDA [Online], Office of Pharmaceutical Science (OPS) Process Analytical Technology (PAT) Initiative, Available: http://www.fda.gov/AboutFDA/CentersOffices/CDER/ucm088828.htm, accessed 7 December 2008
2. FDA [Online], Q8(R1) Pharmaceutical Development Revision 1, Available: http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm073507.pdf, accessed 16 March 2009
3. A.J. Rein, S.M. Donahue and M.A. Pavlosky, *Drug Discovery & Development*, **3**, 734–741 (2000).
4. A. Brindle, *European Pharmaceutical Review Industry Focus 2008*, 54–55 (2008).
5. A. Brindle and L. Lundsberg-Nielsen, *Pharmaceutical Engineering* **27**, 9–20 (2007).
6. M. Trojanowicz, *Flow Injection Analysis Instrumentation and Applications*, World Scientific Publishing Co. Inc., Hackensack, 2000.
7. T. McMahon and E.L. Wright, Introduction in *Analytical Instrumentation; Practical Guides for Measurement and Control*, R. E. Sherman and L. Rhodes (Ed), Instrumental Society of America, Research Triangle Park, 1996.
8. D. Lee and M. Webb, *Pharmaceutical Analysis*, CRC Press, Boca Raton, 2003.

9. P. Puma, Automated Sampling in the Process Environment, in *HPLC: Practical and Industrial Applications*, J.K. Swadesh (ed.), CRC Press, Boca Raton, 2000.
10. J.M. Chalmers, *Spectroscopy in Process Analysis*, CRC Press, Boca Raton, 2000.
11. B. Connell, *Process Instrumentation Applications Manual*, McGraw-Hill, New York, 1996.
12. D.J. Huskins, *General Handbook of On-line Process Analyzers*, Ellis Horwood Limited, Chichester, 1981.
13. F. McClellan and B.R. Kowalski, *Process Analytical Chemistry*, Blackie Academic and Professional, London, 1995.
14. C.H. Gregory, H.B. Appleton, A.P. Lowes and F.C. Whalen, Instrumentation and Control in the German Chemical Industry, British Intelligence Operations Subcommittee Report #1007, June 1946.
15. C. Brosilow and J. Babu, *Techniques of Model-Based Control*, Prentice Hall PTR, Upper Saddle River, 2002.
16. S.R. Covey, *The 8th Habit: From Effectiveness to Greatness*, Simon and Schuster, Inc., New York, 2004.
17. J. A. Crandall, How to specify, design and maintain on-line process analyzers, *Chemical Engineering* **100**, 94–98 (1993).
18. J.A. Crandall, How to choose an on-line analytical system, *Hydrocarbon Process* **74**, 69–74 (1995).
19. S.M. Jacobs and S.M. Mehta, *American Laboratory*, **5**, 15–22 (1987).
20. D.J. Wheeler and R.W. Lynday, *Evaluating the Measurement Process*, SPC Press, Knoxville, 1990.
21. 6SixSigma [On-line], Measurement System Capability Manual, Available: http://www.6sigma.us/MeasurementSystemsMSA/measurement-systems-analysis-MSA-p.html, accessed: 1 April 2009
22. I. Verhappen, Typical process analyzer application justification, in *Analytical Instrumentation*; Practical Guides for Measurement and Control, R.E. Sherman and L. Rhodes (eds), Instrument Society of America, Research Triangle Park, 1996.
23. I. Verhappen, Specification and purchase of process analyzer systems, in *Analytical Instrumentation*; Practical Guides for Measurement and Control, R.E. Sherman and L. Rhodes (eds), Instrument Society of America, Research Triangle Park, 1996.
24. B. Farmer, Interfacing with process analyzer systems, in *Analytical Instrumentation*; Practical Guides for Measurement and Control, R.E. Sherman and L. Rhodes (eds), Instrument Society of America, Research Triangle Park, 1996.
25. Control Engineering [Online], What is a smart sensor? Available: http://resource.controleng.com/article/CA6296119.html, accessed: 30 June 2009
26. K.J. Clevett, *Process Analyzer Technology*, Wiley-Interscience, New York, 1986.
27. J.E. Dickens, M. Ponstingl, S. Christian, C. Dixon, J. Schroeder and S. Horvath, A novel light induced fluorescent (LIF) sensor for process and field applications, *IFPAC Conference*, Baltimore MD, 2009.
28. MicrOptix Technologies, LLC [Online], Available: http://www.microptixtech.com/, accessed: 30 June 2009
29. J. Workman, Jr., M. Koch and D. Veltkamp, *Analytical Chemistry* **79**, 4345–4364 (2007).
30. R.A. Crocombe, *Spectroscopy* **23**, 40–50 (2008).
31. R.A. Crocombe, *Spectroscopy* **23**, 56–69 (2008).
32. A.P. Demchenko, *Trends in Biotechnology* **23**, 456–460 (2005).

2

Implementation of Process Analytical Technologies

Robert Guenard and Gert Thurau

Merck and Co., Global Pharmaceutical Commercialization, Merck Manufacturing Division, West Point, PA, USA

2.1 Introduction to Implementation of Process Analytical Technologies (PAT) in the Industrial Setting

The field of process analytics (PA) has continued to experience a period of steady growth over the last two decades, driven by the need for manufacturing productivity improvement across all industries. This increase in scope and breadth has been partially fueled over the last 5–10 years by an uptake of activities in the previously only marginally served pharmaceutical industry. Particularly in this regulated environment technical advances are accompanied by regulatory initiatives worldwide, such as the United States Food and Drug Administration (FDA) guidances 'GMPs for the 21st century[1]' and 'PAT – A Framework for Innovative Pharmaceutical Manufacture and Quality Assurance (2004),[2]' and more recently by their promotion of Quality by Design (QbD) through the CMC Pilot Program under the umbrella of a new Pharmaceutical Quality Assessment System.[3] While much of QbD has been formally initiated by the FDA it is becoming more widely adopted by other agencies worldwide, including efforts by the International Conference of Harmonisation (ICH) through published guidance ICH Q8 Pharmaceutical Development, ICH Q9 Quality Risk Management and ICH Q10 Pharmaceutical Quality System.[4] ICH Q11 (Development and Manufacture of Drug Substances – chemical entities and biotechnological/biological entities) is currently in development. The purpose of this chapter is to provide a generalized work process that will serve as a foundation of the fundamentals of implementing process analytical technologies in an industrial setting. A collective body of experience from the process analytics (PA) or process analytical chemistry (PAC) community across industries reveals that there are certain tasks that must be completed in a certain order to ensure the success of implemented process analytical technologies. It is our hope to convey a work process general enough to be tailored to fit diverse industries of various sizes. We feel that this topic is under-represented in the literature and would like to add our perspective based upon our own experience as well as that gleaned from the broader PA community (also see Chapter 15.

Process Analytical Technology 2e Edited by Katherine Bakeev
© 2010 John Wiley & Sons, Ltd.

Our objective in this chapter is to give a general description of important considerations and also a general work process for implementing process analytical technologies. We recognize that our viewpoint is, as always, a contemporary snapshot and highly dependent upon the organization in which the technologies are applied. It is not our contention that this chapter will be all encompassing or provide a one-size-fits-all template for strategic implementation of process analytics.

2.1.1 Definition of process analytics

A broadly accepted definition of 'process analytics' is difficult to capture as the scope of the methodology has increased significantly over the course of its development. What was once a subcategory of analytical chemistry or measurement science has developed into a much broader system for process understanding and control. Historically, a general definition of 'process analytics' could have been:

> Chemical or physical analysis of materials in the process stream through the use of an in-line or on-line analyzer

This definition can be described as 'analysis *in* the process' and is closely related to the traditional role of analytical chemistry in process control. The classical scope of a process analytical method is it to supplement the control scheme of a manufacturing process with data from a process analyzer that directly measures chemical or physical attributes of the sample.

Typically, this definition is accompanied by the characterization of the measurement relative to the process as 'in-line' (probe in direct contact with the sample inside the processing equipment), 'on-line' (the sample is withdrawn from process via a sampling loop; probe not directly part of processing equipment) and 'at-line' (the sample is removed from process but measured in close proximity to process, within the timescale of processing).

This previous definition had been broadened after the FDA's issue of the PAT guidance document to encompass all factors influencing the quality and efficiency of a chemical or pharmaceutical manufacturing process. Driven by developments in Six-Sigma and operational excellence programs an extended definition included such items as:

- development of the process, namely the identification of critical to quality attributes and their relationship to the quality of the product
- design of a robust process to control the critical to quality attributes
- simple sensors and more complex process analyzers
- a systems approach to use and correlate all significant process information including the use of different modeling approaches, ranging from mechanistic to statistical models
- data mining approaches to detect long-term trends and interactions
- potent data management systems to process the large amounts of data generated.

This wider definition can be summarized as the 'analysis *of* the process' and had been developing in the pharmaceutical industry[1] since around 2004–2006 to encourage better use of the information content of classical process analytical methods for the improvement of process development and control. Particularly in the pharmaceutical industry, the acronym PAT for 'Process Analytical Technology' was often being used to describe this newer definition of process analytics.

With the onset of the QbD initiative (see Section 2.3.4), first codified in the ICH Q8 document in 2006, the above-mentioned broader definition of process analytical technology has somewhat evolved into parts of the key elements of the broad definition of QbD, resulting in a partial retraction of the term 'PAT' back

to its historic roots of process analytics or process analytical chemistry (PAC) (i.e. meaning in-line/on-line process analysis). At present the partial overlap between the three terms 'process analytics', 'process analytical technology' and QbD is both a source of ongoing confusion as well as an indication that the field of process analytics continues to see dynamic changes.

2.1.2 Differences between process analyzers and laboratory analysis

Several attributes distinguish process analytical methods from classic laboratory analysis. Most prominent are those differences that refer to the 'directness' of the analysis in the process versus that of one in a controlled laboratory environment:

- The speed of the analysis, which opens the opportunity for live feedback.
- The elimination of manual sample handling with the inherent gain in safety and, to some extent, the elimination of operator error (and being able to maintain sample integrity).
- The general ability to overcome the traditional separation of analytical chemistry and the manufacturing process for even more complex analytical problems.

Possibly less obvious, but equally significant, is the fact that the integrity of the sample is more likely retained when it is not removed from the process which is to be characterized: in contrast to the laboratory analysis this approach can offer true 'process' analysis versus merely sample analysis. This inherent characteristic can be a double-edged sword, however. While the signal from a well-designed and operated analyzer contains a variety of information that can be effectively used for process understanding and control, there are also unique challenges for the robust performance of a process analytical method.

Additional advantages of process analyzers are the possibility to automate all or most parts of the analysis to offer the opportunity of a 24/7 unattended operation. On the downside, the level of complexity and upfront effort during the design, implementation and even maintenance stages can be high and require specialized expertise.

2.1.3 General industrial drivers for PA

The growing use of more complex PAT (versus the historically used simple univariate sensors such as pressure, temperature, pH, etc.) within manufacturing industries is driven by the increased capabilities of these systems to provide scientific and engineering controls. Increasingly complex chemical and physical analyses can be performed in, on, or immediately at, the process stream. Drivers to implement process analytics include the opportunity for live feedback and process control, cycle time reduction, laboratory test replacement as well as safety mitigation. All of these drivers can potentially have a very immediate impact on the economic bottom line, since product quality and yield may be increased and labor cost reduced.

In addition to these immediate benefits, the element of increased process information with the opportunity for continuous improvement is an important driver. With the availability of process data of higher population, especially compared to off-line laboratory tests, the analysis of the existing process scenario, which is one of the first steps of process improvements, is greatly enhanced. At face value this driver might not appear to be as readily attractive as those with immediate financial benefits such as that gained through yield improvement. However the ability of a manufacturing organization to continuously improve its processes is not only attractive from a technological point of view, but can also ultimately generate monetary value. This aspect will be discussed in greater detail in Section 2.2.

For regulated industries an additional driver is the use of process monitoring data to support changes in the manufacturing process to show the equivalency of an enhanced process to a previously registered version.

Regulatory agencies worldwide encourage this use of process analytical information. In particular the US FDA has promoted the use of PAT through recent guidance and publications.[1,2]

2.1.4 Types of applications (R&D versus manufacturing)

There are several types of process analytics applications that can accompany the lifespan of a product or process. The drivers, challenges and benefits associated with these types of methods can be very different from process research and development through to full-scale production.

Some methods are only applied to the initial phases of the development with the goal of defining and understanding the process design space and its relationship to product quality attributes. In this early phase of process development, a broad range of process analytics methods can be used to gain process understanding, by making process measurements during process development experiments. This phase is often called process optimization – finding the most efficient and robust operating conditions that allow the process to produce quality product. The requirements for hardware and operating procedures are typically aligned along those of other R&D technologies. In these types of applications, only a subset of stakeholders are involved in the process analytics implementation compared with the number involved in implementation for a full-scale production plant. For example maintenance, automation and control, and quality organizations may not be involved in multipurpose systems used in the research and development environment.

The next stage of process development is typically the scale-up phase. The goal of process analytics in this phase is to continue to gain process knowledge and/or to verify scalability of the process by comparing process data obtained during initial development with pilot plant or full-scale data. Depending on the scale, more stringent requirements for safety, installation and procedures are typical, especially when actual manufacturing equipment is utilized.

Finally, process analytics methods can be used in commercial manufacturing, either as temporary methods for gaining process information or troubleshooting, or as permanent installations for process monitoring and control. The scope of these applications is often more narrowly defined than those in development scenarios. It will be most relevant for manufacturing operations to maintain process robustness and/or reduce variability. Whereas the scientific scope is typically much more limited in permanent installations in production, the practical implementation aspects are typically much more complex than in an R&D environment. The elements of safety, convenience, reliability, validation and maintenance are of equal importance for the success of the application in a permanent installation. Some typical attributes of process analytics applications and how they are applied differently in R&D and manufacturing are listed in Table 2.1.

2.1.5 Organizational considerations

2.1.5.1 Organizational Support

One of the most important factors in ensuring the success of process analytical methods is the strategic organizational support that is afforded to design, implement and maintain a system. Since most process analytical methods have significant up-front costs and effort, they require managerial support at least on the site operations level, but ideally through a company-wide strategy.

Most expertise/resources for all phases of process analytical methodology development are typically available in the different functional areas of a manufacturing plant that has some laboratory capacity. But these need to be effectively organized. Thus, there is not necessarily the need for additional personnel in manufacturing to implement process analytics. However, the coordination of this existing expertise in the different areas, and the approval of the area management to redirect some resources to process analytics and not merely consider this an ‘add-on’ project are critical for long-term success. Alternatively, most expertise

Table 2.1 *Differences in process analytics applications in R&D and manufacturing*

Item	R&D	Manufacturing
Installation type	Temporary, often mobile setup	Permanent, hard-wired and integrated
Flexibility	Very flexible to changing needs, continuous method improvement	Typically dedicated analysis as part of integrated system solution
Analytical support	Readily available specialized analytical support	Site analytical support less specialized and with more limited resources
Method validation	Basic method validation (short-term use, fit for purpose, little robustness), data will see 'expert eye' prior to release	Extended method validation, robustness and ruggedness tests important for unsupervised operation
Training needs	Only initial training – Usually expert users: vendor introductory training + manuals sufficient	Continued training on different levels (administrator, operator) – dedicated training program
System qualification	Minor qualification needs as long as method/system reliable	Qualification as part of larger system, and with more focus on long-term operation
Maintenance effort	High maintenance of equipment acceptable	Complexity and cost of maintenance is significant factor
Data format	Various (proprietary) formats acceptable, conversion to information manually	Need for standardized data format, validated system for conversion of data
Data volume	Limited amount of data	Large amount of data; archiving and retention issues
Owner of policies	Process analytics developer – R&D policies	Manufacturing policies (including GMP if regulated industry)

can be contracted to third-party organizations such as instrument vendors or implementation specialists (though care should be taken in this approach since there will be some loss of competencies related to the individual organization). Process analytics development can either be undertaken on the individual site/production operations level or through the initiative or the support of a central corporate group. Both approaches have their distinct advantages and challenges.

Site/production operations development Process analytics development as initiated by the site will typically allow the highest level of interaction with the plant personnel and management and more direct access to process experience and knowledge. The business case can originate from needs identified by site personnel or management, and evolve in a process that is managed with local organizational structures and work processes in mind. Most often the aspects of ownership and roles and responsibilities are handled very effectively on the site/production operations level as long as there is continuity in management support of the approach. The downside of a pure site development of process analytics methods is the potential initial lack of specialized process analytics expertise until know-how is obtained by virtue of developing the first system. It is important that site-based projects leverage technology and expertise available in other parts of the organization.

Central group development The development of process analytics applications by central groups and subsequent transfer of ownership to the plant can have certain advantages: specialized know-how and

training is immediately available and will allow synergies with other implementations done previously. A dedicated central group with a significant portfolio of successful implementations can afford better leverage of internal and external resources, such as vendor or other third-party support. The development of a process analytics strategy, previously identified as one of the key elements of the successful use of process analytical methodology in any organization, is often much more effectively done by a group that has some overview and oversight over several different areas and implementations, rather than a site group that works with a more limited scope. An easier alignment of process analytical efforts with central research and development objectives is another advantage of central group development.

The most often cited disadvantages of central group development of process analytics have to do with the interaction and technology transfer to the site: the identification of a local technical or business justification, the relaying of ownership of the process analytics system to the site and the long-term support by site operations groups. It is critical to success that the interests of central organizations and sites are aligned to effectively develop and support the process analytical systems, particularly those with relatively high complexity.

Necessary roles Necessary roles in process analytics implementation can be filled by personnel from different departments, depending on the organizational setup of the company and the complexity of the analyzer installation. A list of required expertise/necessary roles includes:

- management – provides overall sponsorship to project, business support and expertise as well as strategic oversight
- procurement – works with vendors to facilitate and coordinate purchases of equipment
- site/plant operations-contributes technical know-how as well as process/manufacturing logistics. This is typically the recipient and final owner of the technology
- central engineering – includes project, process, mechanical and civil engineering services
- process analytics – provides project management, analytical knowledge, analyzer technology development, analyzer engineering, implementation support and training
- automation and control – handles efforts in the area of process control and automation of process equipment. this group or person may also provide it expertise to the project
- maintenance – will be ultimately responsible for care and feeding of process analytical equipment
- vendor – will provide products and services for process analytics implementation
- quality – potential oversight of data and product quality
- regulatory expertise – necessary for regulatory compliance and strategic partnerships with regulatory bodies.

As has been stated before, not all of these roles need to be occupied based on the scope of the project. Nevertheless a subset of these skills is required for most implementations, and the coordination of efforts is a key element of success.

Profile of a process analytical scientist/engineer Based upon our collective experience, successful process analytical personnel have some combination of the following competencies:

1. *Technical:* Good measurement scientist who understands manufacturing processes
2. *Interpersonal effectiveness:* Project management, interacts well cross-functionally, team player, creates strong partnerships and can resolve conflicts.
3. *Initiative:* Fills roles as needed; responsive and not easily discouraged.

4. *Business focus:* Makes value-based decisions; evaluates on an opportunity/need basis, not purely technical requirements.
5. *Innovative:* Looks for new ways to solve problems, monitors pulse of technology.
6. *Learning:* Continues to learn new skills and grows in expertise throughout career.
7. *Overall leadership:* Provides central role in process analytics implementations; guides and influences to keep often diverse implementation teams motivated and on mission.

Currently, there are few universities or graduate programs that address at least the specific technical qualifications for process analytical chemists. The University of Texas Dallas and the Center for Process Analytical Chemistry (CPAC) at the University of Washington are two graduate programs in the United States with specific training in process analytical chemistry. Outside of the United States, the University of Strathclyde, Newcastle University and the University of Hull (in the United Kingdom) offer curricula in PAC as part of the Center for Process Analytics and Control Technologies (CPACT). The above listed competencies that fall into the 'soft skills' category are, however, just as essential as the technical abilities, and most often provide the key to the success of process analytics in an organization.

2.2 Generalized Process Analytics Work Process

There are many ways to execute capital projects, particularly those which ultimately lead to the implementation of process analytical technologies. Differences in project execution arise from the type of industry, the corporate organizational structure, as well as the technology that is implemented. Furthermore, a distinction can be made between implementations in brand new facilities and those in existing facilities. It must also be defined whether process analytics will be done in R&D versus full-scale routine production setting; rules can be very different in each case. Though no two companies will apply process analytics exactly alike, there are some commonalities surrounding the stakeholders that need be involved and the fundamental project stages that must be completed for successful implementation. In this section a generalized work process is presented which illustrates a rudimentary project structure. Figure 2.1 shows a swim lane flowchart giving a high level overview of a process analytics project from concept through to its release to operations (RTO) and continuous operation. Though not a perfect fit for all organizations and projects, this chart gives a basic road map and was spawned through knowledge and experience across several companies and industries. Substeps and roles will change depending on the industry, but the basic structure should be broadly applicable. A fundamental premise of the work process suggested here is that the implementation will be into a production-scale facility for a permanent installation. We chose this type of installation as a basic premise because it typically requires the most comprehensive project management and work process. Other types of implementations will most likely be some subset of this work process.

The basic idea of this flowchart is to graphically depict the order and roles necessary in the various stages of a project. As can be seen in Figure 2.1, the stakeholders (or roles) are listed in the first column of the chart and occupy a unique row or 'swim lane'. Project steps are sequentially orientated from left to right, with the start located at the far left. Those stakeholders participating in a given step are denoted by inclusion in the project step block. If the swim lane is not included in the project block or the block has a dotted line in a lane, this stakeholder does not have primary responsibility in that project step. Flowchart arrow location indicates the primary stakeholder facilitating each step, gives temporal order to project flow, and allows for the use of decision blocks. Laying out a work process this way maximizes project alignment and serves as a blueprint to success. Given this foundation, each of the steps in the generalized work process will be surveyed in the following sections.

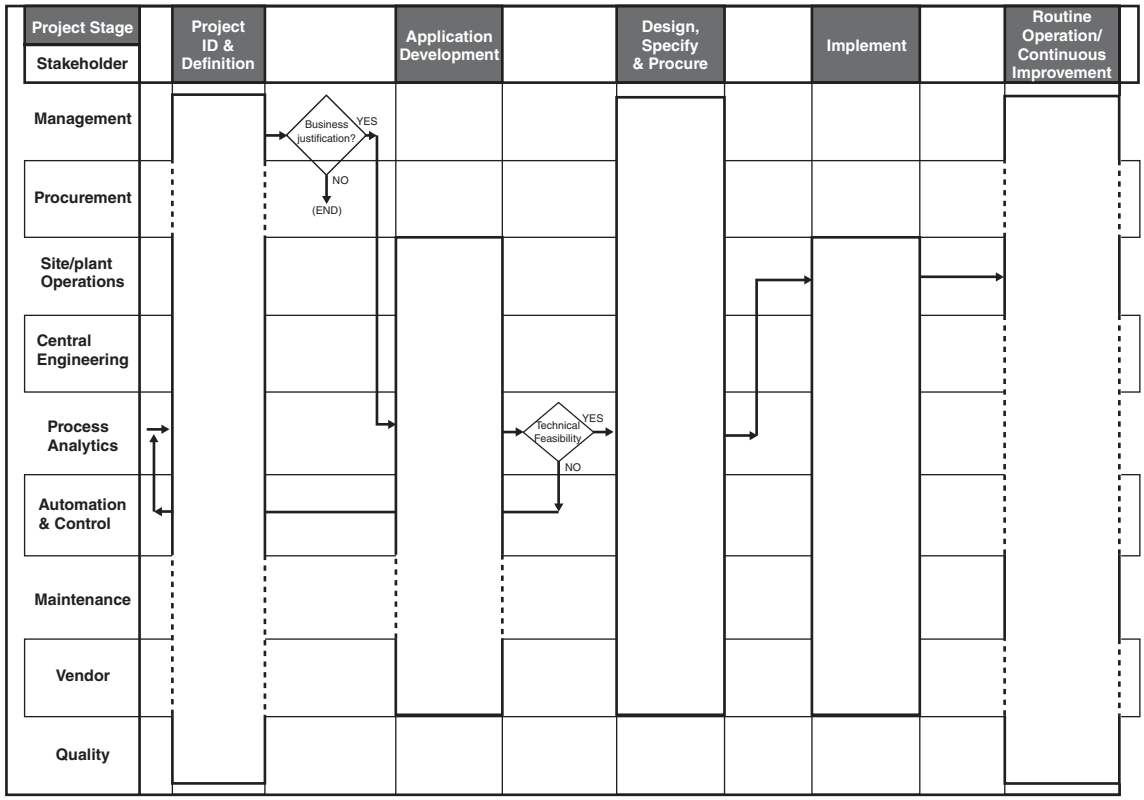


Figure 2.1 *Swim lane flowchart of generalized process analytics work process.*

2.2.1 Project identification and definition

In the initial stage of a process analyzer project, which we call Project Identification and Definition (see Figure 2.1), an assessment is performed to evaluate the business opportunity for application of process analytics, as well as to develop a project plan to ensure a successful process analytics implementation. Specifically, identification refers to the opportunity assessment and definition to codifying the scope and drafting a plan. It is in this project phase that success should be defined and agreed upon by all stakeholders. Ideas for process analytics projects can spawn from many places and people including process control engineers, site production operations and technical support organizations, business leadership, business system improvement personnel, strategic manufacturing initiatives, process analytical chemists and the results of benchmarking within or across industries, to name a few.

A small but diverse team of personnel is typically involved in the activities of the Project Identification and Definition Stage (refer to Figure 2.1). Representatives from groups such as process analytics, process control, project management, central engineering and production operations should populate the team in the first stage. The size and diversity of the team should grow from the identification step to the definition step as resources are allocated and plans developed.

Regardless of the source of the project idea, each idea should be passed through a process analytics opportunity assessment screen of some sort. It has been our experience that a formalized screening process, developed under the premise of consensus with input from all key stakeholders, is the most effective method

of ensuring an implementation that is business prudent, technically feasible and sustainable. An opportunity screen should asses the value added with the process analytics implementation, look at alternatives to implementing process analytics, consider the strategic fit to the production plant and company goals and assess any associated risks (impact and probability). One of the major contributors to the failure of process analytics implementations is a missing or poorly executed opportunity assessment. Projects that proceed based upon an attractive technology alone, without establishing criteria for success, will have a severely diminished likelihood of producing business value. In all cases, the business opportunity must be clearly articulated. Though it is preferable to have implementations that add financial value, some might be considered necessary to conduct business from a safety, regulatory or technical basis (cannot operate the plant without it). Therefore, overall, it is recommended that process analytics projects be treated such that they are screened like every other business proposal.

When calculating the benefit–cost ratio of process analytics, care should be exercised to include all costs and value added over the lifetime of the implementation. Long-term cost of ownership (LTCO) is often used to describe the total cost of an analyzer over the lifetime of the installation; usually this is 10–15 years. Line items in this calculation include costs to engineer, implement, run, maintain and retire the analyzer. Costs as ostensibly small as utilities are included in LTCO because kilowatt hours and purge gas volume, for example, can become significant given an analyzer may be in continuous operation around-the-clock, 365 days a year. Costs are typically much simpler to estimate than the value gained by the implementation, particularly if even only fairly accurate estimates are desired. Cost accounting techniques such as return on investment (ROI), return on assets (ROA), index of profitability (IP), initial rate of return (IRR) and net present value (NPV) are often used to estimate the benefit of capital projects including those of process analytics nature. Net present value is represented by the equation

$$\mathrm{NPV} = \mathrm{CF}^{*}\mathrm{HR} - \mathrm{II} \tag{2.1}$$

Where CF is the cash flow from operation, HR is the hurdle rate or required minimum return and II is the initial investment.[5] Advantages of NPV are that it accounts for the time-value of money, provides a value estimate for the lifetime of the project (10 years) and can always be calculated (unlike IRR). Hurdle rates are set by central finance groups or, in some cases, the project team. If a positive NPV is obtained at a given hurdle rate a project is considered profitable and should be executed based on competing priorities.

Once the process analytics opportunity has been deemed attractive through cost analysis, risk assessment and strategic fit evaluation, project definition can then be executed. In this step the majority of the planning is done and considers resources, technology, process control, engineering and project execution. Logical first steps in project definition are chartering the project team and developing an overall project plan that includes roles, detailed tasks and time lines. Various software tools are available to aid in this planning and tracking. There are essentially three major portions of the plan: the analyzer technology itself, the implementation plan, and a process control plan (if necessary).

Essential elements of the technology plan are completing some type of analyzer questionnaire (AQ), selecting an appropriate technology and surveying potential vendors for the technology. Probably the most critical activity to achieving long-term success of a process analytics implementation from a technical standpoint is defining the application through an AQ. An AQ is a survey that aids in determining the requirements of the application (e.g., desired precision, accuracy, range, components, etc), asks for the physicochemical conditions of the process at the sample point, and inquires into how the data will be used. In other words, this information defines technical success for the application. Armed with the AQ information, developers of the process analytics solution will have a target and can optimize the chosen technology and method for the best fit. Process developers/engineers and control engineers are usually a good source for

the information necessary in the AQ. Clevett describes basic information requirements and the need to have the necessary data early in the project.[6] Once the application target is developed, technology selection can take place. Key operating performance and functional attributes should be addressed in choosing the right technology for the application.[7] Simply stated, the technology should comply with safety and regulatory bodies, give the necessary analytical figures of merit for the application, have acceptable LTCO, be reliable and easily maintained, and be user friendly to those using the data generated. Since there are typically several vendors for any given type of process analyzer, a selection process should be completed. In the definition step, one has only to list potential vendors, with the formal vendor selection process to take place in the design, specify and procure stage.

During the design, specify and procure stage preliminary process control and implementation plans are developed. These plans layout the basic timing/milestones for each effort, as well as define the roles necessary to complete all the given tasks. It is expected that these plans will change over the course of the project with cross-functional coordination and increase in technical knowledge. Key elements of a process control plan include automation considerations and the potential role of process analytics in production recipes. Typically, representatives from automation and control, process analytics and development/central engineering are key players in the development of a process control plan. An implementation plan lays out the specific tasks and timing for receiving, installing, testing and commissioning of the technology. This plan often includes training schedules as well. All of the stakeholders in the project should play some role in the implementation planning because it involves the coordination of so many resources as well as timing with production.

2.2.2 Analytical application development

Using the technical information obtained in the AQ, feasibility studies are done. This stage of the project usually begins as a thought experiment and ends with data showing the technical proof of concept for a process analytics solution. It does not include all of the details necessary to have a successful implementation; these important details are addressed in the design, specify and procure stage of the project. Development of the analytical application is primarily the role of the process analytics scientist, but involves iterative feedback with technical partners in engineering, plant operations and automation and control to assure that the application is on course to fulfill the needs of the process. An experienced and efficient practitioner of developing process analytical applications will have a keen awareness of the available technologies and current science, use designed experiments to gather feasibility data, and perform the experiments necessary to fulfill the needs of the analysis. If available, it is useful to develop and evaluate the technology using a pilot or market development scale production facility. In lieu of these, simulated process conditions in the lab may have to suffice. Closure of this step typically comes with a technical report showing the proof of concept with recommendations for the analytical application that fulfills the needs of the analysis.

2.2.3 Design, specify and procure

Properly designing a process analytical system is crucial to achieving a successful implementation. It is this stage of the project that is arguably the most technically and logistically challenging. Armed with the AQ, business justification and feasibility data, the project team must design, specify and facilitate the purchase and delivery of equipment to the production site. Equally important are the logistics of assuring that lead times are codified and coordinated with site operations milestones, particularly if the process analytical system is being installed in a plant under construction. Key stakeholders in this effort include process analytics, engineering, operations and automation and control on the technical side with management, procurement, project engineering and process analytics groups working on the logistics and resource needs. It is

most critical in this stage of the project that all of the team's expertise is heard from and communication between stakeholders is free flowing. Even the smallest exclusion in this stage of the project can cause large difficulties during implementation and routine operation. A comprehensive description of all the necessary steps to complete the design specifications of a process analyzer system alone could provide enough material for a book. We have decided here to provide a survey of some key points and considerations in this phase of the project. The basic substeps of this stage of the project are designing the process analysis system (drawings included), estimating costs, developing specifications with detailed criteria, selecting suppliers, procuring the equipment and developing logistics and resource plans for implementation.

Process analytics systems consist of the following components: the sampling systems, the analyzer(s), automation as well as a data management system. These components must be carefully selected and tested during the design phase because each part profoundly affects all others. When designing an analyzer system, the following considerations should be addressed:

- the physical layout (must be designed to fit the physical space)
- utilities needs (electricity, gases, etc.)
- equipment identification (tagging)
- sampling system (representative process sample)
- environmental influence (not affected by outside environment)
- user interface (easily learned and operated)
- signal I/O (compatible with control system, fault detection, etc.)
- diagnostics (automated to communicate problems).
- regulatory code compliance (EPA, FDA, CENELEC, etc.)
- reliability (e.g. long mean time between failures 'MTBF')
- safety (to the process and to the operators)
- maintainability (e.g. mean time to repair 'MTTR')
- forward/backward compatibility
- long-term cost of ownership (LTCO, cost over life of installation).

The impact of certain items on this list (reliability, maintainability, compatibility and LTCO) will be felt more profoundly after some time of operation but should be addressed in the design phase as well.

A well-designed sampling system: should provide a sample that is representative of the process; does not introduce any significant time delays; maintains the physicochemical state of the sample; is highly reliable (long mean time between failures); is easily maintained (short mean time to repair); is cost effective and allows safe operation. For general chemical or petrochemical implementations, upwards of 80% of all maintenance problems occur in sampling.[8] A general philosophy of a useful sampling design includes 'the lowest total cost of ownership' over the expected lifetime of the system.[9] Automation and input/output design depend upon the ultimate use of the process analytics and the process control equipment that will be communicating with the analyzer(s). A complex automation scenario would be an analyzer that has an automated sampling system (complete with automated calibration), is used in closed-loop control of the process and is controlled by the distributed control system (DCS) of the production plant. A simple case would be simplex communication from the analyzer to the DCS used in a process monitoring mode. A general rule of thumb when designing process analysis systems is to minimize the complexity while fulfilling the needs of the analysis. Complex implementations can lead to reliability and maintenance problems. The foregoing considerations will take the project team a significant amount of collaboration and time to address. Unfortunately, all too often, operations and maintenance groups are left out of design efforts. Because they are the ultimate owners and caretakers of the systems, they have unique perspective on the long-term considerations that central groups may overlook.

Once the process analysis system is designed, the analyzer specification process can take place. Essentially this involves setting the detailed criteria necessary for the design so that vendors can be selected and equipment procured. Specifications for the analyzer system must be set such that the desired performance of the analysis can be met or exceeded. In particular, this requires a strong operational understanding of the analyzer, which should have been acquired during application development, and how it relates to analytical figures of merit required by the analysis. With this information, a detailed specification for the operating performance of the analyzer and sampling system can be written.

With the specifications for the analyzer, sampling system and automation equipment in hand, a supplier selection process can be initiated. Based upon experience, choosing suppliers is often an emotionally and politically charged event amongst project team members. A nice way of avoiding this, as well as making the best business decision, is to set up a data driven decision matrix that incorporates the needs of the implementation and assesses the ability of the supplier to fulfill them. It is here that procurement groups become highly valuable to the project. They can assist in setting up a decision matrix, sending out RFPs (request for proposals) to potential suppliers and aiding in the selection decision. An RFP from each potential supplier provides much of the basic data needed to make the decision. Once vendors have been selected, quotes can be issued and cost estimates made. These cost estimates and recommendations can then be presented to management for approval. Once approved, procurement can take over to purchase and coordinate delivery of the equipment. An often-requested service of the supplier is a factory acceptance test (FAT). FATs are usually codesigned between the customer and supplier to assure that needs of both parties are met. These tests provide data proving the supplied equipment can meet the specifications, and also serve as a baseline of performance for future trending.

2.2.4 Implementation in production

Site/plant operations and maintenance groups take center stage in the implementation stage of the project; they will become the eventual owners of the process analysis system. Most of the work will be done by operations personnel, usually with support from the central process analytics group and engineering (especially during a new plant startup). A detailed implementation plan should be followed that contains these basic elements chronologically: train personnel, receive and install equipment, safety check, demonstrate operating performance, test automation, document and release to operations. The implementation stage really begins prior to the arrival of any equipment with training of operations and maintenance personnel. If the equipment or technology is new to the site, training is given on the scientific basics of the technique and the details of the hardware. Furthermore, training on the methodology, calibration and routine operation is also recommended.

In collaboration with central groups, and quite often the vendor, operations will install and test the equipment. After receiving the equipment on site and prior to installation, a thorough visual inspection should take place to look for any obvious damage to the equipment. Time line and liability problems can often be mitigated through early detection of shipping damage. The actual installation involves physically fixing the equipment in its permanent location, connecting utilities and signal input/output (IO). Sampling systems and interfaces are installed and connected with the analyzer. Once the analyzer is installed, safety and installation inspections are performed, the analyzer is energized, internal diagnostics are performed and FATs repeated to assure the performance of the instrument remains within specification limits. Sampling systems are inspected and pressure and flow tested to assure the safety and operation of the system. Loop checks of IO connections are performed to test continuity with the control system. In parallel to the analyzer system installation and individual component testing, process automation is implemented by operations, control engineering and automation personnel.

After an evaluation of all of the components of the total analyzer system, the instrument (s) can be calibrated or standardized as necessary. For many types of analyzers (e.g. gas chromatographs, oxygen

analyzers, filter photometers) standards are analyzed and calibrations developed on the installed equipment. On some types of sophisticated spectroscopic analyzers, calibration methods can be transferred electronically by standardizing the instrument.[10] With the calibration/standardization complete, commissioning can take place. Commissioning is essentially the performance check of the total process analysis system. In commissioning, process monitoring is simulated with some type of sample while the analyzer and automation are running. This tests the system as a whole and allows any minor glitches to be addressed before actual production runs. As part of this, there may be some method verification work that occurs by using known samples or with grab sample and lab analysis comparisons to process analyzer measurements (more details on sampling are found in Chapter 3). This process may be referred to as performance qualification, particularly in the pharmaceutical industry (see Section 2.3). After this work has been successfully completed and documented, the ownership of the analyzer is transferred to operations, also known as release to operations (RTO). Support from process analytics does not stop suddenly after RTO; there is usually a transition period when operations gains experience with the system and end users gain confidence in measurement results. In addition, often further method verification data is needed to assure quality of the measured results after RTO.

2.2.5 Routine operation

There are several key factors to the success of the long-term reliability and viability of a process analytical system. Based on experience, it is necessary to have an on-site owner central to facilitating performance monitoring, maintenance and updating of the system to obtain long-term value out of a process analytical system. Overall, production operations are responsible for a large number of systems and equipment. By assigning one or two people as owners of the process analytics system, the issues and intricacies of the implementation will be focused in a small effort. To keep process analysis systems running and reliable, the performance should be monitored, calibrations updated or method verified and preventative maintenance performed. It is optimal if performance monitoring of both the instrument and method performance can be automated. Most modern process analyzers have the capability of performing diagnostics on-the-fly and can communicate these results to the DCS. In this manner, a fault detection alarm is triggered when there is an instrument problem, allowing plant operators to address problems as necessary. Example situations that can be automated as fault alarms are fouled probes, clogged filters, burned-out light sources and low purge air pressure to name a few.

Regardless of the type of analyzer system or application, verification or calibration of methods must be performed periodically or be triggered by some type of event. There are two main reasons for a method to need updating: a change in the instrument response, or a change in the sample. All instruments have some inherent drift, but the amount of that drift and its effect on analysis results is what is really important. Therefore, a periodic check of the analyzer predictions through either check standards or grab sampling and comparison with lab analysis as well as evaluation of outlier diagnostics (such as Mahalanobis distance or spectral F-Ratio) should be done. Furthermore, when adjustments or replacement of analyzer hardware takes place, the method predictions should be verified. Changes in the sample such as raw material supplier, composition, or physical state could affect the calibration method. Thus, anytime there is a change in the process it must be communicated to the analyzer owner so that the effect of the change on the method can be evaluated. This is usually taken care of procedurally in a change control process (see Section 2.3).

Analyzer systems that contain consumable or limited lifetime parts must undergo periodic preventative maintenance (PM) to keep performing at a high level. Typical consumable parts include spectrometer light sources, chromatography columns, sampling filters and pump seals to name just a few. Maintenance organizations are commonly charged with keeping maintenance schedules, logging performance, performing the actual maintenance and shelving spare parts. Central groups and production operations often work closely

with the maintenance group to establish PM schedules and inventories on spare parts. Both of these will be optimized as experience is gained with a particular analyzer system.

2.2.6 Continuous improvement

With the increasingly competitive environment in the manufacturing sector, organizations are always searching for new ways to reduce costs, improve efficiency and product quality. To remain competitive, many companies have devised and implemented corporate-based continuous improvement initiatives. As part of continuous improvement, process analytical implementations should be periodically evaluated to determine if they are still necessary and adding value to the production process.

Implementation of process analytics in the pharmaceutical industry for continuous improvement purposes has received increased attention in recent years. Process analytics or PAT (see previous definition) has been identified as a key factor in continuous improvement initiatives and is seen as a central element of emerging regulatory strategies to enable these efforts as part of QbD.

2.3 Considerations for PAT Implementation in the Pharmaceutical Industry

2.3.1 Introduction

The purpose of this section is to describe the differences between process analytics implementation in what will be broadly differentiated as the 'chemical' versus the 'pharmaceutical' industries. In the present context petrochemicals, polymers, specialty and basic chemicals are examples from the chemical industry. In contrast, this section reflects on process analytics in the pharmaceutical industry, where the manufacturing of product can be divided into two distinct phases: the chemical synthesis of the active pharmaceutical ingredient (API), and subsequently the formulation of the API into a drug product, often a solid dosage form like a tablet or a capsule. While the first phase, the synthesis of the API, is technically very similar to manufacturing processes in the chemical industries, there are still unique regulatory and economic drivers that set the pharmaceutical industry apart. For the second phase, the formulation of the drug product, the distinction from the chemical industry extends also into the technical sector, with an even larger element of regulatory considerations. To some extent the technical similarities to process analytics in the formulation of the drug product can be found within the food and the agricultural industry. The processing methods for biologics, including vaccines, therapeutic proteins, and other biopharmaceutical products, comprise another distinct application area within pharmaceuticals that is outside of the scope of this contribution.

The best practices for the successful implementation of process analytics applications are common to all industries. There are, however, differences in the business model (driver), technical and particularly regulatory areas that are unique to the pharmaceutical industry. The following section seeks to briefly outline these distinctions.

2.3.2 Business model

The business model of the innovative pharmaceutical industry is to transform the results of basic medical research into products that provide health benefits to the patient. This model is characterized by a high risk of failure in the development of new products, coupled with the benefit of market exclusivity for a number of years if a product can be successfully developed and is approved by the regulatory agencies. Only very few chemical entities complete the development process from drug discovery to a commercial product. But those products that do reach the marketplace have to yield sufficient financial rewards to assure that the

overall medical research program is productive. Typically, the manufacturing costs in the pharmaceutical industry account for a much smaller portion of the overall cost during the lifetime of a product, hat is, from discovery to the patient. Of greater influence to the profit over product lifetime is the speed-to-market, by which a product can be moved from discovery through development to a registered commercial product. Delays in gaining regulatory approval for new drugs are costly because they negatively impact the period of marketing exclusivity. And while regulatory approval risks are typically not considered to be determined by the section of the regulatory dossier which describes the manufacturing process development and control (but instead by the results of the clinical trials), new pharmaceutical process applications were usually modeled after existing ones that have previously passed regulatory scrutiny. On the one hand process analytics is seen as a promising tool for manufacturing innovation and continuous improvement, but these benefits must be matched against any potential costs associated with product development time line risks that may be incurred. The proper balance between the risks and benefits of PAT application is being sought within each company in the pharmaceutical industry. The regulatory aspects of the QbD initiative (see Section 2.3.4) were established to support the use of innovative manufacturing approaches in new products by providing the framework to identifying and mitigating risks.

Once approved by the regulators, commercial manufacturing processes are launched and the principles of continuous improvement are applied. These principles have been widely used in the chemical industry due to their large potential for economic benefits and now also have been adopted in the pharmaceutical industry, partially enabled by the QbD initiative.

2.3.3 Technical differences

For the two pronounced phases of drug manufacturing, the technical differences for the use of process analytics between the chemical and pharmaceutical industries are more pronounced in the manufacturing of the formulated drug product than in the chemical synthesis and purification of the API.

2.3.3.1 Active Pharmaceutical Ingredient (API) Manufacturing

The manufacturing of the API is in principle a synthesis of fine chemicals with high quality requirements for the final product and a tight external regulatory framework. In API manufacturing, the process steps, physical and chemical characteristics of the sample and equipment are very similar to the chemical industries:

- a series of synthetic, separation and purification steps
- samples are often liquid or slurries
- if solid samples are present, they are processed as batch units
- common unit operations equipment like reactors, separation columns, large-scale dryers.

Many technical solutions developed for process analytics in the chemical industry consequently also apply in the pharmaceutical industry. Analyzer systems and implementation principles are essentially the same in both industries (see Chapter 15).

2.3.3.2 Drug Product Formulation

The techniques used in the manufacturing processes for formulating final drug products are often very different from those of the chemical industry, with some similarities to food or agricultural manufacturing processes. Although significant research has been invested, and progress made in academia and industry to

develop a better first-principle based understanding of the chemical and physical aspects of individual process steps (often referred to as 'unit operations') there is still an opportunity for improved mechanistic understanding of the influence of process parameters on critical-to-quality attributes in pharmaceutical manufacturing.

With respect to the use of process analytics, the heterogeneity of the material in the process (i.e. powders and granules) often poses the largest technical challenge. The aspect of sampling, particularly with the elements of a representative sample size, sample presentation to analyzers and probe fouling deserve special attention in method development. (See chapter 3 for more information regarding sampling).

Another industry-specific aspect of quality control is both a technical challenge for the implementation of process analytics in the pharmaceutical manufacturing process, as well as one of the great opportunities: the elimination of typical reliance on finished product testing to assess the quality of the product. The potential of PAT to control the quality of the product during and not after processing and, as an added benefit, quantifying and reducing process variability in pharmaceutical manufacturing is more recently being explored as one of the most promising drivers for the use of process analytics. This quality control approach, also commonly referred to as 'real-time release' (RTR), was virtually unknown at the time of issuance of the FDA PAT guidance (2004).[2] Since then there is a small but growing number of registered and marketed products which are tested or released in real time, i.e. as soon as batch manufacture is complete.

A short summary of conceptual technical differences with respect to process analytics between the chemical and pharmaceutical industry is listed in Table 2.2.

Another attribute that favors the use of process analytics and which previously had not been relevant in the pharmaceutical industry also has more recently been conceptually introduced – the shift from traditional batch to continuous processing. Although a broad presence in the routine manufacturing environment is likely still years away most major pharmaceutical companies have internal or external development efforts underway to reap the benefits of continuous processing, such as potential for tighter quality control, reduction in scale-up risks, and more flexibility in responding to changes in demand. Just like in those industries where continuous processing has been utilized for a long time, the pharmaceutical industry will also require the use of process analytics and feedback loops to keep these processes in control.

Table 2.2 *Comparison between chemical and pharmaceutical industry processes with respect to process analytics*

Industry	Chemical	Pharmaceutical
Process type	Continuous and batch	Batch, (some continuous processing efforts underway)
Process inputs	Ranges from crude feed stocks with variable properties to high purity materials	Pure, high quality raw materials
Sample type	Homogeneous or heterogeneous as gasses, liquids, slurries, solids	Heterogeneous, often blends of solids
Control	Approaching control limit that gives the highest yield	On target at the center of the specification range
	Many in-process controls often with limited final product testing. Process monitoring	Reliance on final product testing (release testing) – some examples of real-time release
Product specifications	Broad range of products tailored for distinct applications (potentially very tight tolerances)	Narrow range of product specifications

2.3.4 Regulatory aspects of process analytics in the pharmaceutical industry – the concept of quality by design

The pharmaceutical industry is regulated worldwide through national health authorities such as the FDA in the nited States. One important element of pharmaceutical production regulations are the current Good Manufacturing Practices (cGMPs) – established in the Code of Federal Regulations (CFR)[11] for the United States. In other countries similar regulatory documents exist to ensure that the drugs produced are 'safe, pure and effective.'

While the detailed interpretation of these regulations is the responsibility of individual companies, the overall guidance of the cGMP has ensured the manufacture of high quality products while not specifically encouraging innovation in manufacturing. Since about 2004 however, a significant change in the posture and vision for the regulation of pharmaceutical process development and commercial manufacturing has occurred in the United States and worldwide. In 2004 the FDA developed a guidance 'cGMPs for the 21st century'[1] to provide a risk-based approach to regulations, which was designed to remove the barriers to innovation. This was then followed by the issuance of the FDA PAT guidance developed through cooperation with industry and academia. More recently the concept of QbD has been providing the regulatory and technical framework for a number of innovations including the increased use of process analytics.

2.3.4.1 Quality by Design and the Link to Process Analytics

The QbD initiative within the pharmaceutical industry represents a large change initiative supported jointly by industry, academia and regulatory agencies worldwide. While a common understanding of details, scope and work process of QbD is still evolving one simple and commonly accepted general definition is that, 'quality should not be tested into a product, it should be by design'. This definition was first outlined in the International Conference on Harmonisation (ICH) Q8 guidance 'Pharmaceutical Development', which is often cited as the genesis of the QbD regulatory approach. Together with the subsequently issued ICH Q9 guidance on 'Quality Risk Management', and the ICH Q10 document 'Pharmaceutical Quality System' these three guidances embody the concepts of QbD in process and product development, technology transfer and commercial manufacturing. While the authors of this chapter do not seek to create a comprehensive review of the QbD approach here, it is important to review the central concept of the design space and its relevance to the field of process analytics.

The design space of a pharmaceutical product is defined in ICH Q8 as the multivariate space which defines raw material and process responsible for acceptable product quality. From a regulatory viewpoint, products that are registered based on a design space could offer the opportunity to apply changes to the manufacturing process within the design space. The regulatory approval of these changes could be defined as part of a post-marketing plan (PMP), versus a case-by-case prior approval as required by more traditional interpretations of pharmaceutical regulations. The link between QbD and process analytics is evident in all phases of the life cycle of a pharmaceutical product. While the design space development using process analytics methodologies is essentially a concept based within the product and process R&D phase of the life cycle, the approach to controlling the process to stay within the design space (often referred to as the QbD Control Strategy) extends the impact of process analytics into commercial manufacturing. A particular aspect of process analytics, the application of multivariate modeling for process monitoring and control, has a unique position to enable the use of true multivariate design spaces (and related control strategies) in manufacturing.

Process analytics have already been utilized in process development, realizing all the benefits of the methodology in the R&D environment without facing particular regulatory scrutiny as often assumed for commercial manufacturing applications. With the increased use of concepts like design of experiments,

robustness analysis and accelerated technology transfer, process analytics can provide an enabling role to achieve the benefits of these methodologies.

The true paradigm-changing effect of the QbD concept for process analytics, however, has been in its use in commercial pharmaceutical manufacturing. While prior to the FDA PAT guidance there was a definite perception of risk of regulatory acceptance to using process analytics for control purposes in manufacturing, this scenario has almost been reversed with the realization that for many pharmaceutical processes the methodologies offer a superior way to analyze variability, monitor performance and control quality. The broad consensus among the QbD community is that while QbD is not equivalent to (only) process analytics (or PAT) the use of on-line and in-line testing as well as real-time model control are a substantial element and key enabler to realizing the benefits of QbD. As the industry is anticipating the introduction of more complex and truly multivariate design spaces the use of multivariate data modeling and analysis will allow the use of control strategies for advanced design spaces in manufacturing.

2.3.4.2 Compliance with cGMP in the Pharmaceutical Industry

As established previously, the cGMPs form the regulatory framework for R&D and manufacturing in the pharmaceutical industry. The specific impact of operating under cGMP rules on process analytical work will be discussed in the following section with special focus on current regulatory guidance and the impact of cGMP on new technology introduction in the field of process analytics.

Regulatory guidance on process analytics The specifics of regulatory guidance in the pharmaceutical industry are defined in a number of publications, such as:

- Country- or region-specific government documents such as, for example, the CFR in the United States.
- Pharmaceutical compendia issued in countries or regions that detail procedures to test the quality of pharmaceuticals (e.g. US Pharmacopoeia, European Pharmacopoeia, International Pharmacopoeia, etc.)
- The ICH guidances.
- Other standards organizations such as the ASTM International who has been active in the area of PAT since late 2003 in the development of standards for the use of PAT in the pharmaceutical industry internationally in their standard committee E55 on Manufacture of Pharmaceutical Products

Most guidance documents on testing in the pharmaceutical industry do not presently address the relevant differences of process analytics systems to conventional laboratory systems. For example, the very narrow definition of required performance characteristics for a laboratory analyzer, combined with the often mandated use of certified standards for instrument suitability tests in pharmaceutical compendia, can be challenging to implement with a process analytics system using the same technology in manufacturing. Along the same lines, method validation elements that are scientifically sound and essential in the laboratory setting might be less relevant and more technically challenging to satisfy in the process analytics domain.

Innovative technology introduction in a regulated environment In order to comply with regulations, pharmaceutical companies need to establish quality systems that define in detail all aspects of the process that are used to manufacture product. The main element of such a quality system is the thorough documentation of the manufacturing process and the related product testing. The documentation requirements are in fact a major part of the work process when developing or implementing process analytics in the pharmaceutical industry. This section does not intend to describe these requirements in detail or to serve as a manual for proper documentation. Instead, two elements that are unique to the pharmaceutical industry with particu-

lar relevance for the introduction of modern analyzer technology, as it is often used in process analytics, will be discussed: system qualification and change control.

System qualification is a process that ensures that an analyzer system is installed and operated according to requirements that are aligned with the intended use of the system. The commonly used approach in the pharmaceutical industry is the 'system life cycle' or SLC process. In the SLC approach, the definition of intended use, design, configuration, installation and operation is linked and documented over the lifetime of a system.

The SLC document typically consists of technical documents with system descriptions. It also has references to standard operating procedures (SOPs) and administrative and maintenance systems. A completed SLC documentation set is typically required prior to using analyzer systems in the pharmaceutical laboratory or manufacturing site. It is also required prior to use of the data obtained by the system in GMP.

Some of the concepts of instrument qualification using the SLC concept can potentially delay the introduction of process analytics. With evolving technologies as they are used in process analytics, information for some of the elements of a traditionally designed SLC are likely generated only once the system is being used. The development of a process analytics method might require some feasibility studies with access to real process equipment, as the requirement specification of the system or method might be related to the process capabilities, which can only be quantified by actually measuring them with the process analytics system. In contrast, for laboratory analytical systems that use external standards instead of real process samples to verify system suitability, the SLC concept historically mandates that these specifications are set prior to the start of experiments. It will be important for the application of SLC requirements for process analytics analyzers to not put the 'cart before the horse' and instead to conduct a thorough quality and business risk analysis for the intended use of the instrument rather than prescriptively following the historic practices set forth in the application of the SLC concept to laboratory analyzer systems.[12,13]

Change control is an important principle in both the pharmaceutical and chemical industries. There are several types of change control: for example change control of systems (process analyzer), change control of analytical methods or change control for processes. Change control takes on a regulatory meaning in the pharmaceutical industry. Common to these different types of change control is the need to monitor the status of the system/method/process with respect to the initial qualification/validation and to put measures in place to verify that the changes do not affect the GMP status of the system and the validity of the assumptions under which the system/method/process was operated prior to the change. One rationale behind change control has obvious practical benefits for system integrity as it allows the operation of systems independent of a particular owner, and minimizes the risk of unwarranted modifications to the system by nonexpert or unauthorized parties. In the field of process analytics, where the performance of systems/methods/process can be influenced by a large number of factors, the formal documentation helps to identify and manage the risk involved with a modification. Typical quality systems in pharmaceutical companies handle modification through equipment, analytical method and process change requests. However, similar to the concept of process validation discussed previously, it can also have a negative impact on the use of continuous improvement principles and puts a high resource cost even on necessary changes to the system.

One practical example of the consequences of change control is that in the regime of system/instrument change control, hardware, software and operational aspects such as routine operation, maintenance and data integrity are locked down in fixed configurations. Any modification that has not been anticipated in the initial qualification requires testing and documentation.

Analytical change control is the monitoring of any changes to analytical methodology, which has to be justified and evaluated for its performance and impact on the quality of the product. For process analytical methods, the change control approach requires significant documentation and can impede the efficient optimization of the method in the early phases of implementation if parameters have to be locked down based on an insufficient data set.

The goal of this section was to outline current differences between the implementation of process analytics in the chemical and pharmaceutical industry, specifically in the areas of business model, technology and regulations. The reader of this chapter should be aware that this status quo is changing rapidly in the pharmaceutical industry due to a changing business climate, combined with a more technology-aware regulatory environment in the major pharmaceutical markets. The authors of this chapter would expect that future editions of this book would see major revisions to this comparison of the industries.

2.4 Conclusions

It was the intent of the chapter to present a generalized work process for implementing process analytical technologies in the industrial setting. Based upon a work flow captured in a swim lane chart, each of the major stages in a process analytics project have been described, with some detail as to the substeps, necessary organization input and expertise, as well as an introduction to some of the project tools and implementation considerations. As stated previously, it is not claimed that this is a comprehensive, one size fits all approach, but merely a thumbnail sketch of what must occur to successfully implement process analytics. In describing the work process, we spoke to the differences and considerations when implementing in the chemical industry and the pharmaceutical industry, where regulations in manufacturing are significant. We hope that this presentation of a work process will open discussion and spawn further work to improve how companies can optimize the benefit gained through process analytics.

References

1. US FDA 'Pharmaceutical cGMPs for the 21st Century: A Risk-Based Approach' http://www.fda.gov/oc/guidance/gmp.html, Fall 2004, accessed 11 November 2009.
2. US FDA 'PAT – A Framework for Innovative Pharmaceutical Manufacturing and Quality Assurance' http://www.fda.gov/downloads/AboutFDA/CentersOffices/CDER/UCM174306.pdf, September 2004, accessed 11 November 2009.
3. M.M. Nasr, US FDA, A new pharmaceutical quality assessment system for the 21st century, available http://www.fda.gov/downloads/AboutFDA/CentersOffices/CDER/ucm105873.pdf accessed1 December 2009.
4. International Conference on Harmonisation, http://www.ich.org/cache/compo/276-254-1.html, Guidances ICH Q8, Q9, Q10 and Q11, accessed 11 November 2009.
5. R. S. Polimeni, S.A. Handy, and J.A. Cashin, *Cost Accounting*, 3rd edn. McGraw-Hill, New York, 1994, pp. 158–160.
6. K.J. Clevett, *Process Analyzer Technology*, Wiley-Interscience, New York, 1986, pp. 839–842.
7. J.D. Tate, P. Chauvel, R. D. Guenard and R. Harner, Process monitoring by mid- and near-infrared Fourier transform spectroscopy, in *Handbook of Vibrational Spectroscopy*, volume **4**, J.M. Chalmers, and P.R. Griffiths, (eds), John Wiley & Sons Ltd, New York, 2002, pp. 2738–2750.
8. K.J. Clevett, Process analytical chemistry – industry perspectives – trends in applications and technology, *J. Process Contr. Qual.* **6**, 81–90 (1994).
9. J. Crandall, Sample systems in near infrared spectroscopy, Part 1 *NIR News*, **5**(5), 6–8 (1994).
10. R.N. Feudale, N.A. Woody, H. Tan, A.J. Miles, S.D. Brown and J. Ferre, Transfer of multivariate calibration models: a review, *Chemom. Intell. Lab. Syst.* **64**, 181–192 (2002).
11. Code of Federal Regulations 21 CFR part 210 'Current good manufacturing practice in manufacturing, processing, packing, or holding of drugs; general' and part 211 'Current good manufacturing practice for finished pharmaceuticals, Food and Drug Administration, Department of Health and Human Services.
12. R. Schadt, Process analytical technology – changing the validation paradigm, *Pharmaceutical Rev.* **7**(1) 58–61, January/February (2004).
13. D.C. Watts, A.M. Afnan, and A.S. Hussain, Process analytical technology and ASTM committee E55, *ASTM Standardization News*, **32**(5), 25–27 (2004).

3

Process Sampling: Theory of Sampling – the Missing Link in Process Analytical Technologies (PAT)

Kim H. Esbensen[1] and Peter Paasch-Mortensen[2]

[1] *ACABS Research Group, Aalborg University, campus Esbjerg (AAUE), Esbjerg, Denmark*
[2] *Novozymes A/S, Kalundborg, Denmark*

3.1 Introduction

Process analytical technology (PAT), and its predecessor process analytical chemistry (PAC), includes chemometric analysis and modeling based on appropriate data quality with respect to what sensor signals/measurements [*X*,*Y*] represent in the industrial process setting. Both [*X*,*Y*] data must be acquired by representative sampling, ensuring accuracy with respect to lot characteristics and precision with respect to all non-eliminated sampling and analytical errors. Chemometric data models must closely adhere to reliable performance validation, for example regarding prediction, classification, or time forecasting. Without representative process sampling the chain of evidence versus lot characteristics is flawed because of inherent material heterogeneity at all scales. This applies both to samples as well as sensor signals. It is not possible to correct for nonrepresentativity deficiencies in subsequent data analytical modeling. Since analytical errors typically are one or two orders of magnitude smaller than the combined sampling errors, the dominant aspect of 'data quality' is in practice almost entirely dependent upon sampling.

From the theory of sampling (TOS), a working minimum of principles are delineated for both zero-dimensional (0-D, batch sampling) as well as one-dimensional (1-D) lots (process sampling), including heterogeneity characterization, systematization of eight sampling errors and seven practical sampling unit operations (SUO) with which to combat heterogeneity [1]. A summary of the salient criteria to be met in order to satisfy these demands in the PAT regimen is presented. The effect of nonrepresentative sampling manifests itself as an inconstant sampling bias, which in the sampling instant can neither be estimated nor corrected for. Nonrepresentative sampling also leads to inflated sampling imprecision (reproducibility),

Process Analytical Technology 2e Edited by Katherine Bakeev
© 2010 John Wiley & Sons, Ltd.

which, if unnoticed, leads to futile, and unnecessary, data modeling efforts. In nearly all situations however, sampling errors can be significantly reduced and the sampling bias can in fact be completely eliminated by respecting a simple set of SUOs and guidelines. These issues are not solved by buying a new piece of equipment however, as many current OEM products fail to respect the TOS, as is illustrated here.

A systematic approach is described for practical process sampling which is based on in-depth knowledge of the specific variability of 1-D lots (process lots). An introductory catalogue of typical sampling misunderstandings and equipment fallacies is presented along with the only current setup for representative process sampling (upstream sampling) which is based on existing process technology. Selected practical examples illustrate the use of 1-D variography to design optimal sampling protocols for a variety of typical PAT situations. Variography is also an important quality control (QC) and quality assurance (QA) tool which can be deployed for both absolute as well as relative QC/QA monitoring.

PAT was originally greeted with near euphoria due to modern sensor technologies: instead of having to extract samples and laboriously get them to the analytical facility (be it at-line or in the central laboratory) one now has only to install appropriate sensor technologies directly in the flowing stream of matter. Imagine having only data in the form of signals, ready-made for chemometric calibration. The only facility lacking was a standard multivariate calibration, which explains the enormous initial PAT popularity within chemometrics. Alas, many unresolved issues and problems went unnoticed – gradually manifested by a noticeable reluctance in many communities (spectroscopic, chemometric, process analytical). By 2005 a comprehensive overview addressed these issues squarely: 'Chemometrics and PAT: What does it all mean[1]?'. Among many insights and valuable issues, propositions and solutions for a more synergistic chemometrics-PAT interaction, there was one glaring sin-by-omission:

> Statistics alone does not demonstrate cause and effect. The technical team (…) must explore the underlying principles causing variation within the data (…) to assess the fundamental causal process information. (…) The technical advantages of using chemometrics include: speed of obtaining real-time information from data; the capability of extracting high quality information from less resolved data. (…) Chemometric-based real-time measurements eliminate the greatest challenges to 100% compliance and analytical accuracy for measuring a process – namely, sampling error.

This chapter can be viewed as a massive refutation, complete with all alternative solutions, of the fatal flaw in the last sentence above. Chemometrics most emphatically does *not* eliminate sampling errors [sic].

This chapter deals with the necessity of representative sampling in the context of PAT. All PAT sensors need to be calibrated with respect to relevant, reliable reference data (Y data). This presupposes that representative samples are at hand for this characterization – but *sampled how*? Additionally, X signals (X measurements) need to be qualified as representative of the same volume as was extracted for Y characterization, or at least a sufficiently 'well-matching' volume. How does one demonstrate this in a quantitative manner? If the quality of both X and Y data involved is suspect, how can a multivariate calibration be expected to be trustworthy? This also includes the issue regarding proper validation of the chemometric multivariate calibration(s) involved, which can only be resolved based on proper understanding of the phenomenon of heterogeneity. The TOS delivers answers to all these issues. The TOS constitutes the *missing link* in PAT.

In response to the often heard misconception that sampling is but a statistical and practical issue, the following contextualization is highly relevant for process monitoring and control. Statistical considerations

[1] J. Workman Jr., Chemometrics and PAT: What does it all mean?, *Spectroscopy* special issue: The Role of Spectroscopy in Process Analytical Technologies, 18–20, 22–23 (January 2005).

include the accuracy of the analytical estimation with respect to a preselected level of tolerable risk or uncertainty. It is understood and accepted that the lower the tolerable uncertainty, the more laborious sampling will have to be (the more costly, perhaps somewhat more 'impractical' than today's procedures, which do not originate from in-depth understanding of heterogeneity or representative sampling). It is essential to be able to distinguish between a sampling bias – which can be reduced or eliminated following the TOS, but which is often neglected due to 'practical and economical reasons' – and the remaining 'statistical' sampling variance. These aspects are clearly discriminated in the TOS' definition of 'representativity'. Within the TOS' framework it is fully possible to derive complete, objective, reliable estimates of the total sampling errors (TSE) accompanying existing or suggested sampling plans and decide on the most appropriate sampling procedures. Nonstatistical considerations include such factors as financial, required resources and time constraints. Unfortunately these often dominate or dictate many current sampling protocol designs (ISO, CEN, etc.), with the consequence that more approximate sampling protocols with large risks and *de facto* unknown uncertainties are routinely used – ostensibly 'to save time and money'. While it is not the responsibility of scientists to define the acceptable risk threshold (this is a company, corporate, agency or political responsibility), science would be remiss if it did not elucidate the very serious consequences of irresponsible, voluntary, slack acceptance of *only* these nonstatistical issues.

The purpose of sampling is to obtain a mass-reduced sample which is *representative* with respect to the lot: representativity concerns both *accuracy* and *precision.* The TOS gives strict definitions of all necessary concepts and terms needed to obtain representative samples. However, there will always be a certain minimum influence from sampling errors leading to an inherent uncertainty in the analytical results. The job of the TOS is to reduce this fundamental sampling error to its absolute minimum. While it may appear easy to grab an(y) indiscriminate lot portion conveniently at hand and to analyze it, often with the best analytical method available, such ill-reflected, simplistic extraction and the subsequent analytical results will in reality be worthless if the sampling process is not representative. Contrary to this grab sampling approach, the TOS stipulates always to employ compositing several *increments* from the lot. According to the TOS, all nonrepresentatively mass-reduced lot portions should rather be termed 'specimens' to clearly distinguish them from representative 'samples'.

In order to appreciate process sampling a minimum of the principles behind batch sampling of so-called 0-D lots is needed first. Following is a basic introduction with full referencing to the background sampling literature [1–14].

3.2 Theory of Sampling – Introduction

Representativity is intimately related to the concept of heterogeneity. From a distance, the material in a batch, stockpile, truck load, reactor or pipeline may at first glance appear more or less homogeneous; that is, all constituents are apparently identical. Constituents are the smallest elements that remain indivisible during the selection process. In the TOS these are called *fragments*, which may be grains, particles, molecules or ions, as the case and scale may be. On closer inspection, it will become obvious that visual differences are often observable, and moving down to the level of each fragment, especially as concerns their composition, lots and materials are very nearly always significantly heterogeneous. This is only a matter of scale, as all fragments will in practice *never* be strictly identical. Differences with respect to composition, shape, size, density, etc. will always occur. Within practical sampling, the term homogeneous should be discarded, as it is only a theoretically interesting limiting case, never encountered for any naturally occurring material.

Practical sampling makes use of an *increment*, which may, or may not be identical to a sample. Composite samples are by definition made up of several increments.

From direct inspection of any lot portion alone, it is *not* possible to ascertain whether it is a representative sample, or a specimen only. Focus must instead be exclusively on the sampling process. The basic prerequisite for a representative sampling process is that all potential increments from a batch, container, or pipeline have the same probability P of being selected, and that the elements selected are not altered in any way after the sample/increment has been taken. Furthermore, all elements that do not belong to the batch or container must have zero probability of being selected, for example, leftovers at the bottom of the lot, or leftovers in the sampling equipment from previous increments/batches (cross-pollution).

The relative sampling error is defined as:

$$e = \frac{a_S - a_L}{a_L} \tag{3.1}$$

in which a_S is the analytical grade of the sample (mass of analyte divided by the total mass of the sample) and the analytical grade a_L is the mass of analyte divided by the total mass of the lot.

A sampling process is said to be *accurate* if the average sampling error m_e practically equals zero. Likewise the sampling process is said to be *reproducible* (high precision) if the variance of the sampling error is less than a small predetermined value, $s_e^2 \leq s_0^2$. Accuracy concerns the sufficient reduction of the deviation between the sample grade and the true lot grade. A significant systematic error manifests itself as the dreaded sampling bias.

The notion *representative* is a composite property of the mean square error, which includes both the systematic (bias) and random part (statistical) of the sampling error [2]:

$$r_e^2 = m_e^2 + s_e^2 \tag{3.2}$$

The TOS defines as correct any sampling procedure which is accurate (property of the mean). If it is also reproducible (property of the variance) then it may be qualified as representative. The central issue is then that even for truly representative sampling, in practice any specific analytical result a_S is an estimate of the true (average) a_L only. This is due to the influence of the fundamental sampling error (FSE), more of which below.

The random component of the sampling error, represented by the variance, tends to decrease when averaging over a large number of samples. The systematic part, however, does not. It is essential to assure an accurate (hence unbiased) sampling process in order to cancel the systematic part of the sampling error. When – and *only* when – such correct sampling is in place, the potential in the TOS lies in characterizing and minimizing or eliminating as many of the remaining sampling error variances as possible, that is, bringing the sampling variance in statistical control.

A sampling situation is also classified according to the specific spatial nature of the material being sampled. TOS operates with 0-, 1-, 2-, and 3-dimensional lots (0-D, 1-D, 2-D, 3-D) [1–4]. In 0-D sampling, the lot (batch of material) sampled is definite in space. A 0-D lot can be manipulated – at least in principle – by for example, mixing. But the defining characteristic is that there is no intrinsic spatial autocorrelation between individual potential increments. Examples of 0-D lots include containers, big bags, truckloads of material, heaps, batches. Reactor vessels, while clearly constituting 3-D frameworks, may – for some sampling considerations – also be envisaged to enclose 0-D lots according to the above definition. No confusion need arise in specific cases (however, close attention shall be given below regarding ambiguous cases).

For 1-D sampling, one dimension in space dominates the lot completely as all increments extend fully over the other two physical dimensions, and there is a natural linear order between increments. Conveyor belts and process pipelines transporting a material flux are the two most important examples of 1-D lots.

Process sampling in time forms a complete analogue to sampling along a geometrical/geographical 1-D direction.

Caveat: In 2-D and 3-D lots, where two and three dimensions are dominating, refer to larger and/or more complex, spatial bodies. The dimensions in 3-D lots are all of equal importance and the dimensionality can hence not be reduced. Examples of 3-D lots are geological formations, such as mineral deposits and large, isolated, piles such as shiploads, and stockpiles that are too big to be manipulated for sampling purposes without excessive effort. For 2-D lots the depth dimension can sometimes be considered of vanishing importance. 2-D lots in practice are often flattened. 3-D lots include strata bound formations or the content of a storage bed. Other examples come from the environmental realm, e.g. when taking soil samples of constant depth (or from a constant depth interval) for all increments. True 3-D lots are always difficult to sample representatively. For large material lots, such as mineral deposits, the discipline of geostatistics provides a viable solution. This chapter shall henceforth deal only with 0-D and 1-D lots.

3.2.1 Heterogeneity

All materials are heterogeneous at two distinct levels: compositionally and spatially. It is necessary to fully understand the intricacies of the phenomenon of heterogeneity in order to be able to counteract this characteristic of all lots/materials. Lot/material heterogeneity, together with sampling process deficiencies (measures against which are fully delineated below), are the sole reasons for all sampling errors.

There exist a few classes of materials that would appear exempt from the stringent principles and rules regarding naturally occurring materials delineated in this chapter: 'fine' or 'pure' chemicals, reagents, powders and liquid solutions. Fine powders can be of uniform composition, with sampling variability less than any predetermined threshold, say, 1% (rel.) or less, either because they have been manufactured for this specific purpose (production of reagent grade fine/pure chemicals inevitably includes the most thorough mixing conceivable) or because of exceedingly thorough comminution and mixing as part of the process. Certified reference materials (CRM) fall into this category. Liquid solutions are even easier to mix as they show compositional heterogeneity only at the molecular level. Thus analytical chemistry often differentiates sampling of 'homogeneous' liquid solutions prepared from mixing chemicals in the laboratory. Uniform materials are the only other case in which the sampling bias problems dealt with here are practically absent. Sampling of such materials does not give rise to specific problems and can be accomplished on the basis of conventional statistics, in which the precision is proportional to the inverse square root of the number of samples. Statistical reference works and textbooks on analytical chemistry deal comprehensively with this ideal situation. While the above classes of materials undoubtedly exist, they are but rare *exceptions* to the complete suite of materials that are analyzed using PAT approaches, however. Thus the PAT expert is not exempt from being fully conversant regarding the TOS and its derived practical sampling principles.

In what follows below a distinction is made between two distinct scale levels with which to view any lot: a fragment level and an increment level. It will be necessary to contemplate the lot at these scale levels to be able to define the following two key features:

- constitutional heterogeneity: CH – for the lot, termed CH_L
- distributional heterogeneity: DH – for the lot, termed DH_L.

3.2.2 Constitutional heterogeneity

The CH is an intrinsic property of the material to be sampled, in which the individual units (particles, grains, etc.) are considered unalterable and indivisible (*if* comminuted by the sampling process. The TOS calls all such units fragments.

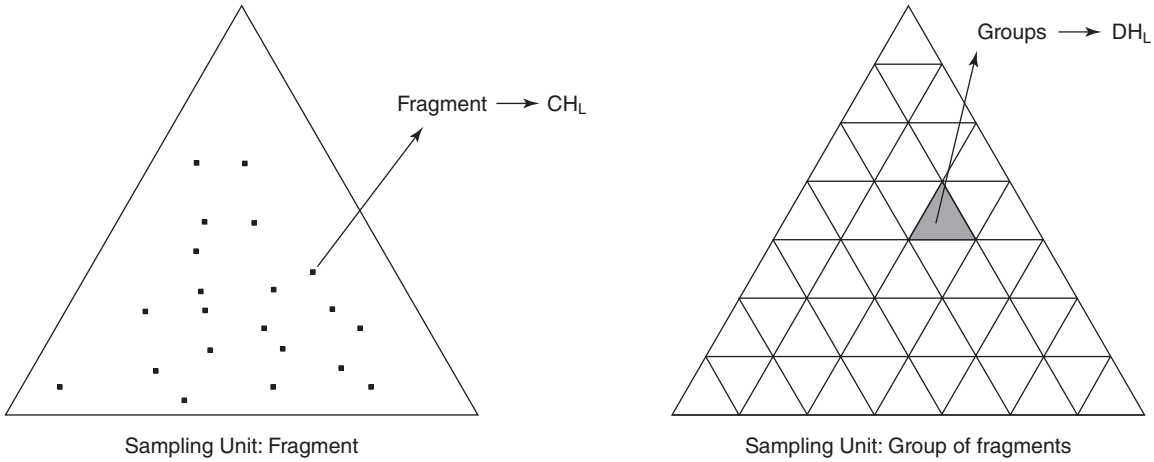


Figure 3.1 *Principle illustration of scale-related concepts of constitutional and distributional heterogeneity. The constitutional heterogeneity, CHL, expresses all between-fragment compositional differences in the entire lot, while the distributional heterogeneity, DHL, reflects all between-increment (between-group) compositional differences, which can be seen as covering the entire spatial geometry of the lot.*

A lot (material) has a homogeneous constitution only when all units are strictly identical, while termed heterogeneous if otherwise.

The amount of heterogeneity h carried by one fragment, F_i, and mass M_i, is defined as:

$$h_i = \frac{(a_i - a_L)}{a_L} \cdot \frac{M_i}{M_{\bar{i}}} \tag{3.3}$$

The constitutional heterogeneity of all N_F fragments in a lot can therefore be described as:

$$CH_L = s^2(h_i) = \frac{1}{N_F} \sum_i h_i^2 \tag{3.4}$$

The essential feature is that CH_L is defined as the variance of all heterogeneity contributions from all fragments which together make up the whole lot (Figure 3.1).

If one lot portion is extracted alone (below to be termed a 'grab sample'), this can never constitute a representative sample by itself, due to the intrinsic constitutional heterogeneity displayed by any heterogeneous material at large. Any two such specimens can never give rise to identical analytical results. The specific sampling error arising from this nonconstant material heterogeneity is termed the FSE. Constitutional heterogeneity is a material characteristic related to the differences in composition *between* the fundamental, smallest inseparable fragments making up the total lot. Since this heterogeneity is intrinsic, mixing and so-called 'homogenization' has no effect on it.

3.2.3 Distributional heterogeneity

The DH appears at a larger scale when dealing with different increments (in the TOS, groups of spatially neighboring fragments) and their interrelationships. Distributional heterogeneity is thus the most direct

concern because all sampling in practice is carried out by extracting increments. DH is nil if the samples consist of single fragments only, which in practice are never relevant. DH can therefore be viewed and used to characterize the added variability stemming from the practical necessity of sampling some 'practical' sampling volume. To be stringent this sampling volume shall be known as the increment volume (or the increment mass) in the following. The increment volume is the singular scale level of importance in practical sampling.

DH is ultimately related to the constitutional heterogeneity. But much more importantly, it takes account of the spatial heterogeneity between all potential increments making up the lot from the practical sampling point of view. This aspect of lot heterogeneity is affected by the omnipresent gravitational force pervading our environment as well as by any agent or force involved in the making of the lot. Macroscopic lot heterogeneity is very often created when the lot is made up/laid up in the first place, while the mesoscopic spatial heterogeneity is more a matter related to the aggregate behavior of the particular material in question. If the lot is subjected to forceful mixing, DH_L is reduced. By contrast, segregation leads to an increase in DH_L.

The heterogeneity carried by an increment within the lot can be described by an analogous equation as for the fragments. Observe however the critical step up in scale, from fragments to increments, for which reason the index now refers to increments, n:

$$h_n = \frac{(a_n - a_L)}{a_L} \cdot \frac{M_n}{M_{\bar{n}}} \tag{3.5}$$

where M_n is the mass of increment n and the average mass of all increments in the lot (for a theoretical analysis, these quantities can be considered known).

The distributional lot heterogeneity, DH_L, can then be defined as the variance of the heterogeneity between all increments (groups), $s^2(h_n)$; there are N_G groups making up the lot.

$$DH_L = s^2(h_n) = \frac{1}{N_G} \sum_n h_n^2 \tag{3.6}$$

In the TOS, DH_L can also be expressed as a derived function of the constitutional heterogeneity (CH_L) in which the spatial, distributional heterogeneity is characterized using a grouping factor Y (for all practical purposes equal to the average number of fragments in each group in the lot) and Z, a segregation factor. Z takes the value 1 for fully segregated material and 0 for completely homogeneous material.

$$DH_L = s^2(h_n) = CH_L \times \frac{1 + YZ}{1 + Y} \tag{3.7}$$

From this formula it transpires that distributional heterogeneity is always larger than zero, and that $CH_L \geq DH_L$, that is, the distributional heterogeneity is always smaller than the constitutional heterogeneity. DH_L is only ever nil, if the lot constitution could indeed have been homogeneous.

With all naturally occurring materials always being heterogeneous, sampling operations will thus inevitably result in an augmented sampling error. The added error (exceeding FSE) associated with the spatial distribution of the groups of fragments within the lot material (increments) is termed the grouping and segregation error (GSE).

Sampling of 0-D lot material will thus be characterized by sampling error variances dependent upon:

- random fluctuations between fragments (CH_L)
- errors generated as a result of the physical sampling extraction of increments. This is tantamount to the sampling process itself producing errors.(as is further discussed below)
- grouping and segregation, i.e. the spatial distribution of heterogeneity (DH_L).

Assuming that sampling is carried out correctly (elimination of all sampling process errors themselves), the 0-D sampling errors associated with the material alone are:

$$CSE = FSE + GSE \tag{3.8}$$

In the process sampling context, this sum is often also termed PIE_1 (see further below).

Heterogeneity in 0-D lots thus covers both small-scale fluctuations within all (potential or realized) increments as well as intermediate and long-range concentration fluctuations spanning the gamut of the entire inner geometry, or volume, of the lot.

In 1-D lots focus is specifically on these long(er)-range variations *along* the dominant space or time dimension. Here GSE is augmented by the possible existence of trends and/or periodicities as specific additional DH_L manifestations. Such lots are either elongated, continuous material streams or nonrandom and discontinuous sample series, such as manufactured units.

The above constitutes the heterogeneity framework for process sampling. Industrial 1-D lots (manufacturing, processing) are generated by chronological operations, for example reflecting raw materials, manufacturing, or processing fluctuations. Besides always including the increment-scale fluctuations (called PIE_1) they will potentially also reflect long-range (trending) (called PIE_2) and/or cyclic fluctuations (called PIE_3), if present. In the process sampling domain, a new, regularized terminology has recently been proposed by Pitard and Esbensen, termed PIE (process increment error, or process integration error)[2].

Thus heterogeneity fluctuations on all relevant scales lead to the following complete array of correct errors for sampling in the general one-dimensional, process scenario:

$$PIE = PIE_1 + PIE_2 + PIE_3 \tag{3.9}$$

The sampling bias is different each time another increment is extracted and it follows no statistical tractable distribution [1–7,14]. It is therefore not possible to use any conventional statistical bias correction rationale [1–7,14]. This issue constitutes the salient distinction between statistics and the TOS. It is an unfortunate myth that 'statistics can resolve sampling issues', no doubt reflecting the possibility that 'sampling' in the strict statistical sense may lead to some form of bias correction. The present chapter makes every effort to distinguish between statistical sampling (from a population of identical, or similar units) and physical sampling (from a heterogeneous lot, in which a similar concept of a set of such units does not exist), for which the TOS must be invoked.

The primary objective regarding sampling of heterogeneous materials and lots is therefore to respect all TOS' specifications needed in order for the sampling process to be unbiased. Failure to do so will by nature generate one or more of a set of three sampling errors specifically related to the sampling process itself. Collectively these are termed the incorrect sampling errors, ISE, which together with FSE and GSE make

[2]This new terminology is proposed jointly in the present publication and in Pitard: 'Pierre Gy's Theory of Sampling and C.O. Ingamell's Poisson Process Approach – Pathways to Representative Sampling and Appropriate Industrial Standards'. Doctoral thesis, Aalborg University, campus Esbjerg (2009). ISBN: 978-87-7606-032-9.

up the basic five 0-D sampling errors. PIE_2 and PIE_3 constitute two additional errors specifically related to process sampling only. All told, a complement of eight sampling errors suffice for all sampling issues to be fully delineated (the occasional increment weighing error (IWE) is commented upon below).

3.2.4 Structurally correct sampling

If handed a sample in the laboratory, it is impossible to tell based on the sample itself if it is representative of the lot from which it was extracted (a true 'sample') or not ('specimen'). Consequently, all efforts should initially be directed to the *sampling process* until it can be unambiguously documented that what results are representative samples only. This dictum applies for both 0-D as well as 1-D sampling; and it applies equally for sensor signals. TOS' main objective is to deliver all the necessary principles behind designing representative sampling processes as well as outlining a foolproof quality assurance system. The sampling process is itself error generating at every sampling stage, including the apparently innocent materials' handling operations such as splitting, sample preparation. A unified approach is presented, in which all sampling errors shall be characterized in an effective, systematic fashion. By focusing on how the eight sampling errors originate and interrelate, it is possible to simplify how to address essentially all types of sampling issues and problems, PAT sampling being no exception.

3.2.5 Incorrect sampling error

The first two errors described above are summarily called the correct sampling errors (CSE) in both 0-D and 1-D sampling contexts.

$$\mathrm{CSE} = \mathrm{FSE} + \mathrm{GSE} \tag{3.10}$$

In contrast, the sampling process itself gives rise to three complementary so-called incorrect sampling errors (ISE), to be delineated in full as these constitute the main culprits with respect to all practical sampling if no action towards their elimination is taken.

Since sampling essentially is a forced statistical selection process (groups of fragments → increments), this must be performed without compromises with regards to lot-sample representativity. Thus far it has been tacitly assumed that it is always possible to follow the basic principle of structurally correct sample extraction to the extreme, that is, all fragments of the lot are submitted to the selection process with an equal probability P of being selected (the *ideal* sampling process). However, for practical reasons, it is necessary to implement every sampling process so that it results in a materialized increment (the *practical* sampling situation) extracted from the lot in a single operation.

This materialization is achieved by first defining the increment to be extracted – an operation termed increment delimitation. After this, the increment must be physically extracted. These two operations are sampling processes each associated with potential errors which are termed the increment delimitation error (IDE) and the increment extraction error (IEE) respectively.

3.2.6 Increment delimitation error

Within process sampling, a correctly delineated increment consists of a cross-section of the stream of matter defined by two parallel boundaries. This ensures that no part of the lot is represented in higher or lower proportions than any other in successive increments, as illustrated in Figure 3.2.

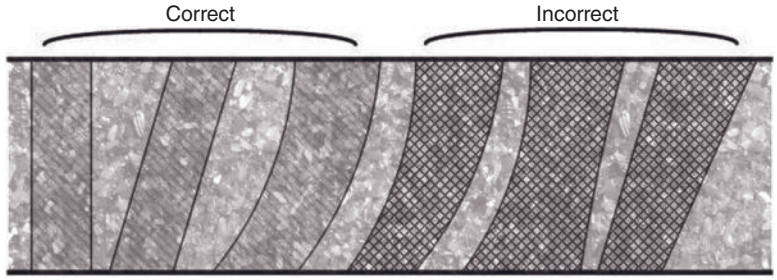


Figure 3.2 *Correct and incorrect increment delimitation in process sampling. Generic illustration (e.g. of a conveyor belt). The three examples to the left are all delineated correctly with parallel sampling implement boundaries, while the remaining three cannot deliver a correct (nonbiased) transverse cut of the 1-D stream. Repeated sampling results in unbalanced transverse proportions of the flux of matter, the effect of which is IDE and IME.*

3.2.7 Increment extraction error

All material located within the correctly delineated increment boundaries must be extracted *in toto*. In order to avoid generating an additional sampling error performing this operation, the IEE, one must ensure that nothing other than what is residing inside the increment is actually extracted. In practice this means that all fragments having their center of gravity inside the delimitated increment must be included in the extraction process and vice versa (Figure 3.2). This principle is known as the center-of-gravity rule (or the rebounding rule for fragment in a falling stream). The actual extraction of an increment from a correctly delimited area is exemplified in Figure 3.3, which presents the general principle. (For full application details, see references 1–7, 9, 14 for a wide range of examples across the entire process industry sector.)

IEE quantifies the errors associated with physical increment extraction from the delineated sampling volume. Any deviation from the center-of-gravity rule will inevitably lead to an extraction error, which can be substantial with many types of particulate materials, or slurries. Obviously, these errors, IDE and IEE, have more influence the more (highly) heterogeneous materials in question.

3.2.8 Increment preparation error

The last ISE error arises if/when the sample is altered after extraction, for instance by absorbing moisture, by spillage, by cross-contamination, or by anything else that inadvertently happens to a sample *after* selection and materialization. The term increment preparation error (IPE) also includes errors induced during operations such as mixing, comminution, and transportation and in addition all handling of increments between the sampling stages as well as all of the processes commonly known in the analytical laboratory as 'sample handling and preparation'. It is better to remove this from the analytical laboratory quality control obligation. To avoid fragmentation, it is preferable to have the complete sampling responsibly in the TOS regimen alone.

The ISE is simply the sum of IDE, IEE and IPE:

$$\mathrm{ISE} = \mathrm{IDE} + \mathrm{IEE} + \mathrm{IPE} \tag{3.11}$$

This triplet gets it collective name (ISE) from the fact that the errors only arise when the sampling process is carried out in an incorrect fashion. The good news is that they can all be virtually eliminated (but only

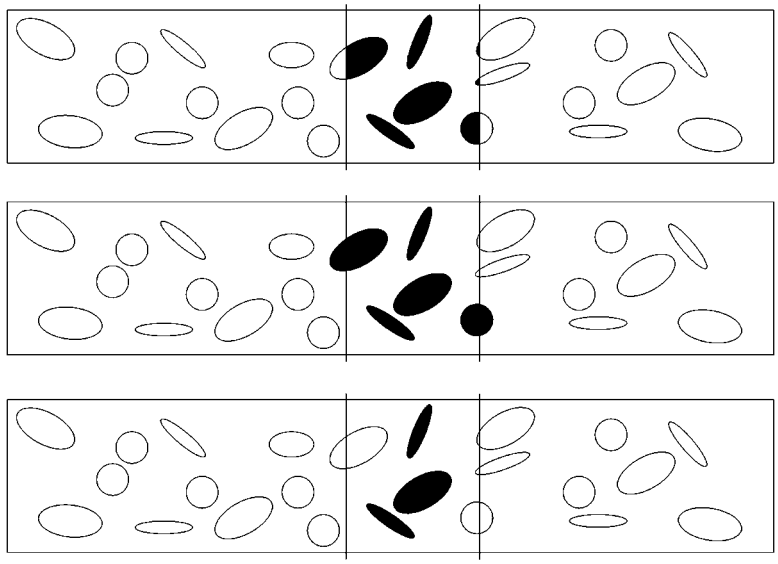


Figure 3.3 *Illustration of correct increment materialization for particulate matter. Open fragments outside the delineated cut should not be part of the increment, while filled-in fragments should be included. In reality, fragments with their center of gravity outside the implement boundaries are not to be extracted – and vice versa. The bottom panel shows two incorrectly excluded fragments. For full details see [1–7,9,14].*

with a dedicated and never inconsequential workload expenditure). By contrast, the correct sampling errors (CSE), a perhaps puzzling name at first sight, are indeed aptly named as these sampling errors occur even in the situation in which all sampling is performed correctly. A correct sampling procedure is consequently defined by the requirement:

$$\mathrm{IDE} = \mathrm{IEE} = \mathrm{IPE} = 0 \tag{3.12}$$

IDE and IEE are random variables, whilst IPE is nonrandom, being generated accidentally and is hence nonstatistical in nature.

3.2.9 Increment weighting error

Whenever increment/sample masses are not identical, estimation of the total lot mass, a_{L} must be done as a *weighted* mean from the analytical results, lest an additional IWE be introduced.

$$a_{\mathrm{L}} = \frac{\sum M_m a_m}{M_{\mathrm{L}}} \tag{3.13}$$

Usually IWE does not pose grave problems, if suitable countermeasures have been contemplated *a priori*. In process sampling obtaining a constant increment mass has many advantages, and is thus often given a high priority. For practical purposes, this can be considered effectuated if the relative variation between increment masses remains below, say, 20% [2,3,9,14]. In order not to confuse the process issues needlessly, below we shall assume that all IWE issues have been dealt with properly (i.e. eliminated).

3.2.10 Total sampling error

Thus the TOS introduces both the correct and incorrect sampling errors, CSE and ISE. At all particular sampling stages, these combine to the total sampling error (TSE):

$$\mathrm{TSE} = \mathrm{CSE} + \mathrm{ISE} \tag{3.14}$$

which (for each sampling stage) can be written out in full (0-D and 1-D):

$$\mathrm{TSE} = \mathrm{FSE} + \mathrm{GSE} + \mathrm{PIE}_2 + \mathrm{PIE}_3 + \mathrm{IDE} + \mathrm{IEE} + \mathrm{IPE} \tag{3.15}$$

PIE_2 and PIE_3 are of concern only for 1-D lots (process sampling), since both the long-range heterogeneity (PIE_2) and the cyclic heterogeneity (PIE_3) have no meaningful counterparts in 0-D lots.

3.2.11 Global estimation error

For the sake of completeness, of course analytical results also include the total analytical error (TAE) in addition to all sampling errors. The sum of these two categories is termed the global estimation error (GEE), which thus includes *all* possible errors which could ever be of interest in PAT:

$$\mathrm{GEE} = \mathrm{TSE} + \mathrm{TAE} \tag{3.16}$$

This chapter will not deal with TAE, as analytical errors are amply dealt with elsewhere in the specific chapters on the individual analytical techniques as well as in the voluminous general pertinent analytical chemistry literature.

Caveat: There is a world of difference in having the full complement of all eight TOS sampling errors with which to analyze, characterize and improve sampling processes and the all-under-one-hat empirical approach termed 'measurement uncertainty', which can only demonstrate the magnitude of sampling error influences. By itself this remains a futile exercise; one still needs the conceptual and analytical understandings from TOS in order to be able to reduce the unacceptably large total errors.

3.3 Mass Reduction as a Specific Sampling Procedure

The focus of process sampling very often is on selecting the final, conveniently small, analytical sample mass as early as possible in the sampling process, preferentially directly as part of the primary sampling.

Contrary to this easy way out (essentially grab sampling, see Figure 3.9 and below) it is infinitely more in compliance with the goal of process monitoring to focus on *how* to make the primary sample be representative – even if this means procuring an apparently 'oversize' sample. This dictum necessitates that primary sample masses subsequently can always be reduced to the desired level commensurate with analysis. This section briefs on the principles and equipment types for representative mass reduction. Indeed this is always possible, although practically never simultaneously with taking the primary sample.

Any mass reduction method must be in compliance with the fundamental sampling principle, that all parts of the lot (or subsample) have equal probability of ending up in the final sample (or subsample) for analysis.

An extensive benchmark study has been carried out in order to identify the optimal mass reduction principle(s) [6]. This was achieved by assessing and validating the almost universally used mass reduction approach, grab sampling, as opposed to a comprehensive series of more advanced techniques and methods, 17 in total. All techniques and methods were tested with regard to a full suite of quality parameters, some scientifically important, others related to cost and minimizing practical operating conditions and expenses. The most important merit of any mass reduction method is the ability to deliver an *unbiased* split of material with the smallest possible variation in repeated runs subject to the best possible accuracy. These two features are summed up in the TOS' quantitative measure of *representativeness*.

The principal types of mass-reduction techniques and methods investigated covered:

- grab sampling (a single increment is used as *the* sample).
- fractional and alternate shoveling (dividing sequentially into two (alternate) or more batches, scoop by scoop).
- riffle splitters, with varying number of chutes, chute width, feeding mechanism, dust minimizing systems.
- rotational dividers in which samples are distributed over radial chutes by a rotating feeding nozzle; rotational splitters are also based on the riffle-splitting principles; rotational splitters are by far the most often used with on-line process sampling mass-reduction equipment.

Petersen *et al*. [6] give a complete evaluation of all approaches and instrument types on the market, which resulted in an overall ranking. The conclusions and recommendations are clear and unambiguous: any method involving manual shoveling, grabbing or similar simplistic selection of some material is by far the worst, in fact totally unacceptable. This includes grab sampling, fractional shoveling, alternate shoveling and all variants hereof. *Only* riffle splitters are optimal under all conditions or methods based on the same underlying principle, namely that all material is distributed uniformly over an equal number of equally sized chutes. This design facilitates a representative split into well-specified portions of the original material (lot). Standard riffle splitters end up with two identical 50% portions, necessitating a stepwise iterative procedure to achieve larger reduction ratios. This is easily evaded using the slightly more technical rotating dividers ('Vario dividers'), where the total sample mass can be reduced to a smaller proportion sample and larger proportion (waste or bypass), as all chutes are used many times over. This principle is identical to using a static riffle splitter in sequence, only many more times over, achieving a very high number of effective chutes in operation. The riffle-splitting principle is illustrated in Figure 3.4.

It is crucial that all mass-reduction equipment is operated correctly. There are a number of very practical rules to be followed when using any of the acceptable devices. Neglecting one or more of these will cause the entire mass reduction to be biased, and result in unnecessarily large variation. Although Petersen and colleagues [6] on the surface primarily deal with stand-alone, laboratory bench type equipment, the principles derived must apply equally well for any type of on-/at- or in-line mass reduction that can be contemplated in the PAT context.

A salient example from the process sampling realm is therefore given here. In the routine PAT laboratory it was desired to perform a mass reduction as part of the process of transporting a powder material. A bypass stream was established and on this the device depicted in Figure 3.5 is supposed to deliver: one 10% subsample, another 25% subsample, with the remaining 65% ('reject'), fed back to the main stream.

Unfortunately the design of this 'splitter' flies against all TOS' mass-reduction principles. This device is unable to counteract radial as well as peripheral heterogeneity in the cross section of the falling stream of matter. (No attempt was made to homogenize the stream before this splitter device was in place either.) This device is designed on the wrong *assumption* that the cross-section of the falling stream of matter is homogeneous at all times. Because this manifestly cannot be the case, potentially severe IDE is inherently

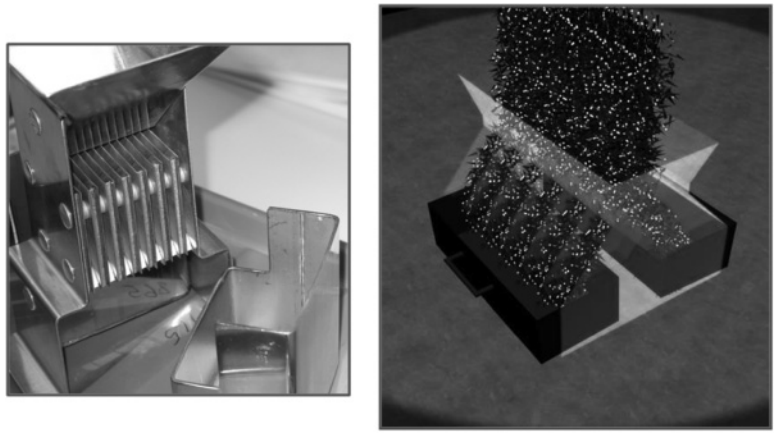


Figure 3.4 *Illustrations of the principle of correct (unbiased) riffle splitting. Riffle splitting (principle and equipment) constitutes the only representative mass-reduction procedure complying with the TOS. See [6] for a comprehensive principles and practice guide.*

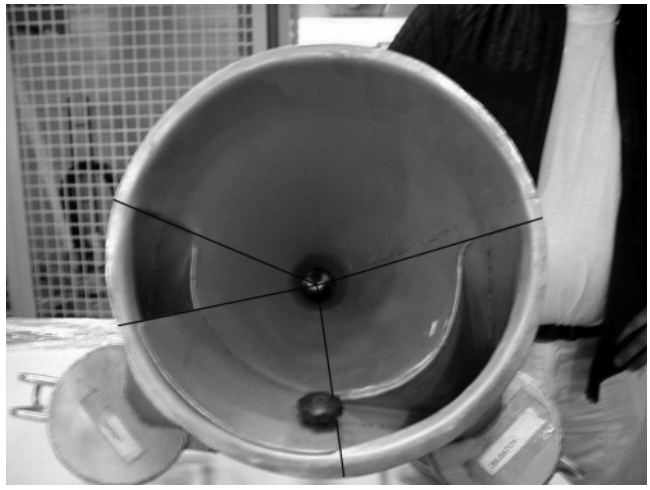


Figure 3.5 *Faulty designed device for 'splitting' a falling stream of powder material into a 10% subsample (left), a 25% subsample (right) and a 'reject' (supposed to cover 65% in the central flow-through channel). The design shown is structurally flawed for several reasons: only the material hitting the outer peripheral annulus is split into the desired 10/25/65% proportions while the central part of the falling stream is selected as reject in its entirety, disqualifying the 'splitter' from ever producing a representative split. See also Figure 3.6.*

generated. Indeed in a particular production setting this faulty in-line subsampler *perforce* must result in the product being 'out of statistical production control', which is quite wrong however – there is ever only a sampling issue here.

Alternatively, a correctly designed, maintained and operated rotational splitter will solve all the IDE weaknesses involved, as illustrated in Figure 3.6.

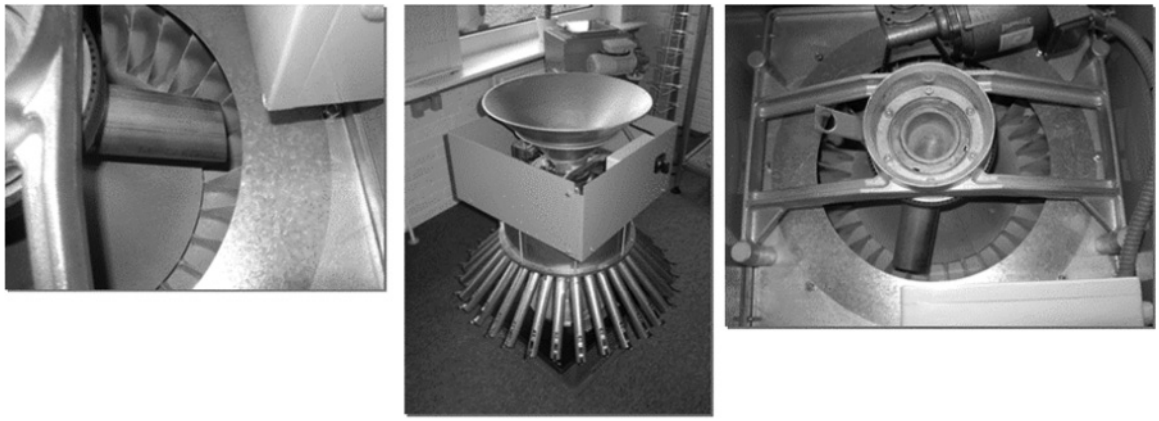


Figure 3.6 *Rotational splitter. The falling stream impacts on the hopper (center panel) which guides the flux through the rotating tube (left panel) meeting, in this case, 32 chutes per rotation. The devise is engineered to be able to maintain a constant rotational velocity irrespective of the load. This feature allows for a vastly extended number of effective chutes by problem-dependent hopper flux versus rotational velocity coordination. Illustration of a so-called 'vario divider' courtesy 'Rationel Kornservice' [15].*

3.4 Fundamental Sampling Principle

TSE has contributions from primary sampling as well as from all subsequent mass-reduction steps, i.e. secondary, tertiary sampling stages. While much of the sampling effort is (or at least can be) under control of the sampler, the part from constitutional heterogeneity is dependent on the material properties only. The FSE cannot be altered for any given system (lot, geometry, material, state, size distribution). Altering FSE necessitates physical intervention such as crushing or comminution, which are generally not options available in the on-line, real-time PAT context. Error contributions related to the spatial distribution of the material are dependent on the specific sampling procedure employed and whether correct counteractive measures against the heterogeneity are invoked (e.g. mixing, composite sampling). There are many interesting possibilities for 'smart' sampling and/or sensor equipment, many of which can be correct. But this domain is also ripe with many structurally flawed solutions which are incapable of delivering a representative increment, 'elegant' process technology engineering notwithstanding.

The imperative is always to respect that all extractions (all possible virtual increments) must have the same selection probability. This is called the fundamental sampling principle (FSP), which must *never* be compromised lest all possibilities of documenting accuracy (unbiasedness) of the sampling process are lost. FSP implies potential physical access to all geometrical units of the lot, which is easily seen as much easier to achieve in the process situation than in batch sampling cases. The TOS contains many practical guidelines of how to achieve compliance with FSP [1–14], as praxis-codified in the seven sampling unit operations outlined below.

3.5 Sampling – a Very Practical Issue

ISEs are the root cause of *all* sampling bias. Elimination of ISE is therefore of the highest priority for all practical sampling. For process sampling the most frequent problems are associated with IDE. After

elimination of ISE, the priority is achieving sufficient reduction of the sum of all CSE to a predetermined level. All of this can only be accomplished by observing TOS' guidelines for representative sampling. There are many practical, mechanical aspects of this issue, all relatively simple, almost all trivial to implement in isolation. The result can only be achieved however if *all* are properly recognized and acted upon (see Figures 3.11–3.20). The TOS literature deals at great length with all these issues, with a focus on explaining the connection between the individual errors and how they can be minimized partially or completely in sampling practice. The selected references below are comprehensive and sufficient for process sampling initiation [1–14,16]. In-depth background references can be found in abundance herein.

3.5.1 Sampling unit operations

A set of sampling unit operations (SUOs) has been formulated recently to constitute the minimum set of principles sufficient for all representative sampling [1,12–14]. From these one can derive specific procedures regarding *practical* sampling, for batch sampling (0-D) as well as process sampling (1-D), although the specific implementation scenarios differ [13,17]. These SUOs may perhaps be viewed analogously to Maxwell's four governing equations covering all of electromagnetism. Both sets of high-level principles must be applied specifically in the relevant cases and instances.

The seven unit operations can be grouped according to their nature and use as follows:

- *Three general, overarching principles* – normally utilized only in planning or optimization of sampling procedures/campaigns:
  - transformation of lot dimensionality - transforming 'difficult to sample' 2-D and 3-D lots to 'easy to sample' 1-D lots (sometimes 0-D transformed into 1-D lots as well)
  - characterization of 0-D sampling variation ('replication experiments')
  - characterization of 1-D (process) variation by variography.
- *Four practical procedures* – often used several times over, in various problem-dependent sequences and combinations during practical sampling:
  - lot, or sample, 'homogenization' by mixing or blending
  - composite sampling (striving for as many and as small as possible increments, subject to the largest total sample mass which is practically possible, assuring the highest possible volumetric coverage of the entire lot)
  - particle size reduction (comminution)
  - representative mass reduction.

For the present introduction to process sampling, the central SUOs will be illustrated and commented upon where appropriate. The specific aspects of the TOS from which the individual SUOs are derived are explained in the literature to any depth desired, complete with many application examples [1–7,12–14].

In the following, the focus is on how these SUOs can be applied to address process sampling in general, and specifically to PAT sampling issues.

3.5.2 Understanding process sampling: 0-D versus 1-D lots

A comparison of the specific setup of correct and incorrect sampling errors for 0-D versus 1-D sampling respectively is used to illustrate the specific process sampling issues. A schematic synopsis of the commonalities and differences involved is given in Figures 3.7 and 3.8.

For the case of process sampling (1-D lots) the corresponding setup is depicted in Figure 3.8.

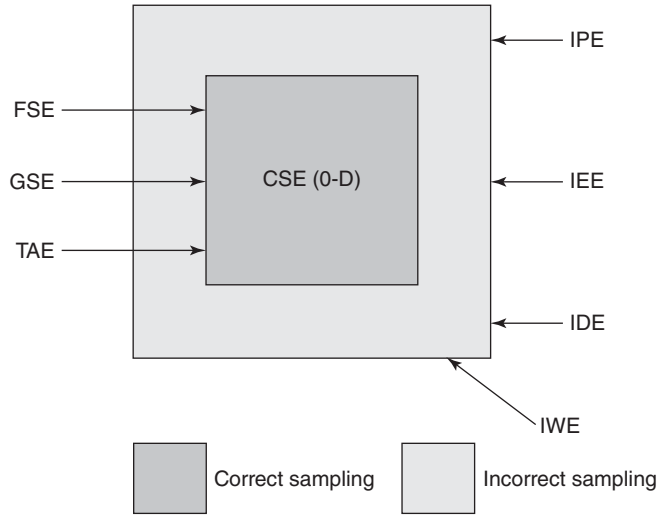


Figure 3.7 *Sampling error relationships for the case of batch sampling (0-D). A core of correct sampling errors (CSE) is embedded in a cloak of incorrect sampling errors (ISE). Representative sampling must start by eliminating all ISE.*

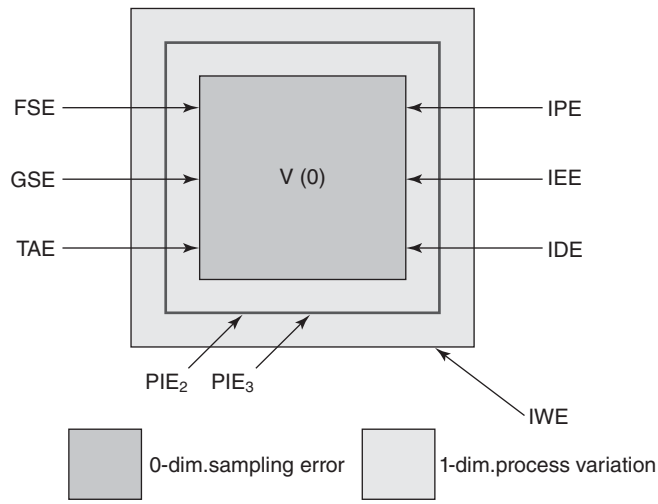


Figure 3.8 *Sampling error relationships for the case of 1-D lots (process sampling) in which the core consists of all 0-D sampling errors (CSE + ICS + TAE), in variography termed* V*(0), i.e. the zero-lag sampling variance (see text). Two additional process sampling errors* PIE_2 *and* PIE_3 *add to the total error variance if not properly eliminated/reduced. The objective of representative process sampling is to reduce* V*(0) maximally, to avoid* PIE_2 *and* PIE_3*, opening up for optimal fidelity in delineating process variability for process monitoring.*

Figure 3.9 *Grab sampling in different practical manifestations (aggregate material), to which must be added all PAT sensor 'alternatives' in Figures 3.12–3.14 dealing with fluids and slurries. All principal characteristics are identical for physical sampling and signal acquisition: there is only sampling from 'some part of the stream cross section'. This creates IDE and therefore cannot result in representative sampling.*

3.5.3 Grab sampling – 0-D and 1-D

Grab sampling is a fatal option in all sampling. Grab sampling has only one apparent virtue: the work expenditure is guaranteed to be minimal. But everything else regarding grab sampling is wrong; it is incorrect and always biased. Grab sampling can never produce a representative sample. Suffice here to refer to the sampling literature for complete documentation [1–14].

> The accuracy of many analytical data reports is a mirage because unwitting negligence and false cost consciousness have ensured that a 'sample' … taken with cursory swiftness has been examined with costly precision.' Kaye, Illinois Institute of Technology (1967)[3].

[3] Quoted in [2].

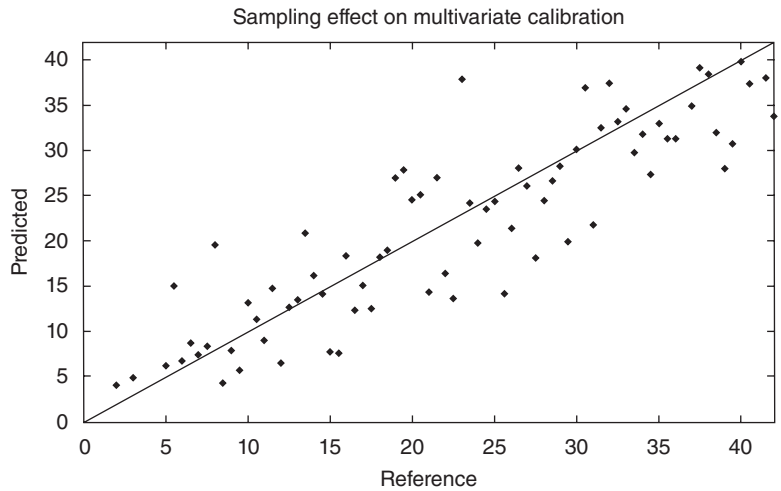


Figure 3.10 *Hallmark signature of significant sampling bias as revealed in chemometric multivariate calibrations (shown here as a prediction validation). Grab sampling results in an unacceptably high, irreducible RMSEP. While traditionally ascribed to 'measurement errors', it is overwhelmingly due to ISE.*

Most process sampling is in fact nothing but blatant grab sampling in the specific 1-D context (Figures 3.9 and 3.10); see also [18] for illustration of the predominant analytical view, which often blurs this understanding.

A chemometric aside: The result of grab sampling, or any other ISE-dominated '*specimenting*', in the context of multivariate calibration is shown in Figure 3.10. The depicted situation, in which RMSEP (root mean square of error of prediction) cannot be brought under sufficient control, can be taken as a hallmark signature of the effects of a sampling process [*X*,*Y*] which includes a significant bias. Contrary to much current chemometric mythology, this is not due to measurement errors alone. Applied TOS contains many examples in which the depicted deadlock was broken only by proper elimination of ISE.

An often heard claim is that this RMSEP is due to measurement errors without specifying whether this is related to *X* (signals) or to *Y* (analytical errors), or both. In chemometric/PAT literature one can find many issues which draw attention away from the grab sampling issue, for example, recurrent discussions on *X*-error assumptions, – distributions, propagations and such like – which at best can only explain a (minor) part of the *X* issue. However, the precision of the analytical equipment used in modern PAT is already of the highest quality, that is, the *X*-signal errors (reproducibility) are already minimized from the outset. High *X*-signal variance (*X* error) much most likely stems from the flux of material having significant heterogeneity at the scale level of the probe head cross section. PAT's stock solution here is to make use of a (very) high number of scans to be averaged (*X*) as modern technology easily allows for hundreds, maybe even thousands of scans, or averaging of consecutive measurements. But PAT probes and sensors are not a panacea, relieving one of the responsibilities with respect to the issue of localization and deployment representativity. These problems are delineated in full measure below. The correct 1-D increment delineation is always, without exception, a complete cross-section of the material flux or stream (Figure 3.2). A high number of scans in the same nonrepresentative deployment location will not eliminate the fact that the probe sensing area only records a multiscan average of the material that happens to stream by the probe's narrow field of view. The requirement for a simple probe head averaging to be representative would be a time-variable demonstration that the entire cross section of the material flux involved is in a state of cross-sectional DH_L homogeneity. This has never been demonstrated for any sensor/probe, but see Figure 3.21 for a close try.

The conclusion is inescapable: sensor/probe heads perform in a manner identical to grab sampling. There is little sense in relying on physical specimens only, or specimen-signals, to characterize the entire cross stream or product in the PAT context. Figure 3.16 is *the* case in point.

3.5.4 Correct process sampling: increment delimitation/extraction

Figures 3.11–3.20 are generic illustrations (representing both conveyor belts and pipelines) used to showcase the specific SUOs needed in the PAT sampling context.

Figure 3.11 illustrates that correct process sampling cannot be realized based on taking only *some* of the stream (cross-section) *some* of the time (top panel), nor by taking *some* of the stream (cross-section) all of the time (center panel). Sampling a correctly delineated and fully materialized increment at regular intervals presents the only option for ensuring representative samples (see Figure 3.2).

Taking these issues into the process sampling domain, Figures 3.12–3.14 show some of the many ways in which 'sampling valves' are designed and implemented in current process technology. A fair estimate

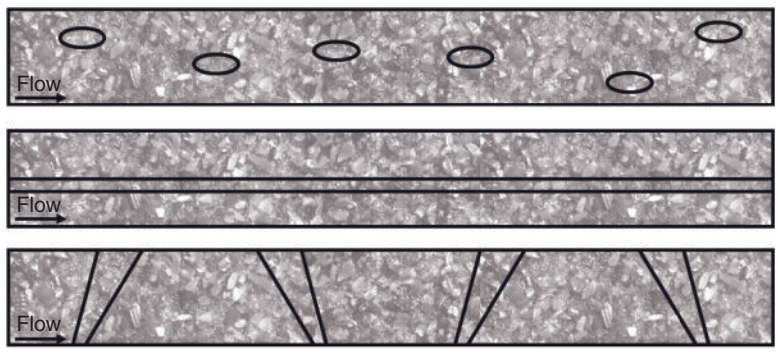


Figure 3.11 *Illustration of incorrect increment delineation leading to IDE (always generating a sampling bias) in 1-D sampling from conveyor belts or pipelines. Top panel illustrates simplistic grab-sampling; center panel shows the situation using a sampling valve with limited cross-sectional coverage (Figures 3.12–3.14); lower panel illustrates time-varying IDE associated with variable cross-stream cutting; compare Figure 3.2.*

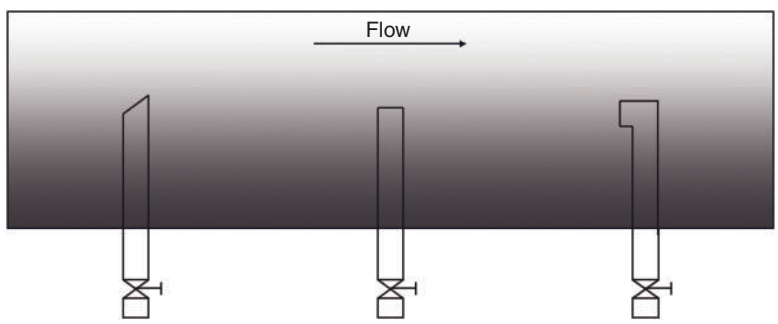


Figure 3.12 *Generic illustration of incorrect valve designs for the case of horizontal flow. In these examples only part of the cross section of the stream is sampled all of the time; intermittent sampling/signal acquisition is in no way better, see Figure 3.11. The sampling valves shown there offer no counteraction for gravitational segregation and/or flow differentiation effects. All 'solutions' shown are structurally incorrect. A specific valve opening design alone does not eliminate IDE.*

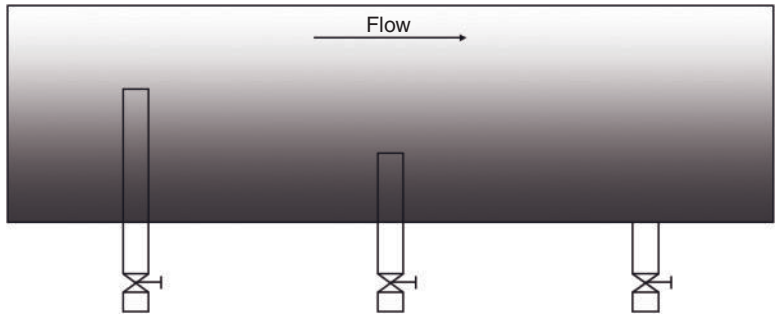


Figure 3.13 *Generic illustration of incorrect valve deployment for the case of horizontal flow. In these examples only part of the cross section of the stream is sampled. The sampling valves shown offer no counteraction with respect to gravitational segregation and/or flow differentiation effects. All 'solutions' shown are structurally incorrect. A specific valve opening that does not cover the entire flow cross-section does not eliminate IDE.*

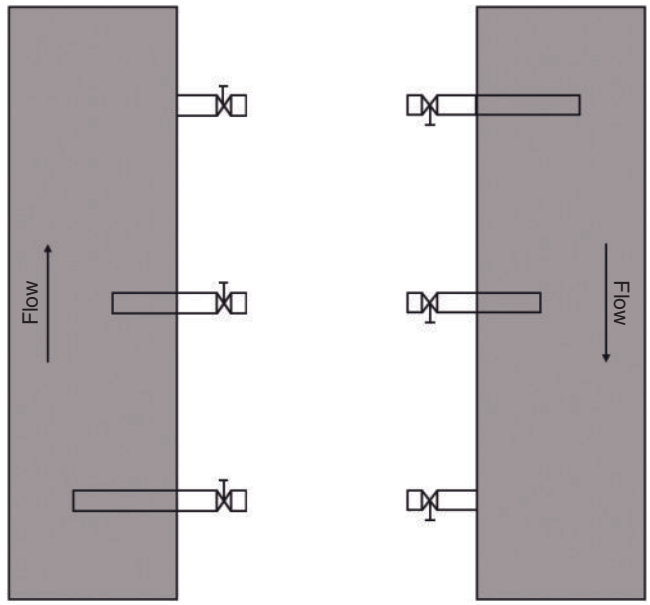


Figure 3.14 *Sampling valves (in reality 'specimenting valves') deployed in vertical flow settings: upward flow (left), downward flow (right). Similar designs as in Figure 3.13 result in flawed, incorrect functionality due to IDE, leading to extraction of nonrepresentative specimens. The alternative PAT situation employing sensors instead of valves, but for horizontal flow, suffers from identical incorrect IDE problems, see Figure 3.16.*

would be that these examples cover more than 90% of conventional sampling instrumentation. Simple TOS analysis allows to state that in reality every 'solution' shown is faulty, failing most blatantly with respect to incorrect increment delineation, for which reason they must *perforce* lead to nonrepresentative *specimens* only. IDEs (sometimes IEE as well) abound, thus leading to an unacceptable sampling bias.

The very same faults are associated with flow in vertical pipeline segments, if based on the same valve designs.

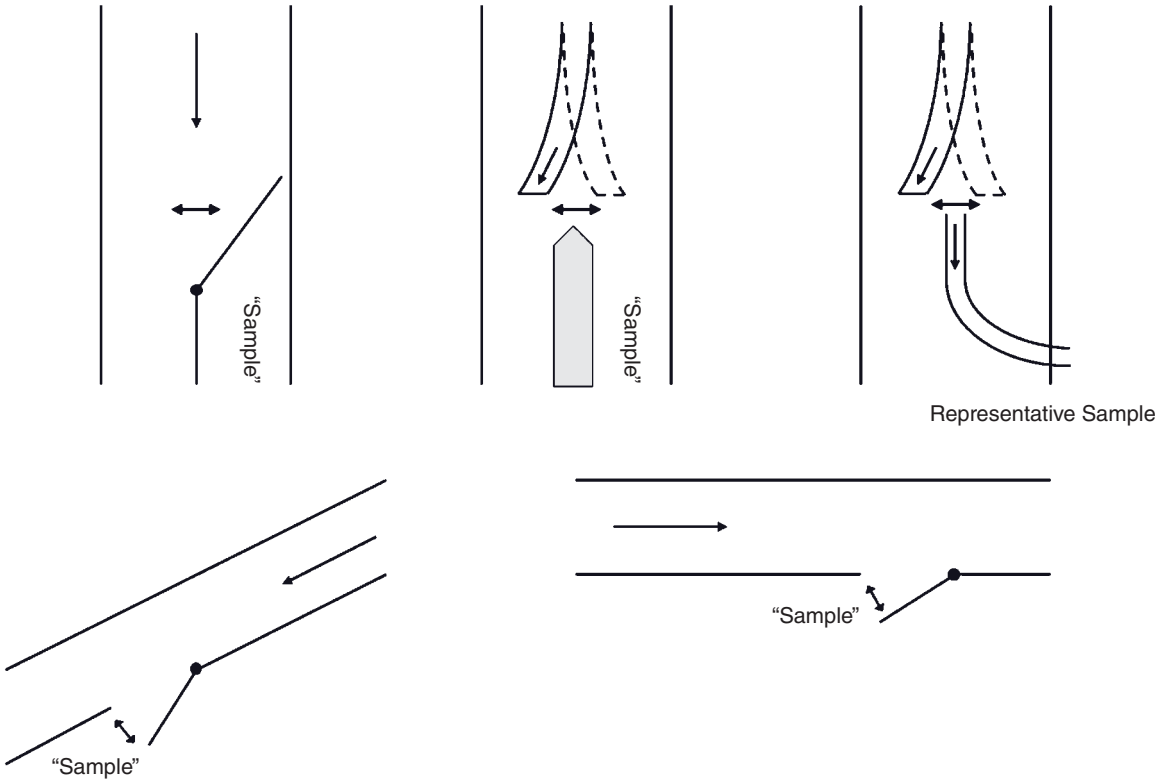


Figure 3.15 *Illustrations of popular 'trapdoor' sampling valve configuration in relation to vertical and horizontal/inclined stream fluxes. The central top panel design purports to siphon off a 'sample' in the right downpipe at intermittent intervals. Detailed TOS analysis leads to the conclusion that only the top right design eliminates IDE. All other implementations are structurally incorrect, leading to nonrepresentative specimens only and must therefore be rejected [1-4,12-14].*

Current process technology is full of similar 'solutions' as shown above. A much used vertical deployment scenario is sampling from a falling stream, in which very many solutions on close inspection also fall apart for the exact same IDE reasons. Thus Figure 3.15 illustrates various implementations of the popular 'trapdoor' sampling valve, four out of five fatally characterized by significant imbalances with respect to the peripheral stream fluxes in relation to the opening/closing of the trapdoor mechanism. Only the top right design allows for correct increment delineation (elimination of IDE) [1–4,12–14].

Current PAT approaches depend heavily on in-line/on-line/remote sensing sensors and probes 'instead of having to extract actual samples from the process'. A fair summary of the most popular PAT sensor/probe designs and deployment strategies in a pipeline is illustrated in Figure 3.16 (and similarly for vertical flows).

From Figure 3.16 the message is clear: the *smaller* the inner piping diameter, the better possibilities for PAT sensors/probes to be able to minimize (*not* eliminate) IDE and consequently also IEE.

3.5.5 PAT versus correct process sampling – what is required?

Close inspection reveals that most probe solutions are fraught with problems identical to those pertaining to the above sampling valve issues (IDE and IEE). In fact, the effective analytical volume in front of the

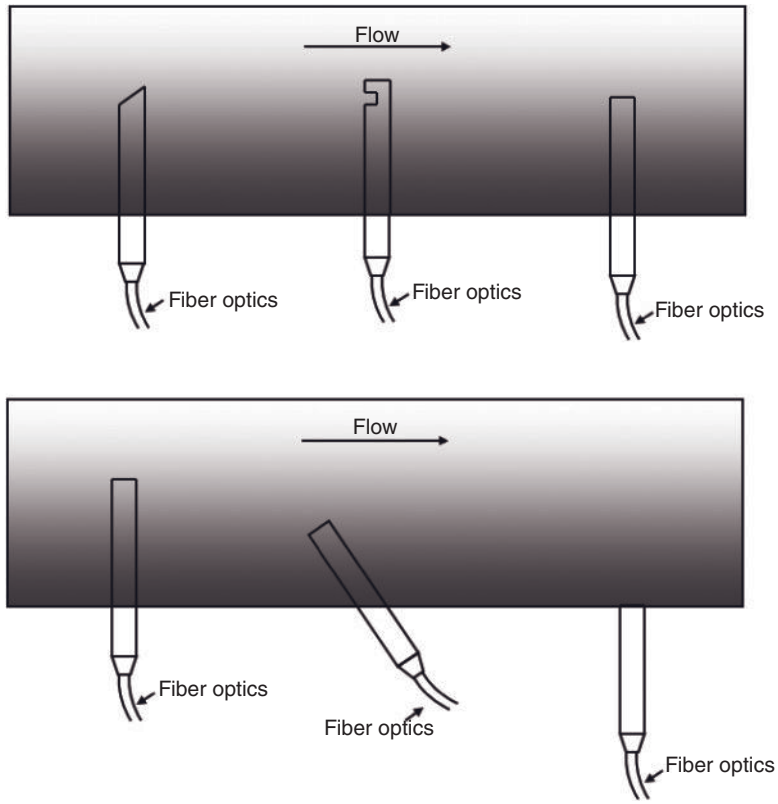


Figure 3.16 *Illustration of incorrect sensor and probe head design (top), and deployment strategies (bottom). There are numerous variations on the same theme(s), e.g. in vertical flows, but all alternatives need not be shown as the design/deployment problems are identical: none are correct – as none eliminates IDE. Indeed none have a field of view larger than a miniscule part of the cross section of the flow. Identical generation of IDE holds also for vertical flow, compare Figure 3.14. Note that real-world conduit diameters are often significantly larger than shown here, which only exacerbates these IDE issues.*

specific sensor or probe head is often significantly smaller than the physical increment volumes extracted by an alternative sampling valve.

Detailed analysis of the specific sampling situations – physical extraction or sensor/probe signal acquisition – reveals the essence of both processes. Noncompliance with TOS' principles for correct sampling leads to unavoidable IDE/IEE with the result that a significant sampling bias is produced by the very sampling process which was supposed to deliver high quality PAT reference data. All multivariate calibrations are critically dependent upon the quality of the training set reference values (Y), which must have originated, one way or another, by sampling. The initial PAT euphoria has come full circle: one is back to square one. How does one procure representative samples *and* representative sensor signals? How does one approach the often glaring mismatch between these different volume supports?

Comprehension of eight sampling errors and the simple set of seven SUOs will allow all process sampling situations to be properly analysed and appropriate solutions (principles, design, equipment, implementation, operation, quality control) to be derived.

When there is a significant IDE present, chances are high that this will produce IEE as well. (In the rare cases this is not so, IDE alone creates a sampling bias.) IEE in the process sampling domain would e.g. be

cross contamination between extracted samples (or specimens), lack of flushing of dead-stocks, clogging effects etc. For PAT, sensor/probe solutions IEE takes the form of for example, clogging (caused by the sensor's obstruction of the flow), fouling or coating of the optical system, etc. Even though these are often interrelated to a high degree, in principle these two sampling errors should always be addressed, and appropriate solutions found, individually [3]. This also applies to IPE, which is often more obvious to identify and easier to correct for [1–4,14].

3.6 Reactors and Vessels – Identical Process Sampling Issues

The process sampling situations delineated above are not principally different regarding sampling from processing reactors and vessels, etc. Figure 3.17 shows the situation regarding sampling in a typical process reactor.

In a typical reactor context, sampling stations may be located at the reactor inlet, or outlet, or it may be desired to sample directly from the reactor volume. Shown here are two valves at two different sampling heights. From TOS' principles it is clear that all sampling solutions shown are incorrect, leading to IDE. It is unfortunately often unrealistic to simply *assume* that conventional stirring will always result in *homogeneous* reactor content. Faith, instead of verification, is all too often put into one form or another of the mixing system as per OEM declarations. Stirring impellers are the most often used solution for reactor mixing. Rarely, if ever, are systematic validations carried out to verify the actual degree of 'homogeneity' attained. These would show that mixing in general is an overrated process. This issue is to a large degree correlated with the specific rheological characteristics of the lot material. In vertical process reactors significant 'dead zones' (shadow zones) are bound to develop, effectively outside reach of the mixing system. For obvious

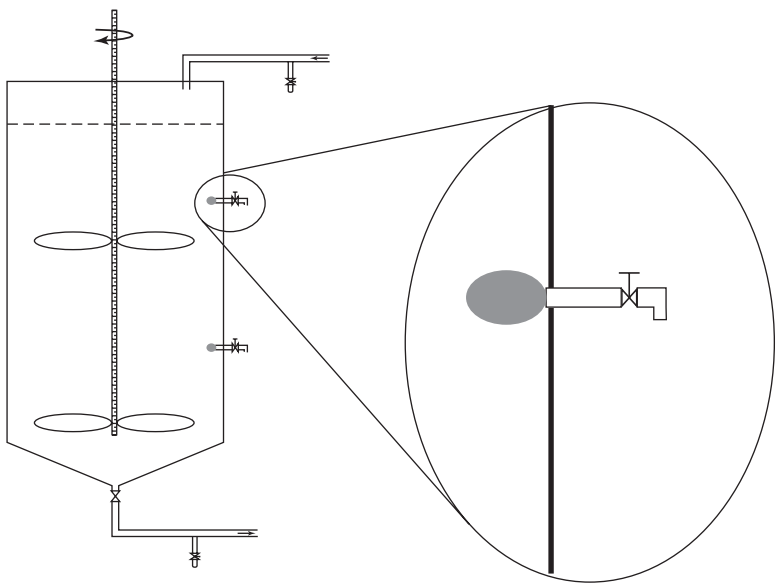


Figure 3.17 *Generic process reactor equipped with conventional sampling valves shown at typical deployment localizations e.g. at reactor inlet, outlet or at different reactor heights. TOS analysis again results in rejection of all sampling options shown here as no representativity with respect to the entire reactor volume can ever be achieved.*

reasons the most important shadow occupies the bottom zone of the reactor. Depending on the details of the design, there may be others. From a strict TOS point of view, sampling of process reactors constitutes an impossible 3-D sampling problem – serious and well-meaning efforts of efficient mixing notwithstanding.

The situation is exactly the same in trying to remedy these problems using alternative PAT sensor technologies. Employing sensors or probes only shifts the issues from physical sampling representativity to identical sensor signal issues. Both approaches are fraught with unsolvable IDE (as well as potential IEE) – Figure 3.18. The result in both cases is that nonrepresentative specimens or *ditto* signals are acquired because of neglect of the critical issues regarding heterogeneity versusIDE/IEE in the process technological design phase (reactor/vessel/piping design versus sampler/sensor deployment).

A classic mistake is well illustrated by the case of building a process analytical model, for example in the form of a multivariate calibration, employing only secondary samples (*X*: spectroscopy; *Y*: laboratory reference data) from the *same* primary sample(s) for the calibration. The likelihood of success is significantly increased because the primary samples are split into *X* and *Y* subsamples which are the only ones appearing in the chemometric model. This will succeed, or not, simply based on whether representative mass-reduction problems are involved and solved, or not. The salient sampling bias, that is, the inaccuracy of the primary samples with respect to the lot, has apparently vanished (at least in the minds of the ill-reflected PAT modeler), but this is a mirage. While the model may well be acceptably validated according to all the traditions in chemometrics, it will not perform as well (it may even be totally useless) when supplied with new primary samples, as their primary sample bias was never incorporated into the modeling. The model will be dysfunctional depending on the magnitude of the left-out bias. References [20–24] discuss these issues in more depth from different perspectives, all related to process sampling.

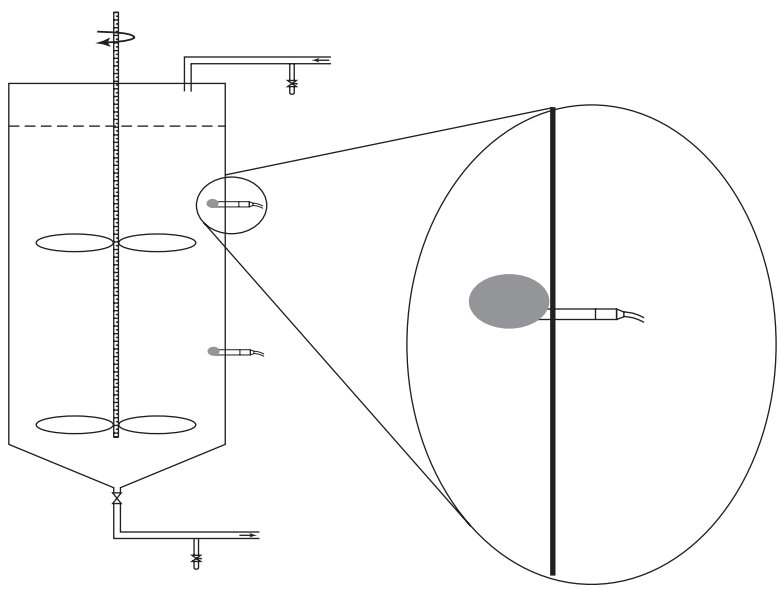


Figure 3.18 *Generic process reactor alternatively equipped with PAT sensors or probes, shown at typical deployment locations at different reactor heights. TOS analysis again results in rejection of all options presented. The critique leveled against incorrect PAT sensor deployment in ducted flow (Figure 3.16) applies in full force for reactor sampling as well.*

3.6.1 Correct process sampling with existing process technology

The general solution to all of problems illustrated above is to apply the most important SUO in process sampling: lot transformation, in this case from 3-D to 1-D. It would be extremely valuable if it were possible to solve (all) the above weaknesses based only on existing process technological elements; and this is actually the case to a significant degree. Below follows a description of the only general TOS-acceptable process sampling solution: *upward flux sampling*.

PAT solutions are to a large degree based upon first successes mainly involving liquid solutions or uniform materials (often powders) with only insignificant diversions from Beer's law characteristics and fair to excellent transmittance/absorbance characteristics for electromagnetic interaction, for example allowing comparatively easy NIR quantification. First generation PAC applications of infrared spectroscopy as well as other analytical techniques (by no means not all spectroscopic), are well covered by McClellan and Kowalski [18]. As the PAC approach developed over time, significantly augmented by the US FDA PAT initiative, applications have steadily broadened with the result that PAT today faces a range of nonideal spectroscopic issues often up against a high degree of material variability. The scientifically prudent and consistent approach is to treat all process sampling issues as if the heterogeneity involved is significant, that is, that there is both cross-sectional as well as longitudinal (process) heterogeneity, and to deal with these issue strictly following only TOS' principles. This general attitude is *inclusive* and will work for all possible cases; while eventual minor cases may crop up, no generalization must ever be based hereupon.

Since the TOS' sampling principles for 1-D lots specifically are designed to cope with *all* types of process sampling/materials issues, a unifying approach is recommended as described in Section 3.6.2. The operative imperative is that all increments *must be* in the form of complete, edge-parallel cross-stream segments in order to uphold the fundamental sampling principle. For today's many cases of IDE-flawed pipeline and reactor sampling (or the equivalent sensor deployments) (Figures 3.11–3.18), there actually is a remarkably easy way out. Coincidentally, the prerequisites for its success are for the most part already in place in many of the current process technological plant designs.

3.6.2 Upward flux – representative colocated PAT sampling

Consider first downward flow in a pipeline where significant heterogeneity must definitely be assumed. In fact, this flow regimen results in the most accentuated differential radial and longitudinal flow segregation possible. These effects will be progressively more adverse with increasing material heterogeneity; a strongly irregular flow will quickly develop, bordering on the chaotic. Downward internal ducted flow is by far the worst sampling scenario that can be encountered, but horizontal flow unfortunately does not involve any principal difference with respect to the critical IDE and IEE issues. In general, all sampling from horizontal and downward directional flows cannot ever meet the TOS criteria.

Consider now the only remaining flow situation – that of an upwardly flowing stream. Because of the constantly opposing force of gravity, the radial expressions of a differential velocity field will be counteracted to a higher degree. Turbulence will also be dampened, and will develop at higher velocities relative to downward flow, but there will of course still be a significant turbulent flow if the velocity is not very low (laminar flow). Because of these characteristics, also helped by wall friction effects, upward-flowing stream fluxes tend to be maximally self-mixing. Consequently, in upward-flowing pipeline streams the most effectively mixed fluxes of matter are found progressively further up along straight vertical pipeline segments. From conventional hydrodynamics of internally ducted flow theory (and practice), the dictum is that it takes a transportation length corresponding to some 40–60 diameters from a turbulence-generating locus (pump, bend, confluence) to develop the most stable flow [25].

From a practical process sampling point of view, this means that the longest possible upward pipeline segment before sampling leads to optimal ISE reduction. For obvious reasons, extracting the sample cannot make use of a trapdoor mechanism or similar. The simplest solution is that depicted in Figure 3.19, which makes use of two co-operated butterfly valves regulating the flow in the main pipeline and the by-pass/extractor respectively. When directed through the bypass, the increment materialized will precisely constitute the stipulated TOS-correct cross-sectional segment of the flow.

The same geometrical IDE solution can be effectuated by several alternative but more complex valve designs. A recent analysis of the issues of severely incorrect sampling valve design and effects for horizontal flow can be found in [20].

Figure 3.19 shows an archetype 1-D stream sampling solution for ducted flow following the TOS principles completely – full elimination of IDE and IEE by use of the upward-flux pipeline sampler. Valve rotations are *prograde* only, taking care of subtle IDE issues otherwise similar to trapdoor flaws. Flushing of the dead stock(s) involved is not a particularly serious issue if/when appropriately recognized (always problem-dependent). Both these features help eliminate IEE significantly; full elimination is always possible.

For the case of 'impossible' 3-D reactor or vessel sampling, this upward-flowing pipeline sampler should be implemented with the reactor in a recirculation loop configuration, as shown in Figure 3.20. The only new requirement is a low-power pump, to be operated full-time for all systems where this is feasible.

This configuration has the important added benefit of ensuring significantly improved mixing of the reactor contents. This feature cannot be over-emphasized. While the upward flow sampler originally was designed

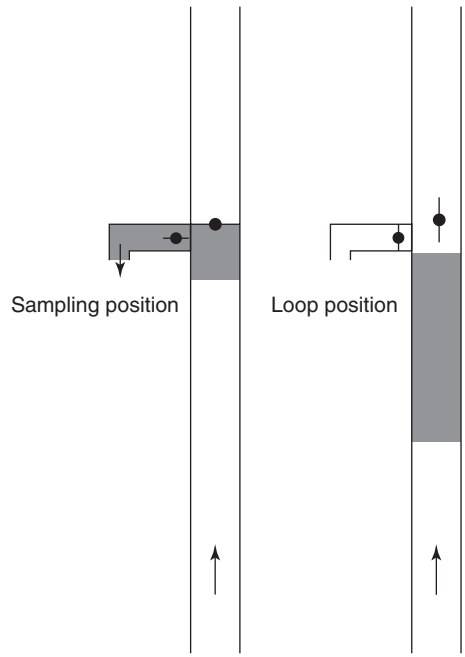


Figure 3.19 *Generic ducted flow sampling solution completely respecting all TOS principles (elimination of IDE and IEE). It is a prerequisite that appropriate flushing/cleaning measures (always problem-dependent) eliminate potential IPE in the bypass extractor segment.*

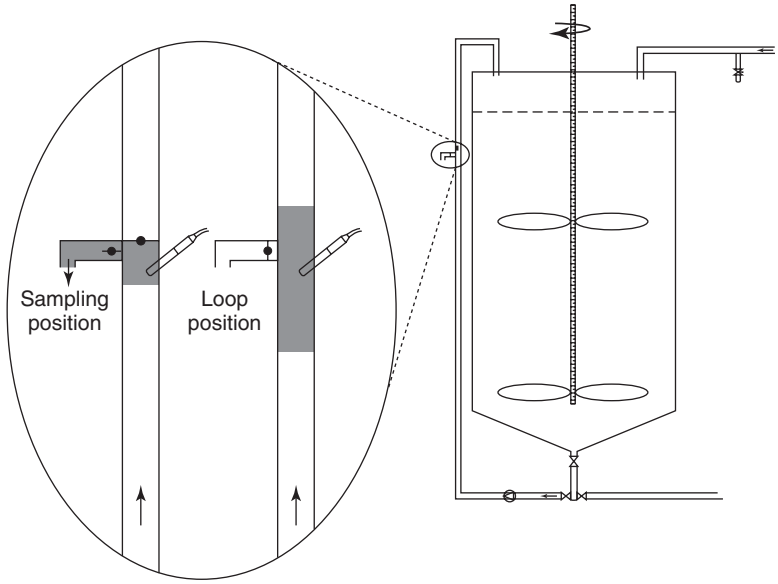


Figure. 3.20 *Colocation of correct upstream sampler and PAT sensor, shown here for a recirculation loop configuration (right). The PAT sensor is deployed in upstream small-diameter piping with a field of view matching as best possible the same stream segment which is being sampled for reference analysis. This setup allows for reliable multivariate calibration because of the colocated [X,Y] modalities.*

to take care of the fatal IDE/IEE problems, it also led to a perhaps equally welcome bonus – a very significantly augmented mixing efficiency:

The material occupying the potentially worst dead zone along the vessel bottom (the potentially most 'deviating' composition on account of gravitative segregation) is the very material first being recirculated by being passed back to the top levels where it will now be subject to conventional stirring with maximal effect.

The self-mixing up-streaming recirculation loop ranks as the prime example of the SUO 'lot transformation' in which a 3-D lot (reactor) has been turned into an 'easy to sample' 1-D configuration (pipeline).

3.6.3 Upstream colocated PAT sampler

This configuration has still another potential bonus. For the upstream sampler, especially in the loop configuration, it is not advantageous to employ large-diameter piping. Indeed it is optimal to specifically use a narrow diameter.

Enter the concept of process sampling colocation. The upstream sampling facility shown in Figure 3.19 can be fitted with any desired in-line PAT sensor or probe facility as well (Figure 3.20). By placing a probe directly *in* the segment being sampled, the differential support mismatch between sampler and probe volume (or field of view) can be deliberately minimized. There is now the largest possible match between the analytical mass (volume) intercepted by the sensor head and the mass (volume) of stream segment characteristics as captured in the PAT signals. Note that it is very difficult to eliminate completely this support mismatch, but the smaller the piping diameter the better. Specific requirements will dictate appropriate diameter specifications etc. In the situation in which this advantage is not enough, it is even possible to go to the extreme and let the PAT-sensor field of view extend to cover the entire cross section (see Figure 3.21).

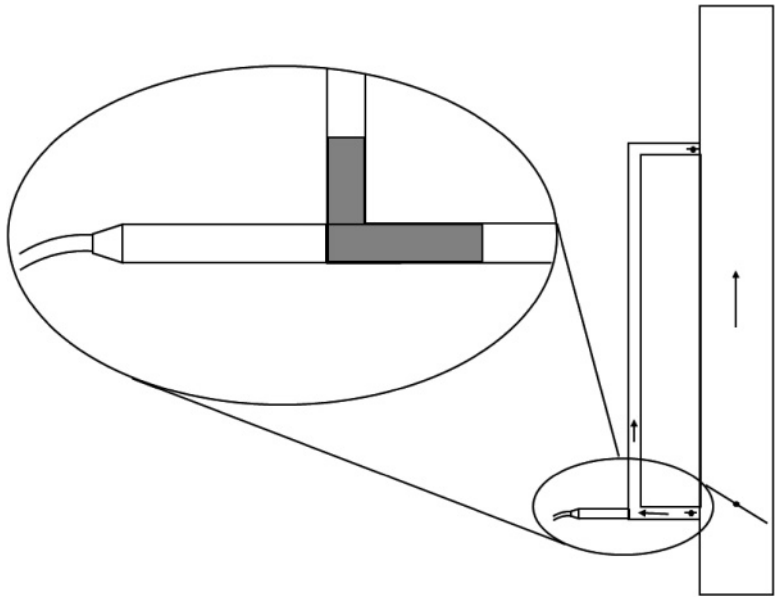


Figure 3.21 *Alternative PAT sensor configuration/deployment in bypass loop configuration. Note that the field of view is here equal to the full stream cross section. Various combinations of the principles illustrated in Figures 3.20 and 3.21 offers convenient process technology flexibility to deal with just about any conceivable scenario otherwise threatened by the traditional fatal IDE/IEE issues illustrated in this chapter.*

The upstream colocated PAT sensor/sampler configuration furnishes the best bid for representative process sampling based only on existing process technology elements. Careful, judicious upstream colocation engineering holds the key to significantly optimized process sampling in current systems. This setup also makes it impossible to succumb to the pitfall of only calibrating on secondary subsamples.

For moving streams of particulate matter (aggregate matter, slurries), typically transported using conveyor belts or similar horizontal agents (slides, trucks, railroad cars, etc.), the equivalent principal solution is the cross-stream cutter, as is illustrated in Figure 3.22.

This case is well described, depicted and illustrated in the pertinent TOS literature, especially with reference to the mining, cement and other process industries [1–3,9–14,21] and need therefore not be dealt with further here. Below the focus is on pipeline and reactor scenarios, specifically for liquid solutions and mixtures, slurries, powders and similar materials of more direct interest to the pharmaceutical and chemical and related industries (food, feed, other).

While the types of lots to be sampled range across a disparate spectrum of materials and phases (1,2,3-phase systems) from analytes in solutions (l), multiphase mixtures (l, s, slurries) to particulate (or solid) aggregates – more often than not with highly significant grain-size ranges and distributions as well (from powders to broken ore), they are specifically *not* to be treated individually on the traditional matrix-by-matrix basis(sic). There is nothing special regarding the main types of materials encountered in the current PAT context, which is precisely the theme developed in this chapter. For two other important process industry domains see Nielsen *et al*. [22] for application of the TOS to the field of biomass conversion while Swanepoel and Esbensen's work [23] concerns wine making, both of which must be treated with identical approaches as those pertaining to other slurries. These references address valuable experience regarding highly heterogeneous materials.

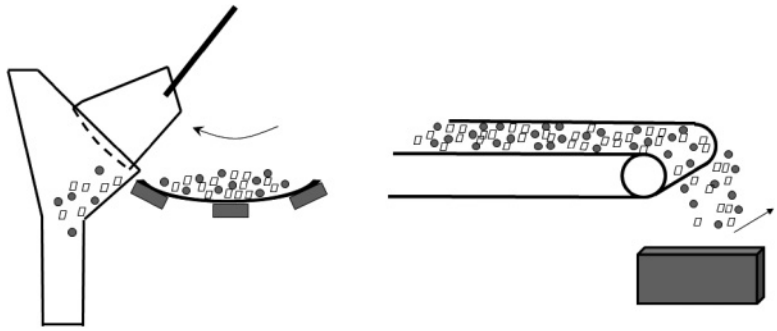


Figure 3.22 *Two realizations of cross-stream cutter solutions to process sampling involving significantly heterogeneous aggregate materials (broken ore, raw material, powders, intermediates, cement, waste, soil, other). The gravity-driven, rotating cross cutter [left] does not comply with TOS' requirements (highly detrimental IDE and IEE), while the automated cross-stream cutter implemented at the terminal end of a conveyor belt [right] can be easily optimized to eliminate all ISEs. See [3,14] for a broad discussion of all issues related to process industry cross-stream sampler principles, and [11,21] for particular details.*

It is the central tenet of the TOS that all lots and materials be treated following the same general set of sampling principles: batch and process sampling both. Indeed *all* sampling situations can be analyzed comprehensively employing only the eight sampling errors. And all practical sampling procedures can be designed, implemented and operated applying only the universal set of seven SUOs.

The remaining part of this chapter deals with the benefits stemming from TOS-guided process sampling. It will be shown how correct primary 1-D sampling employing the powerful data analytical approach called variogaphic analysis will allow unprecedented QC not only of the sampling itself but also of the process variance and its monitoring and interpretation. Variographic analysis allows a breakdown of all sampling errors leading to a unique opportunity for optimizing process sampling. In this endeavor, the 0-D sampling principles outlined above constitute the fundamental bedrock, also in the process sampling regimen.

3.7 Heterogeneity Characterization of 1-D lots: Variography

1-D increments are often correlated along the process dimension (time and/or space), depending on the specific heterogeneity characteristics of the elongated lot. Manufacturing, processing and production lots will by necessity often show strong autocorrelations, while natural processes may or may not exhibit similar characteristics. The salient issue is that variographic analysis is equally applicable for all cases.

In order to sample a 1-D lot, one of three principal sampling schemes – systematic, stratified random or random – must be selected [2,3,8,9,13,14]. Selection of systematic sampling is necessary for initiating variographic analysis. A variogram is used to characterize autocorrelation of 1-D lots as a function of the distance between extracted units (increments). A variogram is superior in identifying trends (increasing/decreasing) and periodic cycles (often hidden to the naked eye) in process data.

1-D sampling requires characterization of the non-random heterogeneity fluctuations *along* the extended dimension of the lot, that is, the heterogeneity between extracted increments. The heterogeneity contribution, h_m, from a 1-D increment is comprised of three parts:

- A *short range fluctuation* component, accounting for the heterogeneity within the particular increment being delineated for extraction (actually the correct 0-D sampling errors FSE + GSE); now known as PIE_1. It is not a requirement that the sampling process is purged of all incorrect sampling errors (ISE) or that ISE has been reduced below a predetermined acceptance level – but the effects of ISE will then of course still dominate, threaten or destroy process insight and/or process control.
- A *long-range fluctuation* contribution that describes the longer-term trends in the process. The new term process integration error (PIE_2) covers what used to be called the time fluctuation error (TFE) in earlier literature.
- A *cyclic* term that describes any periodic variation of the lot that gives rise to a cyclic fluctuation error (earlier called CFE), now called PIE_3.

To characterize the heterogeneity of a 1-D lot, the chronological order of the units must be included in the analysis. Some first-hand knowledge of the process to be characterized is welcome, but *ab initio* variogram analysis is always also possible. Variographic analysis will provide valuable information in the form of the 'nugget effect', the 'sill' and the 'range'. These three key parameters describe different aspects of the process variation and the degree to which this is embedded in incorrect sampling errors. The basics of varigraphic analysis have been extensively described in the TOS literature [1–14].

3.7.1 Process sampling modes

Figure 3.23 illustrates the three principal modes available in process sampling. In systematic sampling a regular interval is used between neighboring samples. With random sampling there is no pattern. The deployment of N_U samples (N units) follows a random distribution over the full process interval to be analyzed. Stratified random sampling first delineates a number of N_U regular strata (N_U intervals) within which random deployment of one sample takes place. All sampling modes are of these three types, or combinations hereof.

3.7.2 The experimental variogram

For a meaningful variographic analysis a minimum of 60 increments are always required. Variographic analysis is problem specific, that is, the lag (interdistance) between any two samples may be measured in minutes, hours, days, weeks, or the data series may be comprised of manufactured units in which a number of equidistant intervals take the role of the elongated dimension. It may at times be necessary to perform two experiments, if both short and long range variation is expected to be of importance, a short range variogram being additional in such cases [2,3,13,14].

A dimensionless lag parameter, j, describes the normalized distance between any two increments:

$$j = \frac{\theta}{\theta_{\min}} \tag{3.17}$$

in which θ is the absolute distance measure (minutes, meters, rank-order index etc.) and $\theta_{\min}$ is the smallest sampling interval selected for the variographic analysis. Figure 3.24 illustrates examples of a set of growing intervals between increments/unit pairs. If the total number of increments extracted is N_U, there will be $N_U - 1$ *increment pairs* with space θ, $N_U - 2$ unit pairs with space 2θ and so on. One usually does not include sample pairs with lags higher than half the maximum number $N_U/2$.

The variogram function, $V(j)$, is defined as ½ times the average squared *difference in heterogeneity contributions* between the sum of pairs of increments, as a function of j:

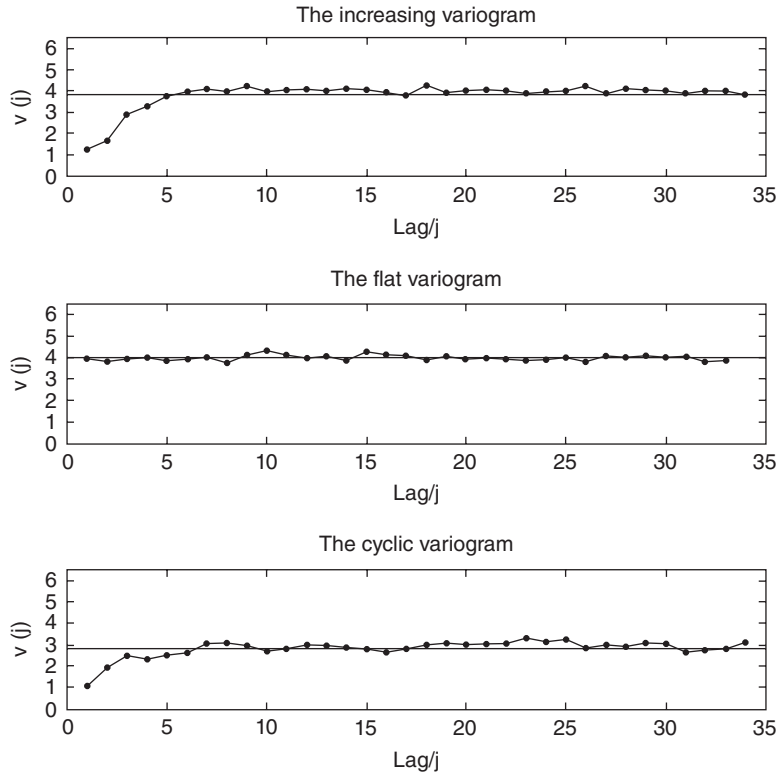


Figure 3.23 *The three basic variogram types: increasing, flat and periodic. In the lower frame a weakly expressed period of approximately 5 lags can be observed superposed on an increasing variogram.*

$$V(j) = \frac{1}{2(N_{\mathrm{U}} - j)} \sum_{m} (h_{m+j} - h_m)^2 \tag{3.18}$$

The variogram function is designed to quantify the process variability in the defining lag domain by expressing the average squared difference between a set of sample pairs, all with identical inter-sample distance, j. In order to comply with conventional statistics, strictly speaking, the above definition quantifies the double of the variance, hence the semivariogram division by two. The variogram can also be expressed by concentrations instead of heterogeneity contributions as shown above. There are no essential distinctions between this so-called absolute variogram (based on concentration) and the relative variogram given here, except in cases where the individual increment masses vary highly (a distinctly unwanted situation), except that the relative variogram to some extent compensates for unwanted large individual variations in increment masses. Further methodological details are found in the literature [2–4,8,9,12–14]; below focus is on the use of variographics in the process analytical technology context.

The proper setting for variographic analysis is a set of 60 representative increments (this is a minimum requirement, 100 samples is always preferable if possible, the minimum number of increments ever successfully supporting an interpretable variogram is 42), in order to cover well the specific process variations to be characterized. It is important to select a $\theta_{\min}$ that is *smaller* than the most probable sampling frequency likely to be used in routine process monitoring and QC. It will be the objective of the data analysis to

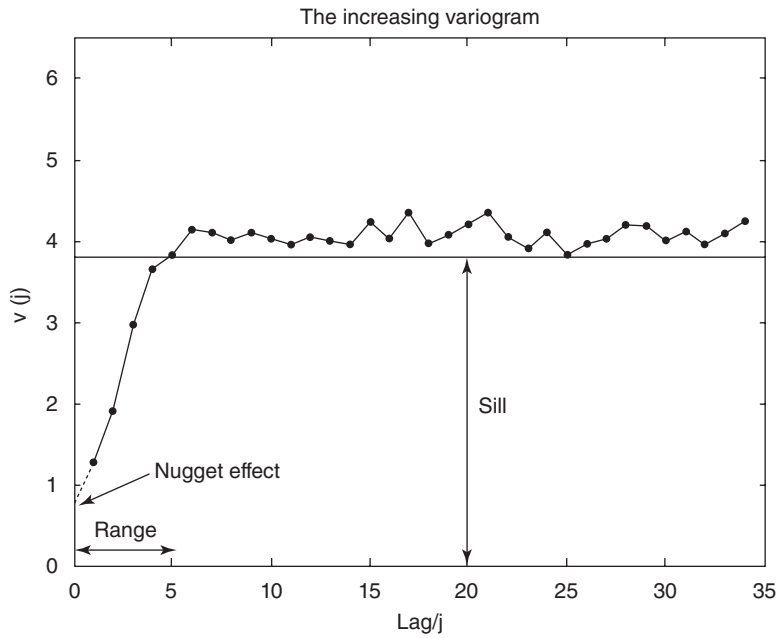


Figure 3.24 *Principal variogram illustrating the three key parameters: nugget effect [MPE, V(0)], range and sill.*

determine exactly this optimal sampling frequency and also to delineate the optimal number of increments to combine into composite samples, that is, a possible reduction of this initial sampling frequency. For this purpose the variographic experiment can be viewed as a pilot study in which this over-sampling will be handsomely rewarded by optimal determination of these sampling parameters. There is therefore very little sense in being frugal with the first deployment of the experimental variogram.

Practical interpretation of the experimental variogram is the first item to address. The variogram *level* and *form* provide valuable information on the process variation captured, as well as the quality of the sampling procedure employed. There are only three principal variogram types encountered in process sampling, but many more combinations hereof:

1. The increasing variogram (the most often, 'normal' variogram shape).
2. The flat variogram (no autocorrelation along the lag dimension).
3. The periodic variogram (fluctuations superposed on an increasing or a flat variogram).

The principal variogram types are illustrated in Figure 3.23. When the basic variogram type has been identified, information on optimized 1-D sampling can be derived. The increasing variogram will be used as an example below.

The variogram function is not defined for lag $j = 0$ – which would correspond to extracting the exact same increment twice. Even though this is not physically possible, it is highly valuable to obtain information of the likely variance corresponding to this 'zero-point variability' (i.e. *what if* it would have been possible to repeat the exact same physical sampling). The TOS identifies this variance as MPE (minimum practical error), which is also known as $V(0)$.

There are many ways to estimate $V(0)$. In the discipline of geostatistics, tractable mathematical models are fitted to the entire variogram, while the TOS takes a pragmatic approach, in which the first five points of the variogram (lags 1–5) are usually back-extrapolated to intercept the ordinate axis. This corresponds to the nugget effect, which constitutes the practical estimate of MPE (usually sufficient as an order of magnitude level estimate only, e.g. Figure 3.24).

MPE encompasses several sampling errors of prime interest in the process context: the FSE, the GSE, the TAE and all ISEs that have not been properly eliminated or reduced, yet. MPE is thus a highly desirable measure of the absolute minimum error that *can* ever be realized in practice. However, MPE is only reached if indeed absolutely *all* vestiges of ISEs have been eliminated, and some form of CSE-reduction has also been applied successfully. Although comparatively easy to estimate (by the variogram), it is usually not without considerable practical effort to actually approach MPE in routine process sampling/monitoring. The objective is always to reach the smallest acceptable TSE, given unavoidable time, effort and economy brackets.

This objective most emphatically cannot be achieved by the simple act of purchasing a particular piece of OEM equipment, whatever marketing declarations notwithstanding. It is essential to be able to analyze professionally the design, functionality, maintenance requirements, etc. of all process sampling equipment and solutions. The TOS constitutes the full comprehensive guarantee for all these stipulations [1–14].

When the increasing variogram becomes more or less flat, the sill has been reached. The sill of the variogram provides information on the expected maximum sampling variance if the existing autocorrelation is not taken into account. This corresponds to an indiscriminate random sampling with time (or along the static elongated dimension).

Conversely, with decreasing lag down to lag one (the increasing part of the variogram, but in the reverse direction) the dropping off of the variance from the sill level is a manifestation of more and more expressed autocorrelation, steadily becoming larger as the intersample lag becomes smaller. The range of the variogram is found as the lag beyond which there is no longer any autocorrelation.

These three characteristic variogram parameters are illustrated in Figure 3.24. To the experienced process sampler, these three parameters contain most of all the primary interesting information needed for sampling error reductions and/or process monitoring and control. Proper process sampling, besides the categorical imperative of securing a documented bias-free sample/signal acquisition, is also based on a competent variographic analysis; for more advanced analysis of the variogram, see Pitard [3].

If significant periodicity is observed in the experimental variogram, it is essential to make sure that the sampling frequency is not close to, and absolutely never identical to this period [2,3,9–13]. Also the specific sampling mode (random sampling: **ra**, systematic sampling: **sy** and stratified random sampling: **st**) becomes important. From the standard variogram (equidistant sampling) it is also possible to simulate the TSE effects of the two other sampling modes [2,3,9,12,13].

From the variographic experiment an optimal sampling plan can be derived. This includes finding practical ways to minimize TSE. Full technical descriptions can be found in [2,3,8,9–13], a practical summary of which would look like this:

- Avoid extracting increments with a frequency coinciding with a period. Doing so will significantly *underestimate* the true variation of the process.
- In general, the stratified random sampling mode (**st**) will always lead to optimal results (absolute lowest TSE), but systematic sampling (**sy**) will in many cases be just as effective, and almost always much easier to implement. When the objective is to derive maximum of reliable information from process monitoring, the TOS is adamant in that random sampling (**ra**) should never be used. The reality is often otherwise in many industrial environments, in which grab sampling (at more or less random time points) often still rule the day (see Figure 3.9).

- Sampling with a frequency below the range will ensure that the process autocorrelation in effect *reduces* the sampling variation, dependent upon the effect of possibly combining increments into composite samples (see further below).
- $V(0)$ is identical to the estimate the minimal practical sampling error MPE. From this further breakdown of the constituent error sources is often possible [2,3,9,12,13,].

3.7.3 Sampling plan simulation and estimation of TSE

Estimation of TSE is now easy, as it follows directly from the same basic calculations attending the experimental variogram. At no additional sampling or analytical cost, it is now possible to reliably *simulate* all conceivable sampling procedures that might be conceived for improvement of a current sampling process. A successful variographic analysis thus allows a complete 'what if' simulation of all possible sampling plans, campaigns, etc.

Any process sampling scheme (**sy**) is fully characterized by two intrinsic parameters only: the sampling rate, r, and Q, the number of increments per sample one is willing to composite in order to reduce TSE if deemed too large. This simulation should be designed so as to evaluate all desired *combinations* of Q at each potential sampling rate of interest $\boldsymbol{r}$. For practical reasons the sampling rate is often quantified by the inversely related lag distance. This simulation can be performed using any existing commercial software for variogram calculation (or by freeware alternatives). In Figure 3.25 the lower right panel delineates the estimated TSE, either in absolute units corresponding to the units for original data series, or in relative [%] units.

A standard range for Q spanning 2, 4, 8 increments can often be used as a starting point (but the user needs to be able to specify any Q). This corresponds to substituting all individual increments in the variographic experiment (initially considered samples in their own right $Q = 1$), by composite samples made up of 2, 4 or 8 increments instead. These simulated composite samples are centered on the average individual sampling times, or locations. In addition, useful variogram software must also allow for completely free specification of the sampling rate, r.

Together, these features allow simulation of any specific sampling scheme that might be envisaged for a particular process sampling situation. Based on only one correctly sampled data set (60–100 samples), all possible sampling scenarios can be simulated at no extra cost. Obviously, the data set for the variographic experiment must be as representative as possible of the prevalent process behavior, hence the mild insistence on the higher end of the interval of 60–100 samples (increments). It will always be a specific, problem-dependent challenge to select such a data set. Local process knowledge and experience is at a premium in all such undertakings, and is much more important than any standard statistical prescriptions that may be gleaned from the literature.

On this basis it is easy to zoom in on an optimal sampling strategy by inspecting the plot of TSE as a function of Q and r. The best approach is to inspect the TSE panels, as illustrated in Figure 3.25 (bottom right panel), to decide on which direction will provide the steepest reduction (steepest descent) in TSE. This will either be by use of a higher sampling rate r, or by using more increments Q. For the variographic experiment represented by Figure 3.25, it is clear that composite sampling is by far the most effective way to decrease TSE. If both parameters appear to influence in an equal fashion, any combination may be chosen which will then also be able better to satisfy external, especially economical or practical, constraints. It may perhaps seem easier to include more increments in a composite sample, or vice versa. Increasing the sampling rate may appeal in other situations, often influenced by the existing sampling equipment (*correct* cross-cutter equipment only). Be this as it may, this is also heavily dependent on the specific lot heterogeneity and other available sampling options. All TSE simulations are real-time, and there is no extra sampling or computation necessary. This is the singular most powerful aspect of variographic analysis. The user should always also

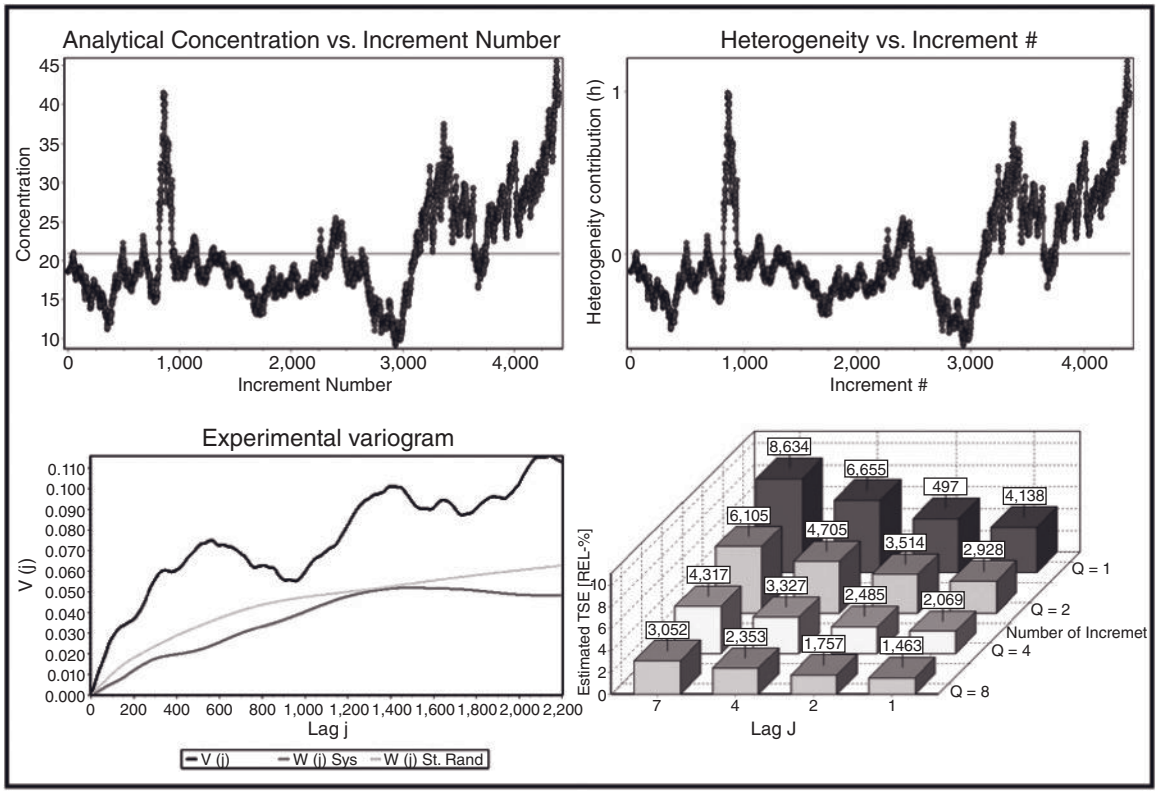


Figure 3.25 *Illustration of comprehensive variographic analysis. Top panels show original data series (left) and its rendition in heterogeneity contributions (right). Bottom panels, (left): the experimental variogram (with two auxilliary functions); (right) TSE estimations for the user-specified set of sampling rate, r, and Q.*

invoke his/her own creativity and sampling experience pitted against the realities of the pertinent economic, practical or other constraints in this context.

3.7.4 TSE estimation for 0-D lots – batch sampling

1-D variography allows verification whether current TSE is below *a priori* determined levels or not. A comparable quantitative sampling effect of DH_L for 0-D lots (batch sampling), that is, the 0-D TSE, which can also be estimated with ease. In practice this is done by extracting and analyzing a number of replicate samples covering the entire geometry of the lot and calculating the resulting empirical variance based on the resulting analytical results. Often a relatively small number of primary samples will suffice, though never less than 10 – preferably more – as the validity of the TSE estimate is directly related to the degree of representativity of this number of samples with respect to the entire lot heterogeneity, DH. This procedure is termed a *replication experiment*. It is critical that replication involves the entire sampling and analysis chain, especially that the primary sampling is replicated in a completely independent fashion, *including* all subsampling and mass-reduction stages, sample preparation etc. It is of course an ideal requirement that all ISEs have indeed been eliminated, i.e. that only correct sampling is employed, lest the estimated total sam-

pling variance be completely inflated by the inconstant bias incurred. Esbensen and Julius [19] give a delineation of the replication experiment in more detail.

3.7.5 Corporate QC benefits of variographic analysis

From a practical sampling point of view, the proportion between $V(0)$ versus sill levels in a variogram provides essential information regarding the quality and reliability of sampling and analysis vs. optimal insight into process variation. Both are important objectives in the process QC regimen. Consider as an example the situation in which a multinational company owns production units in several countries, in which local raw materials are used exclusively (bulk raw materials cannot be shipped between countries for economic reasons). The company is under heavy market obligations to produce identical products within strict specifications, irrespective of country of origin. In contrast to raw materials, products can be, and are, in fact often shipped between countries. The company produces a line of many different products within a wide industrial sector, but the variographic analyses below concerns one specific product only, the volumetrically and economically most important one. The results below are inspired by a real-world company and data sets, which have been suitably cryptized for propriety reasons, as have the product type. But to set the general scene, consider value-added marine products in the food/feed sector. Another example would be finished and semifinished products for the construction industry, or similar. The essential issue here is a bulk product, whose QC is critical.

Figure 3.26 juxtaposes results of variographic analysis from six different production sites in a multinational company context. Sites A, B, C, E, and F all use the same production process, while site D has been allowed to maintain an earlier process – due to local claims of equal efficiency. It was originally decided to allow this, but only *if* a relevant comparison could justify this claim. Thus the regional manager in charge of quality control ordered a 'relevant comparison' to be made. Luckily this coincided with introduction of new competencies in the company regarding chemometrics/PAT/TOS, so six experimental variograms could be ordered and developed with relative ease. Because of an informed central QC officer and local TOS competencies, it was possible to base the comparison on identical sampling rates, etc. For this comparison, analysis based on each individual increment was deemed optimal.

This comparison revealed not only an inconvenient truth regarding the production process at site D, but also other surprising distinctions between the five sites/processes hitherto assumed to be operating with about equal reliability and efficiency. With reference to the basic decomposition features of any variogram (Figure 3.24), the following can be concluded from the results presented in Figure 3.26.

- Site A is by far the best operating site: very low process variability, very low sill (TSE).
- Sites B and F also display low sills (TSE), only a factor 1.5–2 higher than that of A.
- Site C displays a TSE approximately three times that of A. This was deemed too high
- for corporate comfort.
- Sites D and E operate with the highest TSE, approximately four times that of site A, very clearly also out of any acceptance zone.

Further:

- Sites A and D experience TSE to such a magnitude that no process variation can be observed at all (MPE *equals* the sill – a flat variogram). This is caused either by a production process in complete control, site A, or by the least efficient process sampling within this particular corporation, site D, actually caused by a completely unacceptable sampling procedure.

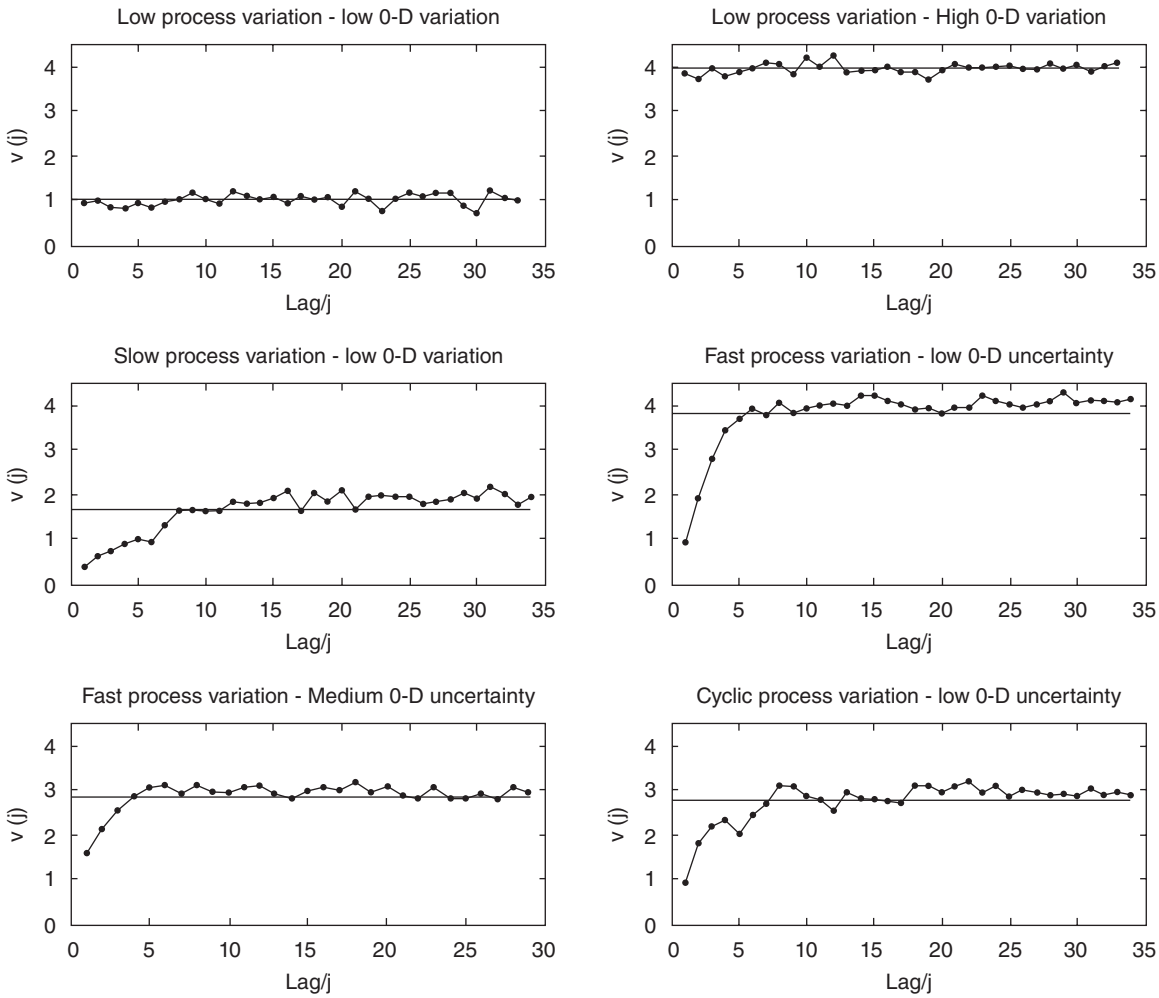


Figure 3.26 *Variogram analysis of one identical product from six different production sites – from six different countries, each employing local raw materials. Six variograms allow for critical corporate comparison of process stability/quality versus sampling (see text for full details). Sill levels for all six cases are highlighted. Panels are labelled top to bottom as A, B, C (left column) and D, E, F (right column).*

- Sites B, C, E, F all, to varying degrees, display a MPE significantly lower than their respective sills. All these sites thus operate with a TSE so much in control that the relevant process variability can indeed be reliably observed. However for site C, half the sill level is MPE, which likely indicate that significant ISE is still present. Contrary to this, sites B, E and F possess such low MPE that this issue warrants no further action. For these sites, most (B, E) or all (F) of the sill variation beyond the range reflects pure, uncompromised process variation.
- For site B the variogram shows a significantly large range (15 lags), indicating that there are ample opportunities available to invoke composite sampling to lower the TSE (approaching the MPE level) if so desired. For sites C, E and F, the range is smaller (approx. 5–7 lags in all cases) - here the autocorrelation is stronger (C, E).

The corporate conclusions from this audit were:

1. The three sites with significant high sill levels were ordered a complete TOS-evaluation of their process sampling (C, D, and E).
2. Sites with low sill levels (as measured by in-house corporate thresholds) were deemed 'fit for purpose' (A, B, F).
3. Site E was directed to look into its marked lack of process stability. TSE issues could easily be brought under control using composite sampling.

Comprehensive quality control could here be carried out at on unambiguous quantitative basis, with full site specificity, based on only six experimental variograms.

Variography constitutes a much overlooked QC tool in the process industry at large, and particularly within PAT. This feature is not restricted to industrial processes. Many natural processes can be characterized by variography as well. Examples include many types of time-series, geographical transects, etc. There is a grading overlap with the science of geostatistics.

Repeated variographic analysis on experimental data sets obtained at relevant QC intervals at the same plant or production unit (machine, oven, reactor, etc.) will allow the same range of possibilities for characterization as shown above. All irregularities in the process (manufacturing, processing, natural,, etc.), e.g. trends, periodicities, upsets, change(s) in sampling error manifestations will lead to a change in the variogram (either in sill level, nugget effect (MPE), range, or in the form of the variogram). Process engineers and other QC personnel must be fully conversant with process sampling and variographic analysis in order to reap these and many other advantages. More details concerning process sampling variograhic analysis can be found in references [3,13,14].

It will become clear that a complete QC check-up should consist of repeated variographic analysis combined with a test set validation, at appropriate time intervals commensurate with the production or plant situation at hand.

3.8 Data Quality – New Insight from the TOS

Data quality is a broad, often loosely defined term. There are many problem- and discipline-related definitions to be found in the literature. This section shall not try to define data quality in any comprehensive, far less complete sense – suffice to denounce any definition that does *not* include the specific aspect of sample representativity however. Data is often equated with information, but this can only be in a hidden, potential form. Only data analysis together with interpretation *may* reveal 'information' – which will always be in a particular problem-specific context only. Such issues are not usually seen as problematic in chemometrics and in PAT, where the pre-history of a data table ('data') in general receives but scant attention. One relevant, major exception is Martens and Martens (2001) [26] who focus comprehensively on 'Multivariate Analysis of Quality'. But even here there is a narrow focus on quality of information only, defined as '… dependent on *reliability* and *relevance*', without further clarifying the definition of these open-ended adjectives.

In as much as the present TOS exposé has specified clearly that 'reliable analytical information' derives from 'reliable analytical results (data)', meaning analytical results exclusively stemming from representative samples and commensurate sampling plans, it stands to reason that any definition of data quality perforce must include reflections/reference to representative sampling. Thus it is crucial also to contemplate the specific origin(s) of *any* data set in the context of the TOS.

From the perspective presented in this chapter, manipulations and 'modeling of data' without proper considerations as to their representativity (strictly speaking the representativity of the samples analyzed)

may not necessarily always end well. Lack of respect for the basic rules delineated in the TOS, i.e. focus on the data alone, constitute a sin-by-omission. In the context of PAT all aspects related to process analytical sampling (0-D and 1-D TOS) are today still very much in their infancy. The TOS indeed constitutes a *missing link* in PAT.

What is usually called 'measurement error' typically only relates to the X-data (instrumental signal errors). However even where the magnitude of this type of error is at a maximum, it is only of the same order of magnitude as the analytical errors, making this type of error almost as irrelevant as TAE *compared* to the TSE (and its breakdown). In spite of this insight, fierce debates over data quality, measurement errors, 'data analysis robustness' and similar break out in chemometric and data analysis circles regularly. This completely misses the mark in view of the relative magnitude of signal/instrument vs. sampling errors as only fully revealed by the TOS' more than 50 years' of experience: TSE is typically can range anywhere between 20–100 times larger than TAE.

3.9 Validation in Chemometrics and PAT

Section 3.8 also has a direct bearing on the issue of proper validation in data analysis/data modeling in chemometrics and PAT. Validation is concerned with validating the performance of data analytical models, be it e.g. for prediction, classification, time series forecasting or similar. Also here the TOS' framework in general, and process sampling in particular, sheds new light on the oldest evergreen in chemometrics: how to validate?

In statistics, data analysis and chemometrics, a much favored method of validation is cross-validation, in which a subset of the training set performs similarly as an apparent 'test set' in a cumulative sequential fashion. Depending on the fraction of training set samples (totaling N) held out of the contemporary submodeling to be used for validation a optional number of no less than $(N - 1)$ potential cross-validation segments can be chosen in the interval $[2, 3, 4 \ldots (N - 1), N]$. Various schools of thought of cross-validation have developed within chemometrics, some favoring full cross-validation (one object per segment; N segments in total), some defining ten as the canonical number of segments (ten segments, each with 10% of N objects), as well as other schemes, while a minority recognize that there is in fact more complexity to the matter than selecting a particular number from this range of $(N - 1)$ variants of cross-validation. But usually there is more focus on strict adherence to *some* form of fixed cross-validation procedure, or other, than openness with respect to the precise assumptions behind cross-validation. This comprehension has hitherto not received much attention in chemometrics and PAT. The TOS illuminates this issue clearly.

Esbensen [27] analyzes the fundamental aspects of cross-validation versus test set validation, further corroborated and effectively illustrated by Esbensen and Lied [28]. The general conclusion is that cross-validation is only a *simulation* of test set validation, in form strikingly similar, but most emphatically *not* in essence. By design, cross-validation specifically works on only one (training) data set, in which context a series of submodels are established based on part of this data set only. All submodels have a smaller basis than the entire training set. By sequential exclusion of one object, or one segment, to be used for validation, the mechanics of cross-validation is often misconstrued as being similar to using an external (test) data set. Unfortunately this is a critically faulty comparison, as left-out data are immediately *replaced* in the training data set when other objects/segments in turn are excluded. Any two submodels in fact have $(N - 2)$ objects in common (full cross-validation); similarly with respect to segments. Many arguments in favor of cross-validation have been marshaled: 'Easy, effective, easy-to-program, saves money (no need for doubling the analytical costs (Y)'. The fact is that all these arguments miss the singular, critical issue: there is only ever *one* data set involved in cross-validation.

The underlying assumption is further that this procedure leads to information as to the future performance of the full *N*-object model; but this is an equally flawed assumption. This procedure has patently no link to any new data set, 'new measurements', generated after the model has been established and cross-validated. In reality, all that cross-validation delivers is a measure of internal training set stability with respect to subsetting (sequential exclusion of one object, or one segment).

The essential characteristic of a proper test set is that it represents a new 'drawing from the population', realized as a new, independent [*X*,*Y*] data set specifically not used in the modeling. It is evident that any *N*-object data set constitute but only one specific realization of an *N*-tuple of individual TSE materializations. It takes a completely new ensemble of objects, the test set, to secure a second manifestation. All new measurements, for example when a PAT model is used for the purpose of automated prediction, constitute precisely such a new drawing/sampling. All new measurement situations are therefore to be likened to a test set – and this is exactly what is missing in all forms of cross-validation.

The main lesson from more than 50 years of experience with TOS is that there is no such thing as a constant sampling bias; it changes with every new extraction from heterogeneous material. From this follows that there can never be a guarantee that the specific training set realization will also be representative with respect to (all) future data sets. It will therefore always be necessary to invoke one test set (at least) in any serious validation of future data model performance. TSE for the new data set is not identical to TSE for the training set precisely because of the heterogeneity of the lot from which both sets originate. Incorporating the second data set in the validation is therefore the only way in which to allow information from both TSE materializations to influence the validation.

While the training set is only used for modeling, the test set is only used for performance testing. In prediction performance for example, the predicted values represent new *X*-data used as input to the training set model, which then predicts the corresponding *Y*-values ($Y_{i,predict}$) – to be compared with the test set $Y_{i,reference}$ values:

$$\mathrm{RMSEP} = \sqrt{\frac{\sum_{i=1}^{n} \left(Y_{i,\mathrm{pred}} - Y_{i,\mathrm{reference}}\right)^2}{n}} \tag{3.19}$$

From the above discussion it follows that all types, variants or schemes of the cross-validation type by necessity must be *inferior*, indeed unscientific [27,28], precisely because no TSE-contributions from any future data set is ever validated. Only test set validation stands up to the logical demands of the validation imperative itself. Cross-validation is therefore logically and scientifically discredited and must never be used as a replacement for a true test; *if* cross-validation *is* used, this must always be accompanied by appropriate warnings [27,28].

Admittedly there do exist a few, rare situations in which no option for test set validation is possible (historical data, very small data sets, other …). In such cross-validation finds its only legitimate application area (NB *None* of these situations must result from voluntary decisions made by the data analyst, however). In historical data there simply does not exist the option to make any resampling, etc. In small data sets, this option might have existed, but perhaps went unused because of negligence – or this 'small sample case' may be fully legitimate.

The salient issue is here: why insist on validating such a small data set in the first place? Even the basis for modeling is already dangerously weak as it is. Better not to validate at all, but put all efforts into securing a more sufficient data basis [*X*,*Y*]. Whenever legitimate cross-validation is applied, only two-segment cross-validation has any merit, though principally still flawed [27,28] contains the proper argumentation in more detail.

Within process sampling, there are many opinions as to optimal definition of the training set versus the test set. In a 5 year data series for example, would the first 3 years constitute a viable (optimal?) training set, with the last 2 years serving as an appropriate test set? Substitute month, week or day for year – the principal issue is identical. Or would pruning over the entire 5-year (-month, -week, -day) set be a better solution, for example taking out every fifth sample to make up the test set (20% relative to the 80% training set)? With process sampling there will always be many local, problem-specific variations of this common theme. But most process sampling scenarios are in fact more complicated than a simple carrying over of 0-D notions and experiences and may include trends, seasonality or even more complex issues (process upsets).

The TOS has made specific allowance for the first two general factors in the form of the time fluctuation error (PIE_2) and the cyclic fluctuation error (PIE_3) respectively. Process sampling schemes not derived from a proper variographic basis will always be in danger of not having captured the essence of the interrelationships between the current (or TOS improved) sampling procedures and the process variations. One excellent validation consists in repeating the variographic experiment at regular intervals. Consistent variograms must reflect that both process variations as well as TSE are under proper control. Deviations can be observed immediately, and variographic analysis allows swift decomposition into the relevant MPE, ICS and CSE, from which the most relevant counteraction will be unambiguously indicated. Repeated variographic characterization in fact constitutes a novel validation approach, specifically in line with the general objectives of PAT. A similar repetition of a test set validation should accompany this – in order for reliable QC in PAT domain to have been fully documented.

3.10 Summary

1. The fundamental principles for representative sampling and effective counteractions towards heterogeneity were delineated for batch sampling (0-D) as well as process sampling (1-D). All principles are contained in the theory of sampling (TOS).
2. First priority is always elimination of sampling bias as generated by ISE (IDE, IEE, IWE, IPE). This principle reigns supreme for both 0-D and 1-D scenarios. If this imperative is not adhered to, sampling will be but an exercise in futility. All principal ISE-eliminating procedures for process sampling were delineated, and further literature was offered in abundance.
3. It is vain to address the subsequent objective of CSE-reduction without unconditional commitment to and respect for the rules for *sampling correctness*. All pertinent specifications were given above. Nobody and no science (chemometrics included) will be able to escape from neglect without suffering significant scientific and economic penalties.
4. CSE reduction is a matter of thorough mixing ('homogenization'), applying composite subsampling with many, small increments or reduction of the top particle diameter (crushing) [0-D]. In process sampling however, the practical possibilities are often limited to optimization of sampling rate, $\boldsymbol{r}$, and/or how many increments, Q, to composite. Process sampling offers a unique, powerful data analytical facility for this, *variography*.
5. Process variography provides verification whether current TSE is below *a priori* determined, acceptable levels or not. For 0-D, this may be accomplished, for example by performing a replication experiment on the complete sampling procedure ten times (independently). It is essential that replication includes all aspects of the current sampling process from primary sampling to analytical result. McClellan and Kowalski [18] gives a brief delineation of the so-called 'replication experiment', which can be employed in the 0-D as well as in the 1-D domains.

6. If TSE levels are verified as higher than acceptable, sampling and analysis is to be terminated with prejudice, as the data generated are unreliable, futile and a waste of practical effort. It is an economic waste to sample, monitor or make decisions based on a documentable inferior basis.
7. Such conclusions (termination) can prevent economic losses or disasters with a minimum of capital outlay. The TOS comprises everything needed to improve on any inferior sampling process or procedure. The present chapter is but an introduction sufficient for the process analytical domain. The reader is also referred to the literature list and the secondary literature referenced herein.
8. Variographic analysis is an effective QC tool in the corporate setting. Variographic nugget effects versus sill levels (absolute and relative) provide insight into TSE and all ISE present, allowing for meaningful comparison between production sites, competing manufacturing process, etc. A variogram can also be used as a quality assurance tool by repeated monitoring of any one process in time guarding against e.g. upsets, trends, unknown periodicities.
9. A full QC check of the PAT-monitoring system employed should consist of repeated variographic analysis *and* test set validation at appropriate time intervals commensurate with the production or plant situation.
10. This chapter introduced application of simple variographic experiments as a basis for problem-dependent deliberations as to an optimal deployment of increment/sample sampling rate, $\boldsymbol{r}$, and/or compositing using Q increments/sample. The variographic experiment constitutes the most powerful process sampling data analysis tool available.
11. There are several variographic software packages available including freeware versions commensurate with the above demand for proficiency. All of these can perform the necessary data analysis shown above; a few sampling consultant programs include more advanced features as well.

References

1. L. Petersen, P. Minkkinen and K. H. Esbensen, Representative sampling for reliable data analysis: Theory of Sampling, *Chemom. Intell. Lab. Syst.*, **77**(1–2), 261–277 (2005).
2. P. Gy, *Sampling for Analytical Purposes*, John Wiley & Sons, Ltd, Chichester, 1998.
3. F.F. Pitard, *Pierre Gy's sampling theory and sampling practice* (2nd edn), CRC Press LLC, Boca Raton, 1993.
4. P.L. Smith, *A primer for sampling solids, liquids and gases – Based on the seven sampling errors of Pierre Gy*, ASA SIAM, Philadelphia, 2001.
5. P.M. Gy, The analytical and economic importance of correctness in sampling. *Anal. Chim. Acta*, **190**, 13–23 (1986).
6. L. Petersen, C.K. Dahl and K.H. Esbensen, Representative mass reduction: a critical survey of techniques and hardware, *Chemom. Intell. Lab. Syst.*, **74**, 95–114 (2004).
7. P. Minkkinen, Evaluation of the fundamental sampling error in the sampling of particulate solids, *Anal. Chim. Acta*, **196**, 237–245 (1987).
8. R. Heikka and P. Minkkinen, Comparison of some methods to estimate the limiting value of the variogram, $v_h(j)$, for the sampling interval j=0 in sampling error estimation, *Anal. Chim. Acta*, **346**, 277–283 (1997).
9. P. Gy, Sampling of discrete materials – III: Quantitative approach – sampling of one dimensional objects, *Chemom. Intell. Lab. Syst.*, **74**, 39–47 (2004).
10. P. Gy and L. Marin, Unbiased sampling from a falling stream of particulate material, *Int. J. Miner. Process*, **5**, 297–315 (1978).
11. P. Gy, Does your mechanical sampler do what it's supposed to? *Coal Mining and Processing*, **18**(12), 71–74 (1981).
12. L. Petersen L and K.H. Esbensen, Representative process sampling for reliable data analysis – a tutorial, *J. Chemom.*, **19**(11–12), 625–647 (2006).
13. K.H. Esbensen, H.H. Friis-Petersen, L. Petersen, J.B. Holm-Nielsen and P.P. Mortensen, Representative process sampling – in practice: variographic analysis and estimation of Total Sampling Errors (TSE). Proceedings 5th Winter Symposium of Chemometrics (WSC-5), Samara 2006. *Chemom. Intell. Lab. Syst., Special Issue*, **88**(1), 41–49 (2007).

14. K.H. Esbensen and P. Minkkinen, Special Issue: 50 years of Pierre Gy's Theory of Sampling. Proceedings: First World Conference on Sampling and Blending (WCSB1). Tutorials on Sampling: Theory and Practice, *Chemom. Intell. Lab. Syst.*, **74**(1), 236 (2004).
15. Rationel Kornservice A/S website. http://www.rako.dk/ accessed 12 November 2009.
16. O. Miserque and E. Pirard. Segregation of the bulk blend fertilizers. *Chemom. Intell. Lab. Syst.*, **74**, 215–224 (2004).
17. VARIO – Freeware for 1-D variographic calculations. Available at www.acabs.dk, accessed 12 November 2009.
18. F. McClellan and B.R. Kowalski (eds), *Process Analytical Chemistry*, Blackie Academic and Professional, London, 1995.
19. K.H. Esbensen and L.P. Julius, Representative sampling, data quality, validation – a necessary trinity in chemometrics, in *Comprehensive Chemometrics*, S. Brown, R. Taulor, and B. Walzak (eds), vol. **4**, 1–20. Elsevier, Oxford, 2009.
20. L. Boland and K.H. Esbensen, TOS evaluation of process technology standard sampling valve designs – apparent reproducibility but fatal accuracy, in *Proceedings of the 4th World Conference on Sampling and Blending* (WCSB4), Cape Town, October 21–23 2009, South African Institute of Mining and Metallurgy (SAIMM), Symposium Series S59. Johannesburg 2009. ISBN 978-1-920211-29-5.
21. R. Holmes, Correct sampling and measurement – the foundation of accurate metallurgical accounting, *Chemom. Intell. Lab. Syst.*, **74**, 71–83, (2004).
22. J.B. Holm-Nielsen, C.K. Dahl and K.H. Esbensen, Representative sampling for process analytical characterisation of heterogeneous bio-slurry systems – a reference study of sampling issues in PAT, *Chemom. Intell. Lab. Syst.*, **83**, 114–126 (2006).
23. M. Swanepoel and K.H. Esbensen, Representative sampling at the grape intake: a critical issue – often forgotten, *Australian and New Zealand Grapeowner and Winemaker*, **521**, 32–37 (2007).
24. P.P. Mortensen and R. Bro, Real-time monitoring and chemical profiling of a cultivation process, *Chemom. Intell. Lab. Syst.*, **84**(1–2), 106–113 (2006).
25. A.J. Ward-Smith, *Internal Fluid Flow*, Clarendon Press, Oxford, 1980.
26. M. Martens and H. Martens, *Multivariate Analysis of Quality. An Introduction*, John Wiley & Sons Ltd, Chichester, 2001.
27. K.H. Esbensen, *Multivariate Data Analysis – in Practice. An Introduction to Multivariate Data Analysis and Experimental Design*, 5th edn, CAMO AS, Oslo, 2001.
28. K.H. Esbensen and T.T. Lied, Principles of image cross-validation (ICV): representative segmentation of image data structures, in *Techniques and Applications of Hyperspectral Image Analysis*, H. Grahn and P. Geladi (eds), Chap. 7. (155–180), John Wiley & Sons, Ltd, Chichester, 2007.

4

UV-visible Spectroscopy for On-line Analysis

Marcel A. Liauw[1], Lewis C. Baylor[2] and Patrick E. O'Rourke[2]

[1] *ITMC, RWTH Aachen University, Aachen, Germany*
[2] *Equitech International Corp., New Ellenton, SC, USA*

4.1 Introduction

The use of ultraviolet (UV) spectroscopy for on-line analysis is a relatively recent development. Previously, on-line analysis in the UV-visible (UV-vis) region of the electromagnetic spectrum was limited to visible light applications such as color measurement, or chemical concentration measurements made with filter photometers. Three advances of the past two decades have propelled UV spectroscopy into the realm of on-line measurement and opened up a variety of new applications for both on-line UV and visible spectroscopy. These advances are high-quality UV-grade optical fiber, sensitive and affordable array detectors, and chemometrics.

Nonsolarizing (or at least solarization resistant) optical fibers make analyses at wavelengths shorter than 280 nm possible by fiber-optic spectroscopy. Prior to this improvement, optical fibers quickly developed color centers when exposed to intense UV radiation from either a deuterium lamp or a xenon flash lamp. The light transmitting ability of the fiber degraded quickly, often in a matter of a few minutes. Current optical fibers maintain their light transmission at or near that of a fresh fiber for months or years. The length of the fiber run, nevertheless, must be kept fairly short if the analytical work is to be done in the deep UV range of 190–220 nm as the transmission efficiency of the fibers at these wavelengths is decreased relative to the range of 220–280 nm. Fiber-optic probes built with nonsolarizing fiber make possible *in situ* sampling of a process, at the same time allowing the detection equipment to be positioned safely away from the process hazards.

The emergence of sensitive and affordable array detectors in 1980s and 1990s has also improved measurement capability in the UV-vis. Commercially available UV-vis instrumentation with photodiode-array (PDA) detectors appeared in the mid-1980s. This made it possible to produce a UV-vis spectrophotometer

Process Analytical Technology 2e Edited by Katherine Bakeev
© 2010 John Wiley & Sons, Ltd.

Table 4.1 *Reviews on UV-vis spectroscopy for process analytics in selected fields of application*

Field of application	Specifications	Reference
Process analyses	On-line composition monitoring, closed-loop control	Saltzman, 2000[1]
Pulp and paper industries	Monitoring of bleaching, yellowing	Schmidt, 2000[2]
Pharmaceutical solids	Monitoring of pharmaceutical syntheses, dissolution testing, quality control	Brittain, 2006[3]
Environmental issues	Pollution detection, wastewater monitoring	Bosch Ojeda/Sánchez Rojas, 2009[4]
Catalysis	Monitoring of preparation, activation, deactivation, study under working conditions	Weckhuysen 2004[5]; Jentoft 2009[6]
Clinical chemistry	Enzyme rate assays, colorimetric assays, end-point assays, immunoassays	Upstone, 2000[7]

with no moving parts, with the exception of a shutter for the lamp. PDAs work well in high-light applications, such as absorption spectroscopy. Charge coupled device (CCD) detectors began to appear in commercial process instrumentation in the 1990s. These detectors offer improved sensitivity over the PDA and are two-dimensional, rather than just line arrays. Front-illuminated CCDs may be used in the UV if they are coated by a chromophore that absorbs UV radiation and re-emits the energy as a visible photon. Back-thinned CCDs directly sense UV photons and are about ten times more sensitive in the UV than a front-illuminated CCD or PDA.

Finally, the development of chemometrics over the past 20 years has also aided in the use of UV-vis technology for more complicated chemical matrices than was possible at earlier times. Chemometrics allows large quantities of spectral data to be analyzed and reduced to a useful bit of information such as the concentration of one or more chemical species. In addition, through analysis of residuals, it is possible to detect when something unexpected occurs in a process.

As will become obvious in this chapter, UV-vis spectroscopy is a valuable tool for process analytics in a wide range of industrial applications, be it in production-related R&D or in actual process monitoring and control. An overview of reviews for the various fields is given in Table 4.1.

This chapter is divided into an introductory 'Theory' section where the underlying principles are outlined, and an 'Applications' section that presents selected examples of on-line UV-vis spectroscopic monitoring of industrial processes.

4.2 Theory

4.2.1 Chemical concentration

4.2.1.1 Transmission and Absorption

Most liquid- and gas-phase UV-vis spectroscopic measurements rely on the well-known Lambert–Beer Law:

$$A_\lambda = \varepsilon_\lambda bC \tag{4.1}$$

where A is absorbance, ε is the molar absorption coefficient in units of L/(mol cm), b is the path length of the measurement in units of cm, and C is the concentration in units of mol/L. The index λ indicates that the

Lambert–Beer law is strictly valid only for monochromatic radiation. Equation 4.1 is actually a combination of two individual laws. The first one is Lambert's Law (which was originally discovered by Bouguer), relating the absorption to the path length, the second one is Beer's Law relating the amount of light absorbed by a sample to the concentration of some chemical species that is present in that sample. It should be noted that this only holds for sufficiently low concentrations. At higher concentrations, ε is no longer independent of C, for example because of mutual interactions of absorbing molecules with each other (aggregation, dissociation), with the solvent (degree of solvatation) or the like. The upper concentration limit for a linear $A(C)$ relationship must be individually determined for every new system. If the concentrations of interest are above this upper limit, a calibration curve rather than a straight calibration line may have to be used.

Spectra are sometimes recorded in units of transmittance (T) or percent transmittance ($\%T$), which do not scale in a linear manner with chemical concentration:

$$A = \log(1/T) \tag{4.2}$$

4.2.1.2 Reflectance

Spectra of solid samples are usually recorded in the units of reflectance (R) or percent reflectance ($\%R$), which is analogous to percent transmittance in that reflectance equals the ratio of the reflected radiation to the incident radiation. With diffuse reflectance, the reflected signal is attenuated by two phenomena: absorption (coefficient k) and scattering (coefficient s). Following the Kubelka–Munk theory, these two coefficients are related to the reflectance of an infinitely thick sample, R_∞, by

$$\frac{k}{s} = \frac{(1-R_\infty)^2}{2R_\infty} \tag{4.3}$$

Hence, if the scattering coefficient remains constant during a measurement, the absorption may be deduced from the reflectance.

4.2.1.3 Isosbestic Points

UV-vis spectroscopy typically yields broad, overlapping bands. In certain systems, this gives rise to an interesting feature: isosbestic points. Let A and B be two different substances, each with its own UV-vis absorption spectrum. At those wavelengths where the two single-component spectra intersect, any proportion of n A and $(1-n)$ B will yield a spectrum that again passes through these points (Figure 4.1). An example is a pH indicator in its protonated (acid) and its deprotonated (basic) form.

The concept of isosbestic points can prove helpful in various ways:

- If A is the starting material in a reaction and B is the product, the appearance of isosbestic points indicates that the reaction does not proceed via an intermediate with a different UV-vis spectrum.
- More importantly, the absence of well-defined isosbestic points indicates the reaction from A to B does proceed via an intermediate.
- If the absorption at an isosbestic point suddenly starts changing in the course of a reaction, this indicates an unexpected process. An example is given by Thurston *et al.*[8] The absorbances at two isosbestic points were constant during a clean, single-step reaction. By observing a sudden change in absorbance after some time, the crystallization of the product was detected.
- If the amount of A has to be followed, and A may convert to the form B with a different UV-vis spectrum, the absorbance at the isosbestic point (spectra of A and B) indicates the absolute amount (see e.g. Agarwal *et al.*[9] in Section 'Discoloration and degradation' below).

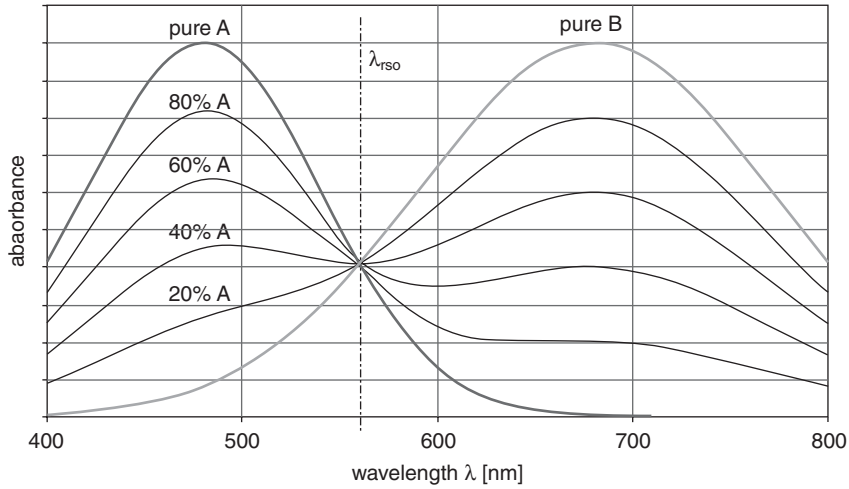


Figure 4.1 *An isosbestic point (schematic).*

On the other hand, the difference beween absorbances becomes dominated by noise close to the isosbestic points, which proved computationally challenging in a study by Thurston *et al.*[10] There, those areas were excluded from the computation.

4.2.1.4 Data Processing

Without going into great detail on any of the different algorithms (details on this can be found in Chapter 12), chemometric models can be divided into a few simple categories. Single-wavelength models simply track the magnitude of the absorbance at a chosen wavelength and correlate that to concentration by Beer's Law. This method works well if there is only one absorbing species present (at the analytical wavelength) and if there is a well-behaved baseline that does not introduce any offsets to the data. If baseline offsets are present but the baseline remains flat even when offset, then a simple model that correlates the difference in the absorbance values at the analytical wavelength from the baseline wavelength to the concentration of the analyte, again by Beer's Law, will work well. There must be no interfering absorptions at either the analytical or the baseline wavelengths. For more complex spectra other multivariate methods are required. These cases involve overlapping absorption bands from two or more components and/or sloping baselines in the data. More sophisticated methods of background subtraction include first and second derivatives (a comprehensive review was published by Sánchez Rojas and Bosch Ojeda[11]), as well as a multisegment background fit and subtraction. Chemometric techniques such as principal component regression (PCR), partial least squares (PLS), and multiple linear regression (MLR) are used to correlate spectral response to analyte concentration.

4.2.2 Color

Transmittance and reflectance data are used in color measurements. Transmittance spectra are used for liquid color measurements, while reflectance spectra are used on solid samples (powders, surfaces) and on opaque liquids (paint). A large number of color scales are in use today. Some are specific to a particular industry, while others enjoy broader application. The most common scales are the CIE L*a*b* (Commission

Internationale de L'éclairage, 1976) and the L*C*h. For L*a*b* coordinates, L* refers to the brightness of a sample, a* is the red–green coordinate, and b* is the yellow–blue coordinate. Regardless of the scale, spectral data from 380 to 720 nm are normally used for color calculations, as this is the wavelength range to which the human eye is sensitive. The only difference between the different color scales is in the wide variety of mathematical transforms that are then applied to that data to extract a particular useful measure of color. Whiteness index (ASTM E313-73)[12] and yellowness index (ASTM D1925-70)[13] are two more examples of different color scales.

4.2.3 Film thickness

The thickness of thin film layers separated by uniform, parallel interfaces can be determined from optical interference patterns that result. These measurements can be made from about 400 nm out through the visible spectrum and on into the near-infrared (NIR) region. Since film thickness measurements rely not on the absolute magnitude of the reflected light, but on the variation of that signal with wavelength, the choice of units is less important. Typically *%R* is used, but in some cases raw intensity is also satisfactory. We will treat thickness determinations in more detail in the applications section of this chapter.

4.2.4 Turbidity

Suspended particles may cause a baseline shift in the spectrum as outlined above. The shift may be evaluated after an algorithm proposed by Huber and Frost[14], yielding the turbidity (see also van den Broeke *et al.*[15]).

4.2.5 Plasmons/nanoparticles

While it has not yet fully entered industrial production, nanoparticle synthesis has spurred considerable interest in the capability of UV-vis-spectroscopy to characterize metal nanoparticle suspensions. The free electrons on the surface of metal nanoparticles may be excited by the incoming light, and the collective oscillatory behavior yields *localized surface plasmons* that can be discerned in the spectrum. In a study of the growth of gold nanoparticles, Salvati *et al.* measure the absorption of the plasmon peak which should scale with NR^3 where N is the number and R is the size of the nanoparticles.[16] A plot of the cubic root of absorbance versus time yielded straight lines for temperatures of 70 °C and above, indicating that nucleation is fast enough to be separated from the growth process, yielding a constant N over time. The nanoparticle size can hence be deduced from the plasmon peak absorbance. A recent review on theory and modelling of light interactions with metallic nanostructures by Montgomery *et al.* illustrates foundations and perspectives in a budding field: the industrial production of nanoparticles.[17]

4.3 Instrumentation

Instrumentation for UV-vis process analysis falls into four categories: scanning instruments, diode-array instruments, photometers, and fiber-optic spectrophotometers with photodiode-array (PDA) and charge-coupled device (CCD) detectors. The former two are more typically encountered in at-line or near-line applications, whereas the latter two are better suited to actual on-line analyses.

- Conventional scanning UV-vis spectrophotometers come in two configurations: single beam and double beam. In general, scanning UV-vis spectrophotometers offer superior stray light performance to an

array-type instrument. Their chief disadvantages for process analysis are the presence of moving parts and the difficulty in getting the sample to the instrument for analysis.

  - Single-beam instruments are found in at-line or near-line analyses where a sample vial or cuvette may be inserted into the instrument. While these instruments can collect data over a broad wavelength range, they are not capable of sophisticated data analysis and an analysis based upon one or two wavelengths is performed. A separate reference vial or cuvette is usually required. In most cases, a sample without any interference (bubbles, particles, multiple absorbing components) is required.
  - Double-beam instruments are more typically encountered in the laboratory environment. They offer the advantage of a separate reference channel and, in general, better optical performance than a single-beam instrument.

- Diode-array spectrophotometers have no moving parts, or at most a shutter that moves; thus they are more suited to locations near the process. However, these instruments still require the introduction of the sample into the instrument's sample cavity, as well as the necessity for a separate reference measurement. These instruments are usually only found in a single-beam configuration.
- Photometers are relatively simple devices that make use of a light source and one or more filters to present a narrow band of wavelengths to a sample and then to a photodetector. One, two, or even three wavelength analyses are possible with these devices. In addition, photometers are designed for location on-line, with samples generally plumbed into a sample cavity in the analyzer.
- Fiber-optic-coupled spectrophotometers (single beam and double beam) are the best choice for on-line analyses. The advent of nonsolarizing optical fiber has made possible on-line analyses in which the spectrophotometer may be located remotely from the process and light is carried to/from the process by the optical fiber. A rugged probe or flow cell provides the sample interface.

  Photodiode arrays are almost always used in single-beam instruments. PDAs provide high signal-to-noise ratio (S/N) in high-light applications such as absorbance/transmission measurements. They are less expensive than CCDs and come with 512, 1024, or even 2000 pixel arrays. CCDs are more costly than the PDA; however, they offer superior sensitivity and versatility to the PDA. With proper electronic processing, CCDs work well in both high- and low-light (emission, fluorescence) applications. They offer lower dark noise than PDAs and greater sensitivity over the entire UV-vis range. For use in the UV, the CCD must be either coated with a material that absorbs UV light and emits visible radiation, or used in the back-thinned configuration.

  Because a CCD is a two-dimensional array of pixels, it is possible to create a double-beam instrument with a single detector. Coupled in an optical design with an imaging grating, the output of two fibers may be imaged simultaneously and separately onto the CCD and both spectra readout by the electronics.

  Both PDA and CCD fiber-optic spectrophotometers may be combined with a fiber switch or multiplexer to access multiple probes or cells in a sequential fashion. Multiplexers are available to select from among 2, 6, 9, or even more fibers/probes.

4.4 Sample Interface

Design and selection of the sample interface is vital to provide the best-quality data for an analysis. The sample interface may be located in the sample cavity of a spectrophotometer, as in the cases of laboratory cuvettes, vials, and flow cells. The sample interface may also be fiber-coupled and located closer to the process. Fiber-optic sample interfaces include flow cells, insertion probes, and reflectance probes.

4.4.1 Cuvette/vial

In a few cases, analyses may be done on the process floor using a conventional scanning spectrophotometer and the usual laboratory cuvette or a specially made vial that fits into the sample cavity of the instrument. Cuvettes may be made from quartz (various grades depending upon UV transmission requirements), or from plastic (disposable) when UV wavelengths are not required for the measurements. In most cases these at-line analyzers have lower performance characteristics than dedicated laboratory instruments, but may be adequate for the task at hand. However, their use still requires manual sampling of the process, which itself brings on problems such as operator exposure to the process and sample integrity. In some automated applications such as drug dissolution, samples are pumped through a flow cell (cuvette) located in the instrument's sample cavity.

4.4.2 Flow cells

Flow cells are used in fiber-optic applications. They may be plumbed directly into a process, or onto a side-sampling loop. They typically have two opposing lenses with a sample gap in between. The flow path is usually the 'Z' configuration (Figure 4.2). The volumes of the flow cells may vary widely depending upon the sampling requirements.

Standard flow cells with a single-pass optical design have path lengths from 1 mm to 1 m for liquid-phase applications. Alignment of optics and optical throughput become problematic at greater lengths. These flow cells usually have flow path diameters from 1 mm up to several centimeters. Flow cells with small flow path diameters have very low cell volumes and are ideal for use in applications where there is very little sample available. Examples of these applications include high-performance liquid chromatography (HPLC), and various combinatorial chemistry techniques.

Short path length flow cells may be used in lower volume, or lower flow situations. The extremely small gap between the optics of fibers limits them to these types of applications. It is possible to make flow cells with sample gaps as small as 25 μm, so that they may be used with highly absorbing species. With even higher absorbing species, attenuated total internal reflection (ATR) probes may be appropriate (see below).

A filtered flow cell offers the advantage of removing particles or bubbles from the optical path. This makes reliable process measurements possible in difficult environments where particles and/or bubbles abound. A fermenter is one example of such an environment.

4.4.3 Insertion probe

When it is not possible to plumb a cell into the process, an insertion probe offers a second method of accessing the sample. These probes may be put in through a port in a pipe or vessel. It is also possible to design the probe port to allow for removal for cleaning while the process continues to run.

Transflection probes (see Figure 4.2) work by sending a beam of light into a sample gap from a collimating lens and then bouncing the light off a mirror at the other side. As such, these probes have a double-pass optical design. These probes may be manufactured with sample gaps from 0.5 mm to 5 cm. At gaps smaller than 0.5 mm (1-mm path length) it becomes difficult to reliably move sample through the probe's measuring cavity.

In a transmission insertion probe (see Figure 4.2) light makes only a single pass through the sample gap. In these probes either the fiber is forced into a 'U' bend at the end of the probe, or corner-cube reflectors must be used to turn the light path 180°. Again, the smallest practical gap is 0.5 mm (0.5-mm optical path).

Like the filtered flow cell, the filtered immersion probe (either transflection or transmission) may be used in sample environments in which bubbles or particles make the use of unfiltered samples impossible.

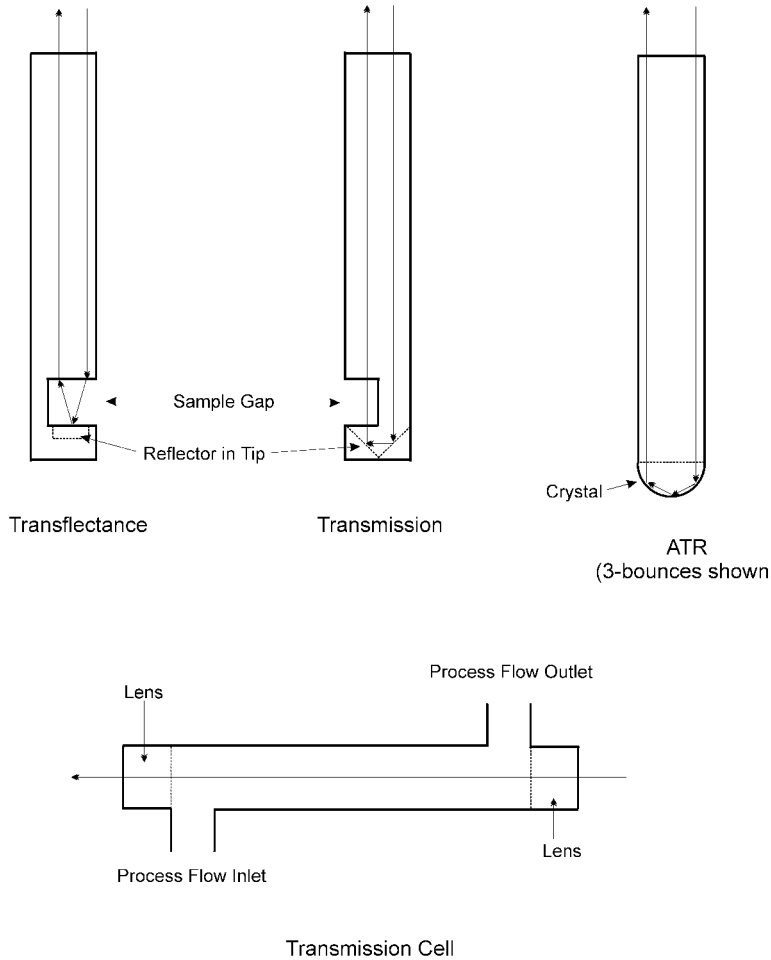


Figure 4.2 *Schematic of basic probe designs: transmission cell, immersion probe (retroreflecting and transmission), attenuated total internal reflection.*

ATR probes offer several advantages over other probe types. ATR is a phenomenon that relies on a difference in the index of refraction of a crystal and that of the solution with which it is in contact to prevent light from escaping the crystal. Only the evanescent wave of the light interacts with the solution layer at the crystal face. The result is an optical path length of only a few microns. Typical designs make use of faceted crystals or hemispheres (see Figure 4.2). The most common ATR material in the UV-vis is sapphire. In rare cases, fused silica may be used. ATR allows spectra to be taken of neat samples with optical density (OD) of 500–1000 or more. ATR is largely immune to particle or bubble effects in the solution. A reference spectrum of an ATR probe is also easily obtained, as a clean probe in air may serve as the reference: once a vessel is flushed, it is not necessary to fill it with a reference solution in order to re-reference the probe. The major disadvantage to ATR is that quantitation of the detected absorption is complicated by the fact that the strength of the absorption band will vary with the refractive index of the sample. The closer the index of refraction of the sample is to that of the crystal (cutoff index) the larger the absorption peak (longer effective path length), and more quickly the absorption cross section will change for a given change in

refractive index. The refractive index may change because of concentration changes or because of temperature changes in the sample. ATR probes make ideal insertion probes because there is no mirror or sample gap that would require cleaning. So long as the crystal is rinsed clean, the probe is ready for use. ATR probes are also generally immune to the effects of bubbles and particles in the sample.

4.4.4 Reflectance probe

A diffuse reflection probe is used to measure the light reflected from a solid surface or powder. This probe is immune to specular reflections from the sample, and detects only diffusely reflected light.

As differentiated from diffuse reflection, we use the term 'backscatter' to mean a probe that detects both specular and diffuse light. The most common designs of these probes incorporate a number of illumination fibers (usually six) in a ring around a central detection fiber. This probe is useful in a slurry or solution high in particulates (e.g. crystallization).

Sometimes, a probe must be shielded by a protecting glass (or quartz) tube because it is not catalytically inert with respect to the reaction mixture, or does not withstand its corrosive influence. In that case, it must be ascertained that the shielding does not falsify the signal, e.g. by specular reflection on the glass.

4.5 Implementation

4.5.1 A complete process analyzer

We have mentioned many of the components of a complete UV-vis process analyzer system, but have not put them together as a single installation to this point. An example of such a system based upon fiber optics is shown in Figure 4.3. The UV-vis analyzer contains a spectrophotometer and a light source. A fiber multiplexer may be added if multiple sampling points are required. Modern analyzers also contain a control computer, display and some kind of data I/O interface to the process. These interfaces can include analog current outputs, digital alarm signals, or numerical streams sent to the plant's distributed control system (DCS). One or more probes are placed into the process. If the measurements made by the analyzer indicate that the process is drifting towards the control limits, the DCS may now take some action to change the operation of the process and bring it back to the optimal range. For example, the amount of reactant being fed may be adjusted, the temperature may be changed, or mixing speeds may be altered to maintain the process control.

4.5.2 Troubleshooting

During the implementation, numerous problems may arise. Many of those can be readily solved in a straightforward manner. A few issues will be listed here.

4.5.2.1 Baseline Shifts

While baseline offsets are usually not a problem in a laboratory cuvette, they can be a significant issue in on-line measurements. There are several sources of such offsets.

- *Bubbles* scatter light in a sample and induce baseline offsets. Scattering has a greater effect the farther into the UV region (shorter wavelengths) one goes. If too much light is scattered, not enough light will be transmitted through the sample and a reliable measurement cannot be made. If possible, bubbles in

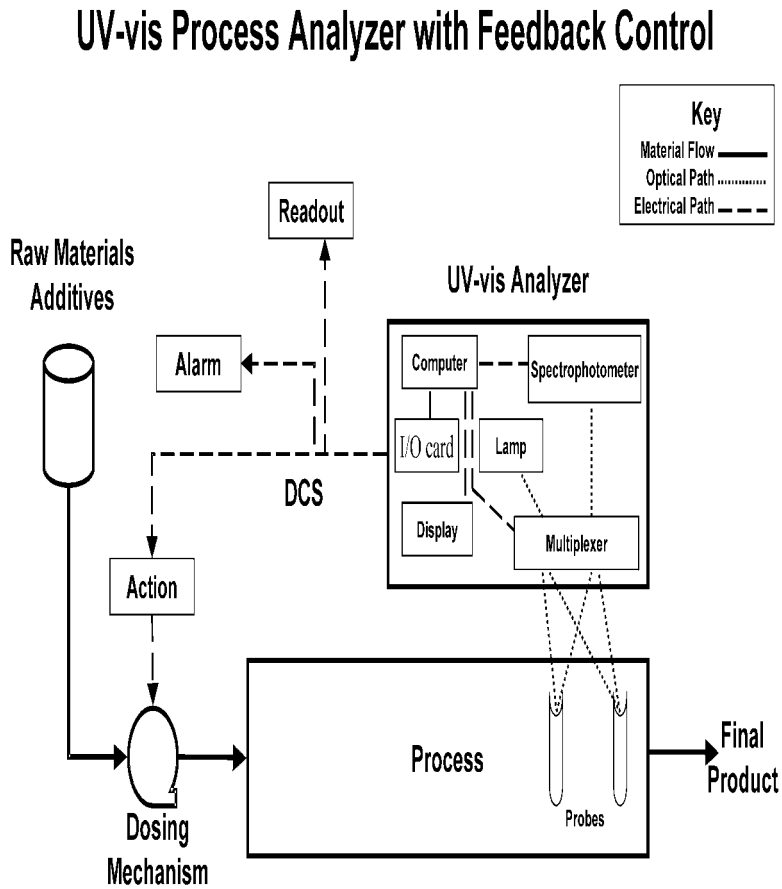


Figure 4.3 *Example of an on-line process analysis.*

the optical cavity of a probe or cell should be minimized, e.g. by sparging the liquid with an inert gas like argon.

- *Particles* also cause the scattering of light in a process sample. Filtering probes and cells are available, which can eliminate most particles and bubbles from the optical path and make reliable process measurements possible for difficult sample matrices (see e.g. Section 4.7). If particle formation is inherent, like insoluble excipients in dissolution studies of tablets, an ATR UV probe may be considered.
- In some cases, the sample will deposit a *film* onto the optical surfaces of a probe or cell. This film might lead to a bias in the analytical results, and eventually could reduce light transmission to make the analysis impossible. Film buildup requires some action be taken to clean the cell or the probe, either cleaning in place or by removing the probe or cell for cleaning by hand.
- The *movement of a fiber-optic probe* and/or the fibers leading to/from a probe or cell can also induce baseline shifts in the recorded spectra. These shifts are usually minor and can be eliminated with proper preprocessing of the data (e.g. derivatives). (A detailed description of preprocessing techniques can be found in Chapter 12)
- The *reference* for the measurements is not always stable. In the case study 'Extruded plastics color' (Section 4.7), the white reference tile was checked about once a week for accuracy.

4.5.2.2 Fiber Breakage

While UV-vis fibers are much less prone to fiber breakage than mid-IR fibers, excessive mechanical stress may still break the fibers. The minimum bending radius as specified by the supplier should be observed, and excessive mechanical stress should be avoided.

4.5.2.3 Nonlinearity

As was pointed out in Section 4.2.1, the plot of absorbance versus concentration becomes less than linear beyond a threshold concentration due to non-negligible interactions between absorbing molecules. The baseline shifts discussed above may equally give rise to deviations from Beer's Law. Upstone discusses fluorescence (which may be removed by inserting a cutoff filter between sample and detector) and peak shifts due to changes in pH.[7]

Note that not only Bouguer–Lambert's law, but also Beer's law can have an upper limit for linearity. If the path length becomes longer and longer, but the light beam is cone-shaped rather than fully focused, the receiving end may no longer capture all of the light. A plot of absorbance vs. path length will then display a negative deviation from a straight line at longer path lengths. This should be checked carefully (see, for example Liu *et al.*[18]).

4.5.2.4 Stray Light

If the probe is not properly shielded from ambient light, spurious effects can interfere with the measurement. Upstone defines stray light as the amount of light present in the instrument at nonanalytical wavelengths or, put another way, the amount of unwanted light present in the system.[7] He points out that it can come from outside entering the optical system (which should be minimal in well-designed instruments) and from the monochromator. Peaks that resemble 'negative absorption signals' may indicate stray light from neon bulbs, which have a discrete spectrum. Johansson *et al.* report on a problem with stray light that may have been caused by problems in the spectrometer/CCD coupling.[19] They were able to partially correct this with a numerical approach. Liu *et al*. (2008)[18] have checked for the influence of stray light with an immersion probe by varying the path length. There are standard tests for stray light (e.g. ASTM E 387-2004).[20]

4.5.2.5 Material Issues

In a process environment, the probes must neither corrode nor display catalytic or inhibitory properties. Quinn *et al.* report that about 30 reactions were run and five different types of metal probes were utilized with a highly corrosive reaction mixture before a custom built PEEK (polyether ether ketone) probe yielded successful *in situ* measurements.[21]

4.6 Applications

UV-vis technology finds use in a wide variety of applications that cut through a cross-section of industry. In this section, an overview on the state of the art is given, ranging:

- from recent laboratory studies that have yet to make their way into industrial applications
- to examples that have been implemented and working for years in actual industrial settings.

Selected industrial case studies are described in more detail in Section 4.7.

4.6.1 Gas and vapor analysis

A number of important pollutant gases and vapors have signatures in the UV region of the electromagnetic spectrum. Among these are NO_x, SO_x, CO, O_3, benzene, toluene, and the xylenes. UV-vis spectroscopy of a host of environmentally relevant gases and vapors is reported by Bosch Ojeda and Sánchez Rojas.[4] In this chapter, two cases are presented: toluene (see Section 4.7) and ozone.

Ozone (O_3) is a triatomic allotrope of elementary oxygen (O_2). It is metastable and reactive with many functional groups. In industry, ozone is used for removing organic pollutants from wastewater, but also for the oxidative cleavage of olefins to carbonyl components (like the ozonolysis of oleic acid to pelargonic acid and azelaic acid). O'Keeffe *et al.* have presented an optical fiber-based UV-vis absorption spectroscopy system for monitoring ozone concentrations.[22] Given that ozone is used in harsh environments and is, by ways of the released oxygen radicals, an aggressive medium itself, the optical fibers as dielectric material are less prone to an attack than metal probes. The possibility for a remote detection via sufficiently long fibers is interesting. When going from UV (254 nm) to the vis region (550–650 nm), working with polymethylmethacrylate-based plastic optical fibers becomes possible. These are low cost sensors, mechanically more sturdy than glass fibers, and have larger diameters (1 mm core), which is beneficial for practical reasons. Visible light has less energy than UV light, so less ozone decomposition is triggered by the measurement. With a combination of UV and visible, a wide range of ozone concentrations can be covered.

4.6.2 Liquid analysis

4.6.2.1 Breakthrough Curves and Alarm Parameters

The power of on-line technology is the ability to monitor a critical point in the process for an event of some kind that intermittent manual sampling and off-line analysis would miss. One such event is column breakthrough on an ion-exchange column, where, for example, recovery of a catalyst may have both economic and environmental impacts. In terms of plant efficiency, on the one hand, it is desirable to use the maximum capacity of the column and then stop or divert the flow only when column breakthrough is detected. On the other hand, it is also desirable to minimize the amount of breakthrough if it has an impact on permitted plant emissions. A practical example, nickel breakthrough on an ion-exchange column, is presented in Section 4.7.

Along the same line, UV-vis based scanning for deviations from a blank reference in industrial wastewaters can indicate problems on an upstream level that remain undetected using conventional parameters.[23] The so-called alarm parameters need not even be based on detailed interpretation of all absorption bands. In a coarse manner, absorption in certain spectral regions may correlate with lump parameters such as total organic carbon (TOC), dissolved organic carbon (DOC) and the like.[15]

4.6.2.2 Vessel cleaning

An excellent application of fiber-optic UV-vis spectroscopy is in the monitoring of a tank or vessel during a cleaning in place operation. This is important in industries where product quality and purity is tantamount: pharmaceuticals, foods, and beverages. Once a batch reaction is completed and the vessel drained, it must be cleaned before the next batch of product can be made. This next batch may be the same product, or a different one. In either case, it is vital that the cleanliness of the vessel be verified. If an off-line analysis is to be used, the cleaning process is delayed while laboratory results are obtained. In other cases, a tank may be flushed for a length of time that has been determined to be satisfactory through trial and error over time at the plant. In fact, much less cleaning may be necessary and the ingrained method in use is overkill – wasting production time and money. An on-line analyzer can pinpoint when the vessel is adequately clean

and may be able to decrease turnaround times for the next reaction batch. Fortunately, in the pharmaceutical and food/beverage industries, many products possess strong UV or visible signatures and the reaction solvent is likely to be transparent to the UV-vis analyzer. It is not even necessary to calibrate for a particular analyte, so long as the decay of a signal peak may be followed over time as the tank is washed.

Fiber-optic UV-vis spectrophotometers are well suited to a clean-in-place (CIP) application. For analytes with signals in the deep UV, best results will be obtained with a deuterium source and short fiber runs. Further out in the UV and visible regions, a xenon flash lamp is the best choice and fiber runs may be much longer. If fiber runs must be kept short, then either the sample must be plumbed into a cell located at or inside the analyzer enclosure or the analyzer must be located within a few meters of the flow cell or probe mounted in the process piping. Because this application relies on following the degradation of a signal to determine cleanliness, it is important to attain the lowest detection limit possible. This may be achieved by improving the S/N through longer-path flow cells and probes (5 cm and up), signal averaging, and using a low-noise detector.

One of the difficulties in quantitative spectroscopic analysis in any part of the electromagnetic spectrum is that a separate calibration model must be maintained for nearly every different analyte. However, in the case of CIP applications this is not necessarily true. Since the goal of this measurement is to verify when no signal is present – that is when the sample spectrum matches the spectrum of the neat solvent(s) – the model may be a simple wavelength type (peak minus background) and the signal peak selected automatically by a peak search algorithm. Alternatively, the integrated area under a peak might be tracked. Another means would be to simply compare the stored solvent spectrum to each sample spectrum, calculate the residual and continue cleaning and making measurements until it reaches a preset minimum level.

4.6.2.3 Liquid Mixing

Ng and Assirelli have carried out a mixing study in batch stirred vessels with working volumes of 3 L to 20 L using a fiber-optic UV-vis monitoring technique.[24] Bromophenol blue sodium salt was used as a non-reactive tracer. The results on traditional Rushton turbines and 45° angled pitched blade turbines showed good agreement with a typical conductivity technique and a correlation proposed in literature.

4.6.2.4 Dissolution Testing

Dissolution testing is a mandated step in the production of all tablets and extended release products in the pharmaceutical industry. There are well-established United States Pharmacopeia (USP) guidelines for performing all dissolution testing. Although none of them are written specifically for fiber-optic methods, they do make allowances for new technologies that are proven to be of equal or greater sensitivity or accuracy than the existing methods.[25] There have been numerous studies in pharmaceutical companies to assess this approach.[26–28] While at a first glance, a lab-based product characterization method appears outside of the scope of a book on process analytics, the aspects and issues are the same, and much can be learned from the thorough studies mentioned.

A dissolution testing apparatus consists of a set of six or eight thermostatted, stirred vessels of an established geometry and volume from the USP guidelines. The dissolution apparatus provides a means to dissolve each sample, but does not provide a means to determine the concentration of the active ingredient in the bath. In the most well-established scheme, sipper tubes withdraw samples from each dissolution vessel and send them through a multiport valve to a flow cell sitting in the sample chamber of a UV-vis spectrophotometer. In recent years, moves have been made to make *in situ* measurements in the dissolution baths by means of fiber-optic probes. There are three possible probe implementations: *in situ*, down shaft, and removable *in situ* (see Table 4.2).

Table 4.2 *Probe implementations in fiber-optical dissolution apparatuses*

	Probe implementation	Advantages	Disadvantages
In situ probe	Slender probe permanently inserted into each dissolution bath in an apparatus	Measurement at same location where the sipper tube had been positioned	Disturbance of the laminar flow in the stirred vessel. Lean probe desirable
Down shaft	Probe positioned down the hollow shaft of the stirring blade in each vessel	Eliminates the problem of flow disturbance at the sampling point	Sampling depth not as USP-prescribed; flow effectiveness through shaft ill-determined
removable *in situ*	Probe dipped into the vessel only for measurement. Once the data have been collected, the probe is withdrawn	Minimize flow disturbance in the vessel, still allow sampling at the same location as the sipper tube	Film may form on the drying optical surfaces of the probe while it is suspended in air over the dissolution bath

For a simultaneous measurement of several dissolution samples, different approaches can be taken that have to trade off between high cost and high crosstalk between neighboring fibers. If numerous probes are coupled to just one common photometer, the setup is less expensive than one where each probe has its separate CCD spectrophotometer, but it suffers from more crosstalk, that is, spilling-over of light from one fiber onto the pixels designated to receive light from a neighboring fiber. This effect can be minimized by careful design of the grating and choice of pixels on the CCD itself. With a mechanical fiber multiplexer (fiber switch) to allow one spectrophotometer to interrogate sequentially each of the probes, there is no crosstalk between the different fiber probes, and just a single light source and detector in use for all of them. The key to success of this design is that the mechanical multiplexer must be both very accurate and precise in its repeated movements to each probe.

Gemperline *et al.* determined the multicomponent dissolution profiles of pharmaceutical products by *in situ* fiber-optic UV measurements.[29] UV-vis spectroscopy yielded the same concentrations in the second half of a dissolution run as an off-line HPLC, but lower concentrations in the first half. It was conjectured that suspended particles were not accurately measured by UV-vis, but dissolved during their transfer to the HPLC, hence leading to an overestimation of the values.

In a recent study, Mirza *et al.* have pointed out that despite the obvious advantages, fiber-based dissolution studies are not yet implemented in industry.[28] They have validated a fiber-optic based dissolution test for linearity, precision, accuracy, specificity and robustness, with excellent results.

4.6.2.5 Discoloration and Degradation

Industrial wastewater may be colored and color removal may be required to comply with regulation. Coque *et al.* describe how various discoloration methods were studied with standard azo dyes solutions and an industrial wastewater from the paper industry as model systems.[30] The liquids were recirculated between the photochemical discoloration reactor and an at-line spectrophotometer (Suprasil quartz cell, path length 2 or 10 mm, resp.). A discoloration yield τ was calculated in CIE L*a*b* space, which permitted monitoring the process and determining completion in an easily implemented manner.

In the pulp and paper industry, UV-vis spectroscopy proves helpful in answering many related questions (e.g. around yellowing of pulp),[31] Kappa number monitoring,[32] alkyl xanthate decomposition,[33] etc.

Interestingly, the discoloration of dyes shows a good correlation with the degradation of aromatic pollutants. Agarwal *et al.* have for example, followed the degradation kinetics of calmagite, a commercially available dye, as a model substance for the abatement of harmful pollutants.[9] They used a diode array spectrophotometer with a flow-through cuvette of 1 mm path length (Shimadzu UV-1650 PC). The UV-vis spectrum of calmagite is dependent on the pH, which renders the absorbance dependent on both concentration and prevailing pH. A variation of the pH revealed an isosbestic point at 566 nm, however. The absorbance at that wavelength could hence be used to determine the concentration independently of the pH.

4.6.2.6 Reaction Monitoring

Inline UV-vis spectroscopy may certainly be used for monitoring the course of liquid-phase reactions. The concentrations of all UV-vis active components may be followed by a proper insertion probe.

Quinn *et al.* followed the reaction of a nucleoside with trityl chloride in pyridine at 50 °C on the laboratory scale.[21] This reaction is the first step of an industrially significant process. The UV-vis spectra were analyzed with chemometric analysis where automatic window factor analysis (WFA) yielded better results than PLS. A reactive intermediate, a solvated carbocation, was identified that had not been found with HPLC (quenching upon aliquot dilution) or NIR, respectively. Small, sudden process upsets could be detected.

Gurden *et al.* studied the monitoring of batch processes using spectroscopy.[34] As a case study, they followed a pseudo first-order biochemical reaction with an intermediate using UV-vis spectroscopy. Following statistical process monitoring, process disturbances could be detected in a series of batches.

A comparison of different analytical methods has been presented for an esterification of crotonic acid and butan-2-ol as model system.[35] The reaction was followed with in-line NIR, Raman and UV-visible spectrometries, as well as with at-line low-field nuclear magnetic resonance (NMR) spectrometry. The Raman und UV-visible data were analyzed as first-derivative spectra and evaluated with univariate calibration models, NMR and NIR with multivariate analysis by PLS regression. UV-visible spectroscopy gave the poorest between-run precision of about ±50% relative error. Interestingly, using PLS regression to develop multivariate calibration models did not yield better results. The authors explain that the solvent (toluene) and the reaction product (2-butyl crotonate) were the only absorbing species. Since only the latter changed in the course of the reaction, the univariate method minimizes the influence of noise for this component alone. A PLS regression may try to model some of the noise if too many factors are used, thereby introducing errors. It is pointed out that the relatively poor performance of the univariate approach may however be due to the selection of a wavelength where the signal displays a steep change.

The extraction of kinetic parameters from in-line UV-vis spectroscopy may suffer from three sources of error: namely, instrumental noise, error in determining initial concentrations, and error in the calibration of pure standards, as is pointed out by Carvalho *et al.*[36] These authors have studied the influence of these errors on the determined rate constants for a simulated second-order reaction and compared twelve methods to cope with the errors in five groups. They find significant differences in accuracy and precision.

For liquid-phase reactions it is known that they may be enhanced by applying higher pressures if the activation volumes are strongly negative, for example, with Diels–Alder reactions. At pressures up to 600 bar (6×10^7 Pa), Benito-Lopez *et al.* used fiber-based on-line UV-vis spectroscopy to monitor the nucleophilic aromatic substitution of 1-fluoro-4-nitrobenzene with a ten-fold excess of pyrrolidine.[37]

During crystallization procedures, undesirable 'oiling-out' (i.e. liquid–liquid demixing) may occur. A UV-vis study of such systems with highly dispersed vesicles, or crystals, in transmission mode will capture

the increase of turbidity as described above. With an ATR–UV-vis probe, however, the detected signal stems from a very thin layer close to the probe surface. The thickness of this layer is of the order of the wavelength of the radiation – a few hundred nanometers. Hence, the spectra reflect the state of the continuous phase, the solvent, and the solutes, but not the vesicles or crystals, respectively. Deneau and Steele have combined the ATR-UV-vis technique with focussed beam reflectance measurements (FBRM) and particle vision monitor (PVM) to elucidate the phenomena in a pharmaceutical crystallization procedure suffering from oiling-out, and to develop conditions where this effect was eliminated.[38]

The challenge of monitoring batch processes for deviations was addressed by Gabrielsson *et al.*[39] (also taken up by Wong *et al.*[40]) with the hydrogenation of nitrobenzene to aniline as a model reaction. The study followed a factorial design of experiments. Process deviations could only be detected when the UV spectroscopic data set and the process data set were used in a combined manner, not seperately.

4.6.2.7 Distillation

At Merck KGaA in Darmstadt (Germany), in-line UV spectroscopy has been used for monitoring a distillation setup for solvent purification.[41] A transmission probe was implemented at the top of the column. Solarization-resistant UV fibers guided the light to a diode array spectrometer. From the spectra, a quality parameter was extracted and fed to the process control system (PCS). As an example, the quality parameter would exhibit a transient behavior upon column startup. Below a defined threshold value of the parameter, the PCS would assume sufficient product quality and switch the exit stream from waste to product collection. The operation of the spectrometer does not take place in a direct manner, but rather via the PCS.

4.6.2.8 Bioprocesses

In reviews on the use of *in situ* sensors[42] or optical sensor systems[43] for bioprocesses, UV-vis does not play a major role. An example for the application of at-line UV-vis spectroscopy was presented by Noui *et al.*[44] The selective flocculation processes of *Saccharomyces cerevisiae* homogenates were followed with a newly developed direct UV spectrophotometer. The results from a PLS regression model were in good agreement with those from off-line chemical assays.

4.6.2.9 Polymer Melts

Foulk *et al.* have assessed the task to determine the stabilizer content in polyolefin melts in a fast and affordable manner in a systematic way.[45] The stabilizer molecules (e.g. phenols) display characteristic UV absorptions. The polyolefin melt was transported from the extruder through a Swagelok cross to the coolers and pelletizers. Two fiber-optic cables (10 m each) were attached to the cross such that the light was conducted from the light source, passed through the melt in transmission mode, and was then guided to the UOP Guided Wave Model 300 analyzer. The information was processed with the spectrometer software and then with the UNSCRAMBLER multivariate calibration software. Very promising results were obtained every 30 seconds in a nondisturbing manner. This is an example where chemometric analysis allows for modeling the UV-vis data even though it is highly overlapped.

4.6.3 Solid analysis

Dry-solid particles and powders may have different reflectivities or colors, respectively. In this case, nondestructive sampling is possible by measuring the intensity of the reflected light. This may be used in an industrial environment in various ways:

4.6.3.1 Degree of Solids Mixing

Blending of powders is a unit operation that is used in a wide range of the chemical and pharmaceutical industries.[46] The degree of mixing of particulate solids with different reflectivities may be taken from the reflectivity of the mixtures. This was studied by J.B. Gray at Dupont in 1957[47] and taken up by Ashton *et al.* in 1966[48] in order to overcome removing samples for off-line analysis. More recently, Weinekötter and Reh have used laser light (at 514 nm) to detect tracers like black SiC or green Irgalite with fiber-optic probes to assess the performance of different mixers with white $Al(OH)_3$ as the main component.[49] It should be noted that NIR has become more frequently used in these kind of studies, particularly in the pharmaceutical industry.

4.6.3.2 Slurry Synthesis – VPO Catalyst

The solids analysis described above can be taken to yet another level by correlating the color measurement to chemical properties. An excellent model system is vanadium pyrophosphate (VPO), which is a well-known catalyst for butane oxidation to maleic anhydride. During the synthesis of the catalyst precursor, solid V_2O_5 particles are dispersed in a mixture of benzyl alcohol and i-butanol. In this slurry phase, the vanadium is partly reduced. Addition of phosphoric acid leads to a further reduction and the formation of the VPO structure. With a diffuse reflectance (DR) UV-vis probe by Fiberguide Ind., the surface of the suspended solid particles could be monitored during this slurry reaction.[50] Four points can be noted from Figure 4.4:

- The oxidation state deduced from the DR spectra (open symbols) agrees well with the oxidation state determined by off-line reference titration experiments (full symbols).
- The DR spectra yielded, in a noninvasive manner, a significantly better temporal resolution than the off-line titration.
- A slight increase in oxidation state is detected by DR UV-vis spectroscopy after 1500 min. However, it was not detected when the experiment was repeated under nitrogen rather than in air. Oxygen is likely to dissolve in the solvent, oxidizing the VPO particles' surface, which is not captured by the bulk method titration.
- About 10% of the V^{5+} has been (irreversibly) reduced to V^{3+} (Figure 4.4b).

A more recent example for UV-vis spectroscopy in a slurry of particles was reported by Stemmet *et al.*[51] With insoluble pH indicators deposited on the particles, the UV-vis spectra obtained from the moving particles could be used to characterize the state of the pH indicator, the transition point and the concentration of particles.

4.6.3.3 Reaction Monitoring – Heterogeneous Catalysts Under Working Conditions

Due to the sturdiness of UV-vis backscatter probes, heterogeneous catalysts may be monitored under working conditions. Coke deposition can readily be followed by UV-vis spectroscopy. Transition metal redox catalysts often display color changes during the reaction. The UV-vis spectrum of a vanadium pyrophosphate (VPO) catalyst was monitored during butane oxidation to maleic anhydride on a laboratory scale.[50] The backscattering probe supplied by Fiberguide Industries Inc. (Ø = 1/16′) was led into the catalyst fixed bed and collected spectra at 400–460 °C. The spectra could be correlated with the oxidation state of the vanadium under working conditions. Similar monitoring was reported by Philipps *et al.* during the V-Ti-O catalyzed oxidation of isoprene to citraconic anhydride.[52] More examples are given in reviews by Weckhuysen[5] and Jentoft.[6]

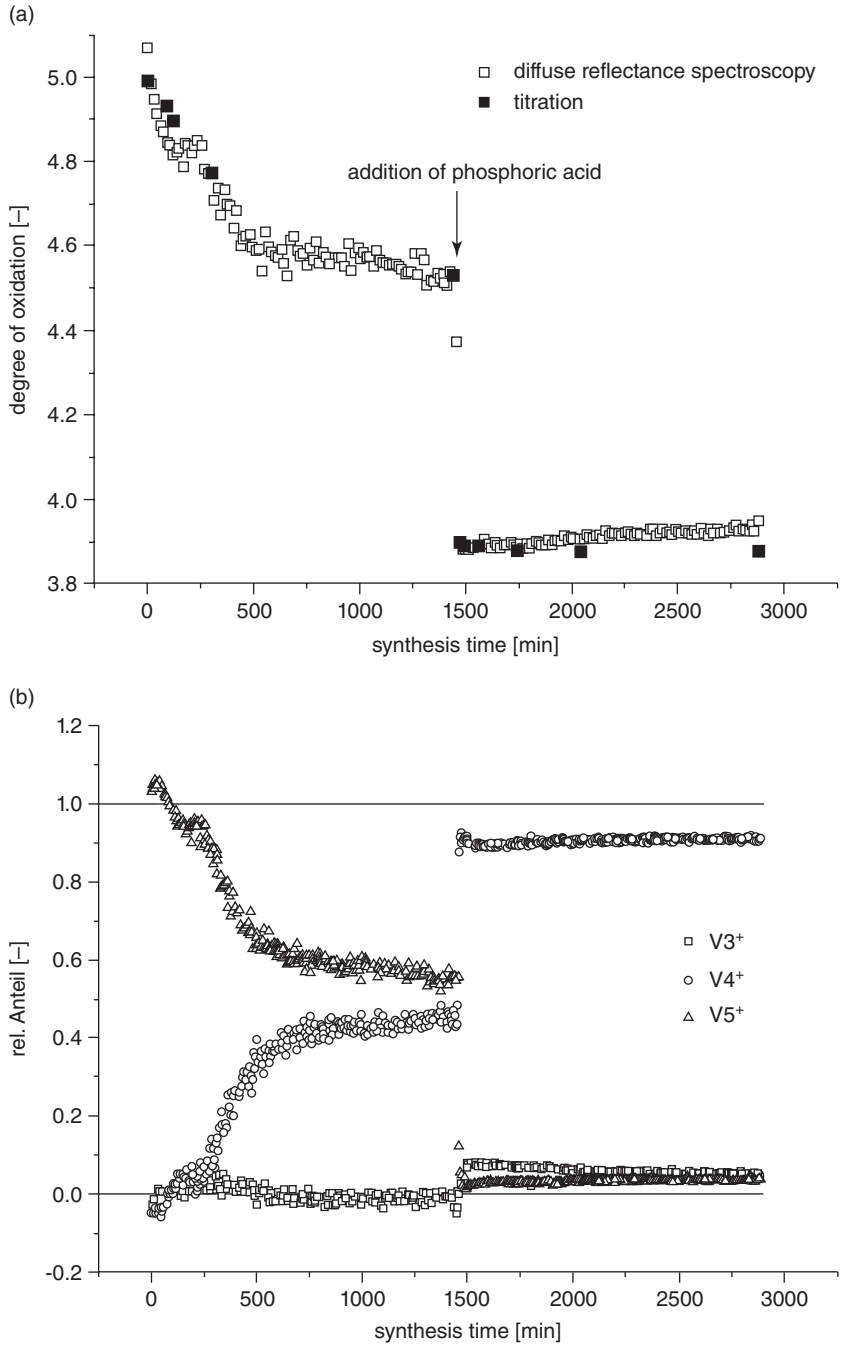


Figure 4.4 *DR UV vis during a suspension VPO synthesis (adapted from Weiß[50]): (a) average oxidation state versus synthesis time; (b) proportion of three-, four- and five-valent vanadium versus synthesis time.*

4.6.4 Other applications

4.6.4.1 Film Thickness

With the appropriate fiber-optic probe and data processing techniques, UV-vis spectroscopy may be used to determine the optical thickness of a transparent thin film. It is possible to simultaneously measure thickness of different layers in a multilayer structure as long as each layer falls within the analysis range of the instrument. Typically, this means layers in the 0.5–150 μm range. A further constraint on this technique is that the layer structure of the film must be smooth on the scale of the spot size of the fiber-optic probe. Neighboring layers must have different indices of refraction in order for them to appear as distinct layers to the analyzer.

The preferred light source for this analysis is the tungsten-halogen lamp, for it offers a smooth emission envelope and high stability. The highly structured nature of a xenon source's emission spectrum makes it unsuitable for film thickness measurements. White light from the fiber-optic probe strikes the sample, and reflected light is collected from each layer interface in the film. For example, a two-layer film returns signals representing each individual layer as well as the total film thickness (assuming all are within the range of the instrument) (Figure 4.5).

Depending upon the particular sample, all or a portion of the spectrum in the 400–1000 nm wavelength range may be used. In contrast to a color measurement or a chemical concentration measurement, maximum

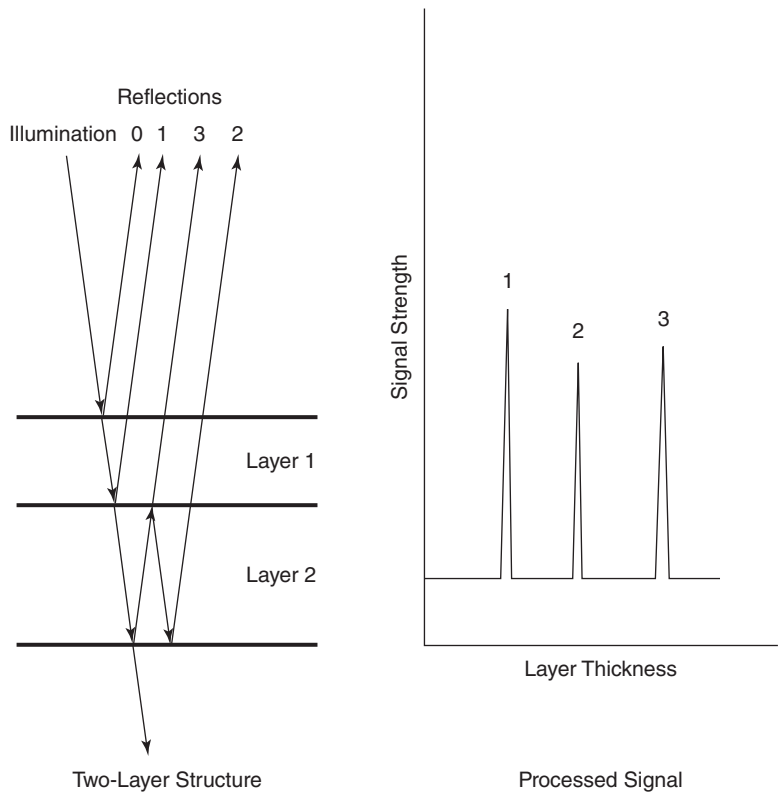


Figure 4.5 *Diagram of reflections from a two-layer film.*

light return from the sample does not guarantee the best results. It is better to optimize the probe placement based upon the ringing pattern in the data, rather than upon the reflected light intensity. The signal returned from the sample represents the optical thickness of the sample's layer or layers. This optical thickness may be converted to a true physical thickness by dividing the result by the index of refraction of the layer in question. The case study 'Film thickness determination: polymer' in Section 4.7 goes into more detail.

4.7 Detailed Application Notes

The examples below are based on experiences made when adapting the process analyzer presented in section 4.5 (Equispec Chemical Process Analyzer, Equitech Int'l Corporation, New Ellenton, SC, USA) to industrial problems.

4.7.1 Gas and vapor analysis: toluene

UV-vis spectroscopy was used to monitor toluene in a chemical manufacturing plant (Dr William Napper, Ciba Specialty Chemicals, McIntosh, AL, USA) for the purpose of preventing the formation of an explosive mixture of toluene and air. This is a critical safety analysis for the plant and must be accurate and reliable at all times. In this particular case, sulfur compounds in the gas stream acted to poison electrochemical sensors, so the decision was made to use a UV analyzer. The analyzer consisted of a fiber-optic UV-vis CCD spectrophotometer, a xenon flash lamp, fiber multiplexer, three 100-cm transmission cells and a control computer. The goal of the analysis was to keep the toluene level below the lower explosive limit (LEL) of 12 000 ppm. The alarm level for the analyzer was set to 60% of the LEL. Typical toluene concentrations were in the 1000–2500-ppm range. As toluene concentrations rose, nitrogen was bled into the system to keep the mixture inert. If the UV-vis measurements indicated that the alarm level was reached (~7000 ppm), the process went into shutdown. Some key points of this installation are that all the electrical components of the analyzer were kept in a general-purpose area, and that transmission cells were heated to 45 °C to prevent condensation on the optical surfaces inside. The sample stream was at ambient temperature, so prior to entry into the cell, the sample was sent through a 100-mL 'knock out pot' to catch any droplets in the gas stream, and then through a 5-μm filter to remove any particulates. The careful preparation of the sample stream for the gas cells was one of the keys to the success of this application. The sample was drawn into the cell by a vacuum venturi; the negative pressure in the process was created by a blower system on the carbon bed that was used to collect the toluene.

The calibration spectra from one of the gas cells are shown in Figure 4.6. Duplicate samples at the same concentrations have been omitted for clarity. These spectra have been preprocessed by subtracting the baseline offset (average response over the 360–370-nm range) from each spectrum. A second-order MLR (quadratic) fit to the data at 266 nm gave satisfactory results, with a correlation coefficient of 0.99875 and a standard error of 11.4%. While the standard error of the fit may seem high, it was adequate for the purposes of this application. Extremely accurate toluene concentrations were not required since the alarm level for process shutdown was set at 60% of the LEL, well outside of the error bars on the high-concentration calibration points.

This instrument ran very reliably while sampling every 30 seconds for 4 years from installation until the process was discontinued, with the need for only one lamp replacement and one replacement of the CPU cooling fan. Keeping the downtime to a minimum was essential, as this was a safety analysis; if the analyzer was down, the process did not run.

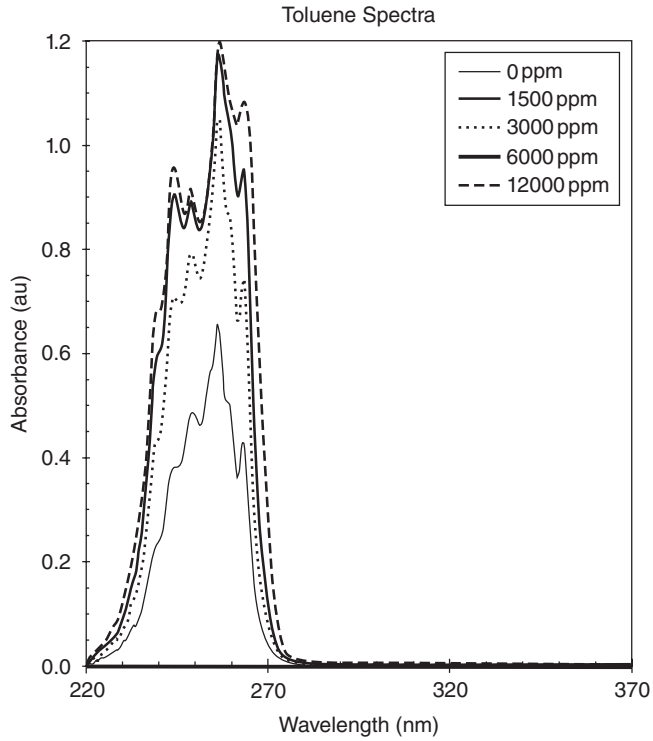


Figure 4.6 *Toluene calibration spectra.*

4.7.2 Liquid analysis: breakthrough curves

The process analyzer has been successfully implemented for the detection of nickel breakthrough on an ion-exchange column. Once breakthrough is detected in this particular installation, process flow is diverted into a second ion-exchange column while the first one is being regenerated. In this case, the process analyzer consists of a fiber-optic UV-vis CCD spectrophotometer, a xenon flash lamp, a long-path high-flow transmission cell and a control computer. The analyzer runs in a two-channel (dual beam) configuration in which a reference fiber loop serves as real-time compensation for lamp intensity fluctuations. The nickel peak has a wavelength shift due to changes in the pH of the process flow, so a chemometric model is used to look for the nickel peak within a 'window' along the wavelength axis. A first-order MLR model was built based upon the area of the nickel peak, using baseline correction due to the potential for bubbles in the process flow. The larger inner diameter of this flow cell (5 cm) allows for the presence of some bubbles without negatively affecting light transmission through the cell. Nickel is being monitored in the range of 0–20 ppm using a 50-cm path length transmission cell. Figure 4.7 is a plot showing an example of breakthrough on the column, followed by switch-over to the backup column.

4.7.3 Solids analysis: extruded plastic color

Proper control of the color of extruded plastic is a key quality issue for many consumer products, including vinyl siding, fencing and decking materials and flooring. Traditionally, samples of the product are measured

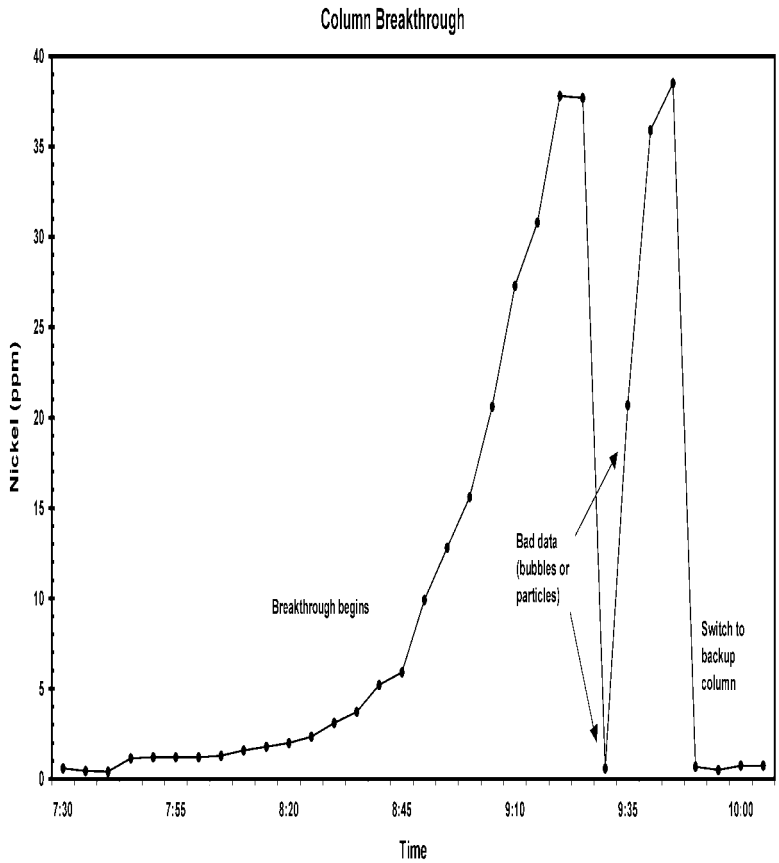


Figure 4.7 *Data showing nickel breakthrough.*

by off-line analysis which can cause a delay on the order of 10–30 minutes before notification of out-of-spec material might reach the manufacturing floor. During that lag period between sampling and results, quite a lot of bad products will have been made. The addition of an on-line color analyzer can greatly improve efficiency by reducing scrap and can also improve product quality by allowing for control of the color to a tighter specification. The hardware configuration of the process analyzer for this application is identical to that for the above nickel application, with the exception of the fiber-optic probe and the software algorithms. In this case, the probe is a specular-excluded reflectance probe with a 45/0° (illumination/detection) geometry that is mounted a short distance (<0.5 cm) above the moving sample web. Such probes can be extremely rugged and may be placed only a short distance outside of the extruder head. Again, the analyzer runs in a two-channel (dual beam) configuration in which a reference fiber loop serves as real-time compensation for lamp intensity fluctuations. A white reference tile serves as the instrument's color calibration. In a vinyl siding installation, it is possible to reliably measure ΔE of 0.3 with this type of UV-vis system (ΔE is a measure of the distance between two points (colors) in color space). Gloss changes in the product do not affect the measured color coordinates. As long as product embossing depth is ~50% or less than the optimum probe-sample distance, it too does not affect the measured color coordinates.

4.7.4 Film thickness determination: polymer

In this application, the process analyzer is used in the vis-NIR spectral region to measure the clear top layer on a co-extruded polymer film. The bottom layer is pigmented to an opaque white color and its thickness cannot be determined by this method. Prior to the installation of the fiber-optic spectroscopy system, film samples were measured manually in the laboratory by a subtractive scheme. First, the total thickness of a sample was measured on a manual profilometer. The top layer of the polymer was removed with methylene chloride. The sample was then repositioned on the profilometer as closely as possible to the originally measured spot and the thickness of the second white layer was determined. The thickness of the top layer was then determined by difference.

There were a number of problems with this analytical scheme: difficulty in measuring the same physical spot before and after solvent treatment, use of a hazardous solvent and generation of an additional waste stream for the lab. The fiber-optic spectroscopy system eliminates all three of these problems, at the same time providing more accurate and more rapid analysis of the samples. The film thickness analyzer consists of a fiber-optic CCD spectrophotometer, tungsten-halogen light source, and a diffuse reflectance probe to interrogate the sample. The tip of the probe is held at a constant distance from the film sample by means of a sample-holder jig. A comparison of the optically determined thickness to the manual laboratory method is shown in Figure 4.8.

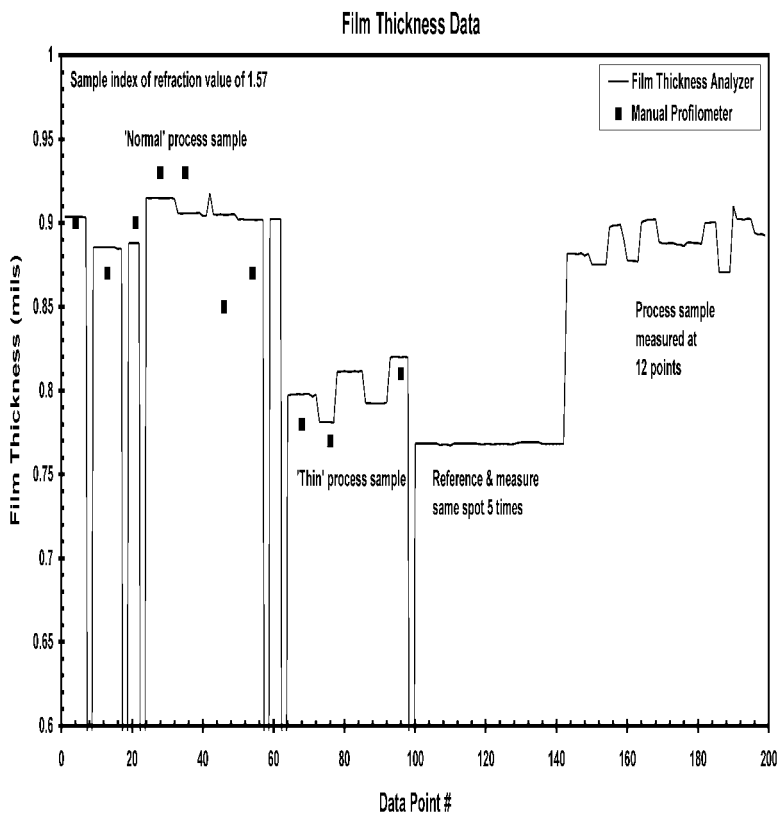


Figure 4.8 *Comparison of optical thickness measurements to a physical technique.*

4.8 Conclusion

The present review illustrates that UV-vis based process analytical technology is widely discussed, but published examples are still scarce. The pharmaceutical industry, driven by the FDA's PAT initiative, can be seen as a forerunner. This is reflected by the observation that many references in this chapter are (co-) authored by representatives from pharmaceutical companies like Glaxo Wellcome[21,29] and GlaxoSmithKline,[8,10,36,39,40] Astra Zeneca,[19,38] Novartis[28] and Eli Lilly.[18]

As with other spectroscopic techniques, UV-vis spectroscopy can readily be used in lab-based R&D. The implementation as process monitors, scanning for deviations from the set state, is certainly an added value and low in maintenance requirements. UV-vis based process control is more demanding on precision, reliability, exactness and operability. The steady progress both in hardware and software is facilitating the growth of UV-vis process analytics in both the pharmaceutical and chemical industries. Given the vast increase in information and knowledge-based control, this may just be what the industries need.

References

1. R.S. Saltzman, Ultraviolet/Visible Spectroscopy in Process Analyses, *Encyclopedia of Analytical Chemistry*, John Wiley, Chichester, 8328–8334 (2000).
2. J.A. Schmidt, Ultraviolet/visible spectroscopy of chromophores in mechanical pulps, in *Encyclopedia of Analytical Chemistry*, John Wiley & Sons Ltd, Chichester, 8388–8406 (2000).
3. H.G. Brittain, UV/vis reflectance spectroscopy, in *Spectroscopy of Pharmaceutical Solids*, J. Swarbrick (ed.), Taylor & Francis, London, 121–149 (2006).
4. C. Bosch Ojeda, F. Sánchez Rojas, Process analytical chemistry: applications of ultraviolet/visible spectrometry in environmental analysis: an overview, *Appl. Spectrosc. Rev.*, **44**(3), 245–265 (2009).
5. B.M. Weckhuysen, Ultraviolet-visible spectroscopy, in B.M. Weckhuysen (ed.), *In situ Spectroscopy of Catalysts*, American Scientific Publishers, Stevenson Ranch, 255–270 (2004).
6. F.C. Jentoft, Ultraviolet–Visible–Near Infrared Spectroscopy in Catalysis: Theory, Experiment, Analysis, and Application Under Reaction Conditions, *Adv. Catal.*, **52**, 129–211 (2009).
7. S.L. Upstone, Ultraviolet/Visible Light Absorption Spectrophotometry in Clinical Chemistry, *Encyclopedia of Analytical Chemistry*, John Wiley, Chichester, 1699–1714 (2000).
8. T.J. Thurston, R.G. Brererton, D.J. Foord, R.E.A. Escott, Principal components plots for exploratory investigation of reactions using ultraviolet-visible spectroscopy: application to the formation of benzophenone phenylhydrazone, *Talanta*, **63**, 757–769 (2004).
9. Sh. Agarwal, Ph. Cluxton, M. Kemper, D.D. Dionysiou, S.R. Al-Abed, Assessment of the functionality of a pilot-scale reactor and its potential for electrochemical degradation of calmagite, a sulfonated *azo*-dye, *Chemosphere*, **73**, 837–843 (2008).
10. T.J. Thurston, R.G. Brererton, D.J. Foord, R.E.A. Escott, Monitoring of a second-order reaction by electronic absorption spectroscopy using combined chemometric and kinetic models, *J. Chemom.*, **17**, 313–322 (2003).
11. F. Sánchez Rojas, C. Bosch Ojeda, Recent development in derivative ultraviolet/visible absorption spectrophotometry: 2004–2008; A review, *Anal. Chim. Acta*, **635**, 22–44 (2009).
12. ASTM E313-73, (reapproved 1993), Standard Test Method for Indexes of Whiteness and Yellowness of Near-White, Opaque Materials.
13. ASTM D1925-70 (1988, withdrawn 1995): Standard Test Method for Yellowness Index of Plastics.
14. E. Huber, M. Frost, Light scattering by small particles, *J. Water SRT – Aqua*, **47**, 87–94 (1998).
15. J. van den Broeke, G. Langergraber, A. Weingartner, On-line and *in-situ* UV/vis spectroscopy for multi parameter measurements: a brief review, *Spectroscopy Europe*, **18**(4), 15–18 (2006).
16. R. Salvati, A. Longo, G. Carutenuto, S. De Nicola, G.P. Pepe, L. Nicolais, A. Barone, UV-vis spectroscopy for on-line monitoring of Au nanoparticles size during growth, *Appl. Surf. Sci.*, **248**, 28–31 (2005).
17. J.M. Montgomery, T.-W. Lee, S.K. Gray, Theory and modelling of light interactions with metallic nanostructures, *J. Phys.: Condens. Matter*, **20**, 1–11 (2008).

18. L. Liu, G. Fitzgerald, M. Embry, R. Cantu, B. Pack, Technical evaluation of a fiber-optic probe dissolution system, *Dissol. Technol.*, **15**, 10–20 (2008).
19. J. Johansson, M. Cauchi, M. Sundgren, Multiple fibre-optic dual-beam UV/Vis system with application to dissolution testing, *J. Pharm. Biomed. Anal.*, **29**, 469–476 (2002).
20. ASTM E387-04: Standard test method for estimating stray radiant power ratio of dispersive spectrophotometers by the opaque filter method.
21. A.C. Quinn, P.J. Gemperline, B. Baker, M. Zhu, D.S. Walker, Fibre-optic UV/visible composition monitoring for process control of batch reactions, *Chemom. Intell. Lab. Syst.*, **45**, 199–214 (1999).
22 S. O'Keeffe, C. Fitzpatrick, E. Lewis, An optical fibre based ultra violet and visible absorption spectroscopy system for ozone concentration monitoring, *Sens. Actuators B*, **125**, 372–378 (2007).
23. G. Langergraber, J. van den Broeke, W. Lettl, A. Weingartner, Real-time detection of possible harmful events using UV/vis spectrometry, *Spectroscopy Europe*, **18**(4), 19–22 (2006).
24. D.J.W. Ng, M. Assirelli, Mixing study in batch stirred vessels using a fibre-optic UV-VIS monitoring technique: a novel method, *Chem. Eng. Res. Des.*, **85**(10), 1348–1354 (2007).
25. *The United States Pharmacopeia (USP)* **26**, (2003), United States Pharmacopeial Convention, Inc., Rockville, MD: General Notices, 6–7; General Chapter <711> Dissolution, 2156.
26. X. Lu, R. Lozano, P. Shah, *In-Situ* dissolution testing using different UV fiber optic probes and instruments, *Dissol. Technol.*, **10**, 6–15 (2003).
27. C.A. Martin, Evaluating the utility of fiber optic analysis for dissolution testing of drug products, *Dissol. Technol.*, **10**(4), 37–39 (2003).
28. T. Mirza, Q. Liu, R. Vivilecchia, Y. Joshi, Comprehensive validation scheme for *in situ* fiber optics dissolution method for pharmaceutical drug product testing, *J. Pharm. Sci.*, **98**(3), 1086–1094 (2009).
29. P.J. Gemperline, J. Cho, B. Baker, B. Batchelor, D.S. Walker, Determination of multicomponent dissolution profiles of pharmaceutical products by in situ fiber-optic UV measurements, *Anal. Chim. Acta*, **345**, 155–159 (1997).
30. Y. Coque, E. Touraud, O. Thomas, On line spectrophotometric method for the monitoring of colour removal processes, *Dyes and Pigments*, **54**, 17–23 (2002).
31. A.-St. Jääskeläinen, T. Liitiä, UV/vis reflectance spectroscopy reveals the changes in fibre chemistry during ageing, *Spectroscopy Europe*, **19**(5), 11–13 (2007).
32. W. Shen, X. Chen, Online monitoring of kappa number during batch pulping by visible spectroscopy, *Ind. Eng. Chem. Res.*, **48**, 4872–4876 (2009).
33. E. Ayranci, B.E. Conway, Adsorption and electrosorption of ethyl xanthate and thiocyanate anions at high-area carbon-cloth electrodes studied by in situ UV spectroscopy: development of procedures for wastewater purification, *Anal. Chem.*, **73**, 1181–1189 (2001).
34. S.P. Gurden, J.A. Westerhuis, A.K. Smilde, Monitoring of batch processes using spectroscopy, *AIChE J.*, **48** (10), 2283–2297 (2002).
35. C.A. McGill, A. Nordon, D. Littlejohn, Comparison of in-line NIR, Raman and UV-visible spectrometries, and at-line NMR spectrometry for the monitoring of an esterification reaction, *Analyst*, **127**, 287–292 (2002).
36. A.R. Carvalho, R.G. Brereton, T.J. Thurston, R.E.A. Escott, Influence of different sources of error on estimated kinetics parameters for a second-order reaction, *Chemom. Intell. Lab. Syst.*, **71**, 47–60 (2004).
37. F. Benito-Lopez, W. Verboom, M. Kakuta, *et al.*, Optical fiber-based on-line UV/Vis spectroscopic monitoring of chemical reaction kinetics under high pressure in a capillary microreactor, *Chem. Commun.*, 2857–2859 (2005).
38. E. Deneau, G. Steele, An in-line study of oiling out and crystallization, *Org. Process Res. Dev.*, **9**, 943–950 (2005).
39. J. Gabrielsson, H. Jonsson, J. Trygg, C. Airiau, B. Schmidt, R. Escott, Combining process and spectroscopic data to improve batch modeling, *AIChE J.*, **52**(9), 3164–3172 (2006).
40. C.W.L. Wong, R. Escott, E. Martin, J. Morris, The integration of spectroscopic and process data for enhanced process performance monitoring, *Can. J. Chem. Eng.*, **86**, 905–923 (2008).
41. K. Beckenkamp, K. Germann, J. Knab, *et al.*, In-line UV-Spektroskopie zur Überwachung und Steuerung einer Destillationskolonne, *GIT Labor-Fachzeitschrift*, **6**, 393–395 (2002).
42. V. Vojinović, J.M.S. Cabral, L.P. Fonseca, Real-time bioprocess monitoring, Part I: In situ sensors, *Sens. Actuators B*, **114**, 1083–1091 (2006).
43. St. Marose, C. Lindemann, R. Ulber, Th. Scheper, Optical sensor systems for bioprocess monitoring, *Trends Biotech.*, **17**(1), 30–34 (1999).
44. L. Noui, J. Hill, P.J. Keay, *et al.*, Development of a high resolution UV spectrophotometer for at-line monitoring of bioprocesses, *Chem. Eng. Proc.*, **41**, 107–114 (2002).
45. S.J. Foulk, K.L. Miller, R. Spatafore, On-line monitoring of polymer melts using UV fiber optic spectroscopy, *Proc. SPIE Int. Soc. Opt.*, **2069**, 85–94 (1993).

46. L. Pernenkil, C.L. Cooney, A review on the continuous blending of powders, *Chem. Eng. Sci.*, **61**, 720–742 (2006).
47. J.H. Gray, Performance of dry solids mixing equipment, *Chem. Eng. Prog.*, **53**(1), 25–32 (1957).
48. M.D. Ashton, C. Schofield, F.H.H. Valentin, The use of a light probe for assessing the homogeneity of powder mixtures, *Chem. Eng. Sci.*, **21**, 843–849 (1966).
49. R. Weinekötter, L. Reh, Continuous mixing of fine particles, *Part. Part. Syst. Charact.*, **12**, 46–53 (1995).
50. St. Weiß, Kinetische Untersuchungen bei der Oxidation von Kohlenwasserstoffen an oxidischen Katalysatoren, Logos, Berlin, 2007. Also: PhD thesis, Erlangen University.
51. C.P. Stemmet, J.C. Schouten, T.A. Nijhuis, In situ UV–vis spectroscopy in gas–liquid–solid systems, *Chem. Eng. Sci.* (2009), doi:10.1016/j.ces.2009.05.031
52. R. Philipps, St. Walter, M.A. Liauw, Selective oxidation of isoprene to citraconic anhydride, *Chem. Eng. J.*, **90**, 107–116 (2002).

5

Near-infrared Spectroscopy for Process Analytical Technology: Theory, Technology and Implementation

Michael B. Simpson

Industry Manager FTIR for Refining & Petrochemicals, ABB Analytical Measurements, Quebec, Canada

5.1 Introduction

Process analytical technology (PAT) applications of near-infrared spectroscopy (NIRS) range in difficulty from the tolerably simple (e.g. moisture content) to the hugely ambitious[1–3] (e.g. full characterization of hydrocarbon composition and properties in refinery process streams). The various demands of these applications, in order to yield useful process control information, may be quite different. The key criteria may be in some cases speed of response, in others sensitivity, and still in others long-term analyzer response and wavelength stability. It is therefore not surprising that implementations of NIRS for chemical and pharmaceutical process monitoring[4–6] cover a very wide range of NIR technologies, sample interfaces and calibration methodologies.

However, what unites all applications of NIRS for PAT are the unique features of the near-infrared spectrum. The near-infrared region is in effect the chemical spectroscopy[7] of the hydrogen atom in its various molecular environments. The frequency range of the NIR from about 4000 cm^{-1} up to 12 500 cm^{-1} (800 to 2500 nm) covers mainly overtones and combinations of the lower energy fundamental molecular vibrations that include at least one X–H bond vibration. These are characteristically significantly weaker in absorption cross section, compared with the fundamental vibrational bands from which they originate. They are faint echoes of these mid-infrared absorptions. Thus for example NIR absorption bands formed as combinations of mid-infrared fundamental frequencies (e.g. $\nu 1 + \nu 2$), typically have intensities 10 times weaker than the weaker of the two original mid-infrared bands. For NIR overtone absorptions (e.g., $2\nu 1$, $2\nu 2$) the decrease in intensity can be 20 to 100 times that of the original band.

Process Analytical Technology 2e Edited by Katherine Bakeev
© 2010 John Wiley & Sons, Ltd.

It follows that all mid-infrared fundamental absorptions in the fingerprint region (roughly below 1700 cm^{-1}) arising from the molecular skeleton (C–C, C=C, C=O, etc.) or from functional groups containing heavier atoms (C–Cl, C–N, etc.) are already into their second or third overtone above 4000 cm^{-1}, and hence too weak to contribute significantly in the NIR. The single practically useful exception to this is the appearance of weak C=O absorption bands in the second overtone just below 5000 cm^{-1}, where they can be observed among the normal hydrocarbon combination bands.

However, what saves the day, in terms of the analytical specificity of NIRS, is the (sometimes extreme) sensitivity of the frequency and intensity of these X–H NIR absorption bands to near neighbours in the molecular structure. The local electronic environment has a particularly strong influence on the X–H bond force constants, and from this derives a remarkably high information content in NIR spectra. Furthermore some particular functional groups (e.g. O–H, N–H) are very strongly affected by both intermolecular and intramolecular H-bonding effects, with sometimes dramatic influences on the intensity and band-shapes in the NIR. Nuclear magnetic resonance (NMR) functions analogously where much molecular structure is deduced from neighboring influences on the nuclear magnetic resonance frequency for a particular atom type except that in NIR one can view a number of marker frequencies such as O–H, C–H and N–H simultaneously.

The functional groups almost exclusively involved in NIRS are those involving the hydrogen atom: C-H, N-H, O-H (see Figure 5.1). These groups are the overtones and combinations of their fundamental frequencies in the mid-infrared and produce absorption bands of useful intensity in the NIR. Because the absorptivities of vibrational overtone and combination bands are so much weaker, in NIRS the spectra of condensed phase, physically thick samples, can be measured without sample dilution or the need to resort to difficult short-path length sampling techniques. Thus conventional sample preparation is redundant, and fortunately so, because most PAT applications require direct measurement of the sample[8,9] either *in situ*, or after extraction of the sample from the process in a fast loop or bypass.

A further benefit of the low absorbtivity of most samples in the NIR is that measurements involving scattering effects (both diffuse transmission and diffuse reflectance) are possible. The penetration depth of the sampling beam in diffuse reflectance measurements of powders in the NIR can be on the scale of millimeters, despite the numerous refractive index interfaces and extended optical path involved. Thus relatively large volumes of material can be interrogated, avoiding problems of surface contamination and sample nonhomogeneity.

Whilst it is true that the majority of NIR analyzers installed in the field for quality monitoring purposes are based on measurements of discrete wavebands using multiple interference band-pass filters, these analyzers are generally in agricultural, food or bulk commodity material QA applications. This article is focused on NIRS applications in process analytical technology or chemistry, and will concentrate on NIR technologies capable of yielding full spectral information – the starting point for the complex data reduction and calibration methods used for PAT quantitative applications. Even with this restriction, there are a surprising number of distinct NIR technologies[10] available, which will be discussed in detail, but which are briefly described here to set the scene.

The most conventional analyzer format used to achieve full-spectrum information is the scanning grating monochromator (Figure 5.2). In its current modern format this is a robust piece of equipment, normally based on a concave holographic grating, controlled via an industrial motor drive and optical encoder for precise positioning. Under some circumstances this format, using a standard tungsten–halogen source at 2800 K and suitable detector can achieve a near-perfect match of the monochromator throughput with detector sensitivity, and hence very high signal-to-noise ratio (SNR). However this system may have limitations; wavelength registration is mechanically derived, and under real-life circumstances (for example in multiplexed fiber-optic sampling) the geometric etendue limitations of a slit arrangement (even when the input

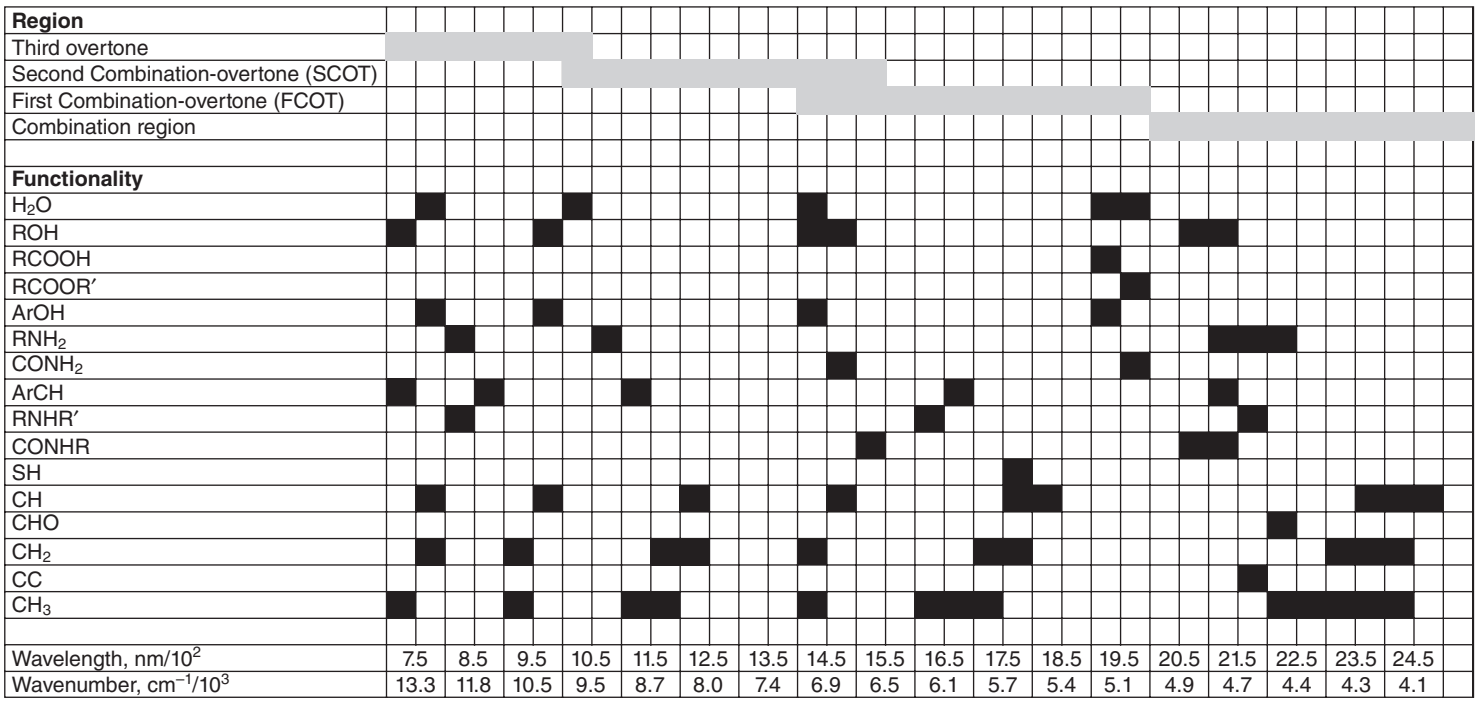


Figure 5.1 *NIR correlation chart.*

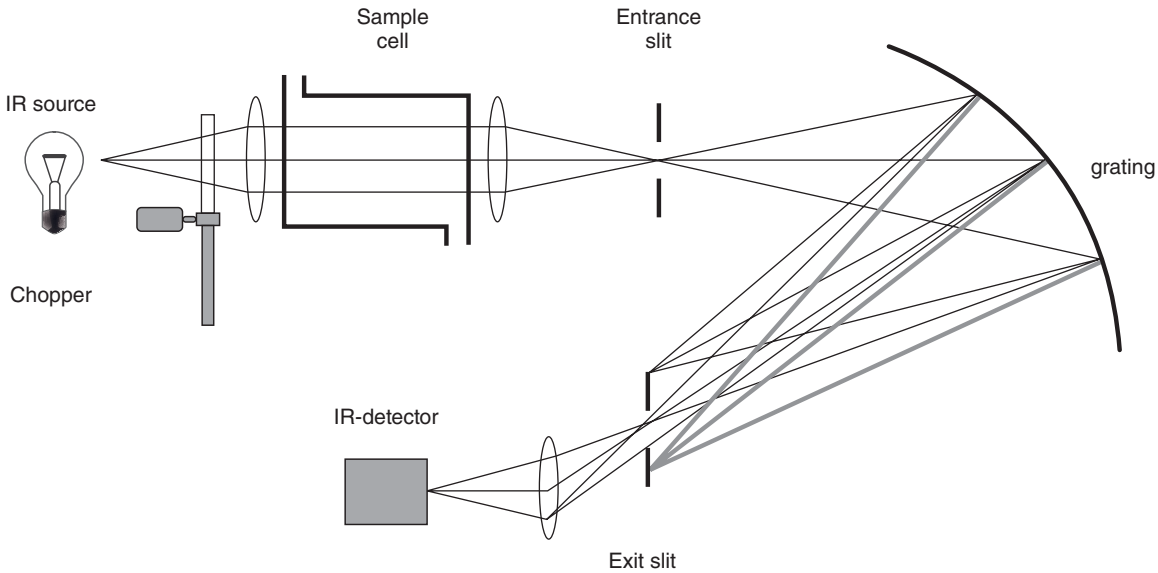


Figure 5.2 *Scanning grating monochromator.*

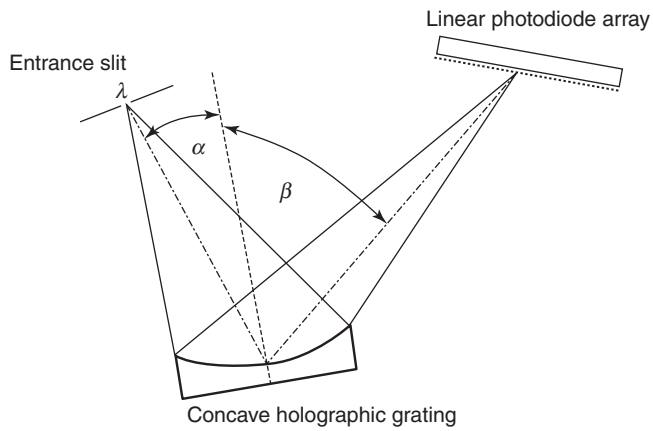


Figure 5.3 *PDA grating polychromator.*

and output fibers are used to define the slit) can make it beneficial to look at higher optical throughput alternatives[11,12].

An obvious extension of the scanning monochromator is the grating polychromator (Figure 5.3), fitted with a photodiode array. Whilst this system still requires an entrance slit to the concave grating, the exit slit is replaced with a photodiode array, allowing measurement in a number of detection channels simultaneously. This allows for a multiplex measuring advantage (all frequencies are measured at the same time, rather than scanned sequentially as in a scanning monochromator), so response time can be short and SNR kept high. However, as with all diffraction optics systems, there are issues of stray light and diffraction order-sorting to be addressed, which in practice restrict the spectral free range of the diode array. A signifi-

cant advantage of the PDA grating polychromator is that it continuously views the full spectrum. This makes measurements immune to spectral distortions due to intensity variations resulting from inhomogeneous samples passing through the optical beam

A further example of a high-throughput full-spectrum device is the tuneable filter instrument (Figure 5.4), normally based on an acousto-optical tuneable filter (AOTF). In this case no entrance slit is required. As much of the source radiant energy as can be collected and collimated, and which the thermal properties of the AOTF filter can withstand, can be used. The AOTF filter is based on the use of a carefully oriented birefringent crystal, optically transparent across the NIR (TeO_2 is a typical example). The crystal is driven by a radio frequency (RF) powered piezoelectric transducer, which sets up an acoustic wave that propagates through the crystal and interacts with the incident broadband NIR radiation. Most of the NIR light passes through unaffected, but one specific narrow waveband is diffracted – with a change of orientation due to the birefringence of the crystal. This allows the broad-band light (which passes straight through) to be physically blocked (or partly used as a reference signal), and the diffracted narrow band pass to be channeled into the sample. In principle, the broad-band and diffracted narrow band-pass elements can be separated using crossed polarizers on the input and exit of the crystal, but this requires extremely high rejection efficiency to avoid stray light issues. In practice commercial devices use physical beam blocks. AOTF devices can be used in scanning or random frequency access mode, depending on the way in which the RF power input is driven.

Finally, in the field of full-spectrum NIRS methods, Fourier transform near-infrared (FTIR) analyzers are included (Figure 5.5). FTIR techniques are predominant in mid-infrared spectroscopy because there are clear and absolute advantages for the FTIR analyzer in the mid-infrared compared with any other available technology. This arises because of the low power output of mid-infrared sources and the low specific detectivity

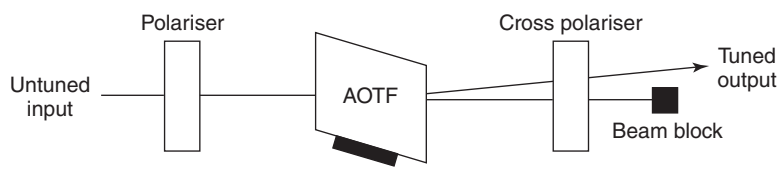


Figure 5.4 *AOTF monochromator.*

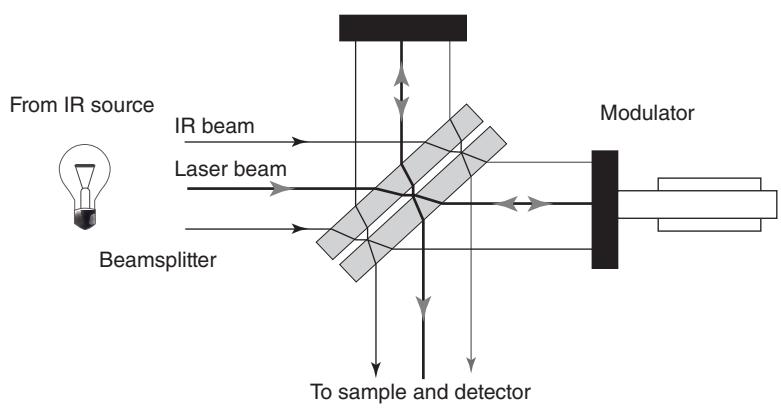


Figure 5.5 *FT-NIR modulator.*

of mid-infrared detectors, which mean that the classic Jacquinot (optical throughput) and Fellgett (multiplex sampling) advantages can be fully realized. The argument is a little more complex as applied to NIRS, and will be discussed in more detail later. An FTIR device is in effect an optical modulator which allows wavelength encoding. The input beam is unmodulated broadband NIR, and the exit beam from the interferometer is still broadband NIR, but with each optical frequency uniquely amplitude-modulated in the acoustic frequency range. This allows the detector signal (which is a combination of signals from all the incident broadband NIR frequencies) to be decomposed using a Fourier transform, and the individual amplitudes of each optical frequency in the broadband input signal to be obtained. The system thus has no input or output restrictions other than a resolution defining Jacquinot stop. Thus, in the conventional arrangement using one exit beam of a two-port interferometer (the other beam returns to the source) 50% of the source power is available at all frequencies all the time. The amplitude modulation of the NIR beam is normally achieved using a beam division/retardation/recombination sequence, but there are a variety of physical methods to implement this. The uniquely defining feature of FTIR methods in NIRS is that they all contain some inherent form of direct wavelength registration. Most practical devices employ a visible frequency laser to define the optical retardation sampling interval. This translates directly into the frequency axis of the recorded spectrum, linked intimately with the known HeNe laser frequency. This final FTIR advantage (the Connes advantage) turns out perhaps to be the major practical contribution of FTIR to NIRS applications.

These four very different NIR technologies represent the mainstream analyzer types used in many process analytical technology applications.[13] They also cover the main types of established commercially available analyzers. Given their various characteristics and the physical principles upon which they are based, it is obvious that they will have different advantages and disadvantages depending on the detailed circumstances of the application. These issues will be covered in more detail below.

5.2 Theory of Near-Infrared Spectroscopy

The atoms in a molecule are never stationary, even close to the absolute zero temperature. However the physical scale of the vibrational movement of atoms in molecules is rather small – of the order of 10^{-9} to 10^{-10}cm. The movement of the atoms in a molecule is confined within this narrow range by a potential energy well, formed between the binding potential of the bonding electrons, and the repulsive (mainly electrostatic) force between the atomic nuclei. Whenever atomic scale particles are confined within a potential well, one can expect a quantum distribution of energy levels.

The solution to the Schrödinger equation for a particle confined within a simple harmonic potential well is a set of discrete allowed energy levels with equal intervals of energy between them. It is related to the familiar simple solution for a particle in an infinite square well, with the exception that in the case of the simple harmonic potential, the particle has a non-zero potential energy within the well. The restoring force in a simple harmonic potential well is $-kx$, and thus the potential energy $V(x)$is $\frac{1}{2}kx^2$ at displacement x, and the Schrödinger equation to be solved for the quantized energy levels in the well becomes:

$$-\frac{\boldsymbol{h}^2}{2m}\frac{\partial^2\Psi}{\partial x^2}+\frac{1}{2}kx^2\Psi(x)=E\Psi(x) \tag{5.1}$$

(where here and throughout $\boldsymbol{h} = h/2\pi$, and $\psi(x)$ is the vibrational wavefunction).

Note that the potential energy $V(x)$ rises to infinite values at sufficiently large displacement. One should expect this boundary condition to mean that the vibrational wavefunction will fall to zero amplitude at large displacement (as in the square well case, but less abruptly). One should also expect that the confining potential well would lead to quantized solutions, as is indeed the case:

$$E_n = \left(n + \frac{1}{2}\right)\boldsymbol{h}\omega \tag{5.2}$$

where $n = 0, 1, 2$ and $\omega = (k/m)^{1/2}$ and $\boldsymbol{h} = h/2\pi$

The first important conclusion from this solution of a simple harmonic potential well is that the gap between adjacent energy levels is exactly the same:

$$\Delta E = E_{(n+1)} - E_n = \boldsymbol{h}\omega \tag{5.3}$$

The only changes in energy level for the particle which are allowed involve a unit change in quantum number.

$$\Delta n = \pm 1 \tag{5.4}$$

At room temperature the value of the Boltzmann factor (kT) is about $200\,\text{cm}^{-1}$, so the vast majority of molecules are in the their vibrational ground state ($n = 0$). Thus in the case of the simple harmonic model above, only transitions to ($n = 1$) are allowed, and NIR spectroscopy would not exist. This is clearly not the case, and the answer lies in the nonconformity of the potential energy well in real molecules with the simple harmonic model. In the simple harmonic (parabolic well) model, it can be seen that the potential well is entirely symmetrical. The potential energy rises equally with displacement in both positive and negative directions from the equilibrium position. This is counterintuitive, because the forces responsible for the rise in potential energy are different in the two cases. In the bond-stretching case the dominant factor is the shifting of the molecular orbital away from its minimum energy configuration. In the bond-compression case, there is the additional factor of electrostatic repulsive energy as the positively charged atomic nuclei approach each other. Thus one would expect the potential energy curve to rise more steeply on the compression cycle, and (due to the weakening of the bond with displacement) to flatten out at large displacements on the decompression cycle.

There are two effects of the anharmonicity of the quantized energy levels described above, which have significance for NIRS. First, the gap between adjacent energy levels is no longer constant, as it was in the simple harmonic case. The energy levels converge as n increases. Second, the rigorous selection rule that $\Delta n = \pm 1$ is relaxed, so that weak absorptions can occur with $n = \pm 2$ (first overtone band), or ± 3 (second overtone band), etc.

Taking account of the anharmonic potential well, for the transition $n = 0$ to $n = 1$ (the fundamental vibration) one sees the absorption now at

$$h\nu = \boldsymbol{h}\omega - 2X_e\boldsymbol{h}\omega \tag{5.5}$$

with the new factor X_e known as the anharmonicity constant $X_e = \boldsymbol{h}a^2/2m\omega$.

However, an overtone transition in an anharmonic well involving $\Delta n = 2$ has an energy of

$$h\nu = \Delta E = 2\boldsymbol{h}\omega - 2(2n+3)X_e\boldsymbol{h}\omega \tag{5.6}$$

where n is the initial state quantum number.

Thus for the transition $n = 0$ to $n = 2$ the first overtone absorption is seen at

$$h\nu = \Delta E = 2\boldsymbol{h}\omega - 6X_e\boldsymbol{h}\omega \tag{5.7}$$

This is not twice the fundamental frequency, but is lower than this, due to the increased effect of the anharmonicity constant in transitions to higher values of n.

So far the effect of anharmonicity in determining the frequency of overtone ($\Delta n > 1$) absorptions, and its effect in relaxing quantum selection rules to allow these transitions to have some absorption intensity have been considered.

However, in polyatomic molecules, transitions to excited states involving two vibrational modes at once (combination bands) are also weakly allowed, and are also affected by the anharmonicity of the potential. The role of combination bands in the NIR can be significant. As has been noted, the only functional groups likely to contribute to the NIR spectrum directly as overtone absorptions are those containing C–H, N–H, O–H or similar functionalities. However, in combination with these hydride bond overtone vibrations, contributions from other, lower frequency fundamental bands such as C=O and C=C can be involved as overtone-combination bands. The effect may not be dramatic in the rather broad and overcrowded NIR absorption spectrum, but it can still be evident and useful in quantitative analysis.

5.3 Analyzer Technologies in the Near-infrared

In this section the mainstream NIRS technologies in current widespread practical use for process analytical technology applications will be examined. These are the scanning grating monochromator, the grating polychromator photodiode array spectrometer, the acousto-optic tunable filter (AOTF) analyzer, and the Fourier transform NIR analyzer. In Section 5.3.5 we will also present an overview of the new emerging technologies which are beginning to have an impact in NIR applications for process analytical technology, but which are not yet in wide use. But first we will look at the features of the common aspects of all NIR analyzer technologies – light sources and detectors.

5.3.1 Light sources and detectors for near-infrared analyzers

5.3.1.1 The Tungsten–Halogen Source and the Halogen Cycle

The tungsten–halogen source is almost universally used for NIR spectroscopy. It has a broadband, pseudo-blackbody emission spectrum with no significant structure. It is inexpensive and remarkably long-lived if operated at appropriate filament temperature and lamp wattage. Current generation FT-NIR analyzers, for example, are able to use suitable source power loadings to achieve >5 years MTBF (mean time before failure). The peak spectral radiance of a tungsten lamp operating at a filament temperature of 2800 K is located at approximately 10 000 cm^{-1} (1000 nm) – very slightly higher frequency than that of a blackbody at the same temperature, but at less than half the peak intensity. Increasing the filament temperature above 3000 K might boost the peak intensity of the spectral radiance to two-thirds of the blackbody at 2800 K, but with a dramatic reduction in source lifetime. For practical process NIR analyzers (of all technologies, for which the tungsten–halogen lamp is the standard broadband NIR source, even though we discuss it here in relation to scanning monochromators) the required source lifetime between changes needs to be of the order of 12 months continuous operation. 10 000 hours of source lifetime is not an unreasonable expectation when operated at 2800 K and at or below the rated voltage for the filament. For a hot filament one would expect the room-temperature resistivity to be much lower than when at operating temperature. This is the case, and unless the source power supply is designed to be current limited, there will be a significant initial current surge, with potential for excessive filament vaporization and redeposition.

The tungsten–halogen lamp operates in a very specific mode, dependent on the detailed chemistry of the atmosphere within the lamp enclosure. The lamp envelope is normally a thin quartz bulb, filled with an inert

gas such as krypton or xenon and a trace of a halogen, often bromine or iodine. Under operating conditions, even with the increased pressure of the inert gas filling, tungsten is vaporized from the filament. Without taking care to deal with this, the evaporated tungsten metal would simply condense on the coolest part of the lamp available, which would normally be the inside of the glass envelope, leading to blackening, loss of emission intensity and premature filament failure. The trace of halogen in the lamp reacts with the deposited elemental tungsten on the inside of the envelope (provided that the lamp is run at a high enough temperature – the envelope needs to be above about 200 °C for this process to work efficiently). This halogen reaction not only cleans the envelope, but also allows for transport of the tungsten halide back into the vapor phase, where it eventually reaches the surface of the filament, where it is destroyed, and the elemental tungsten is redeposited back at the filament, whilst the halogen is released back into the vapor phase – hence the halogen cycle. Incidentally, this is the reason why tungsten–halogen lamps should never be touched with unprotected fingers during lamp changes. The diffusion of sodium (as the salt in sweat) back through the thin glass envelope eventually causes the loss of the trace halogen from the lamp atmosphere, and the interruption of the halogen cycle.

5.3.1.2 Photodetectors for the Near-infrared

The earliest detectors used commonly in near-infrared process analyzers were lead sulfide and lead selenide (PbS and PbSe) photoconductive devices, which are both highly sensitive and inexpensive. They were ideally suited for use in scanning grating monochromators. They do however have limitations in linearity, saturation and speed of response. Consequently the adoption of near-infrared analyzer technologies requiring some or all of these characteristics saw the rise in use of photodiode devices such as the indium gallium arsenide (InGaAs) detector. These are both very linear and very fast and sensitive, making them the idea all-round detector for near-infrared applications.

There are two basic types of photodetectors available for use in NIRS: thermal detectors and photon devices. Photon devices can be divided into photoconductive and photodiode detectors. The latter can be operated in either photoconductive or photovoltaic modes (with and without an applied bias voltage respectively). Thermal detectors are not common in most NIR analyzers, but they have a unique and useful place, as pyroelectric detectors in FT-NIR instruments. The higher overall optical throughput of an FT instrument allows the far less intrinsically sensitive pyroelectric device to be used successfully, with the added advantage that depending on the choice of optical materials and source technology, extended range instruments can be built spanning simultaneously both NIR and mid-infrared applications.

Pyroelectric detectors[14] depend on the use of a thin slice of ferroelectric material (deuterated triglycine sulfate [DTGS], Figure 5.6, is the standard example) – in which the molecules of the organic crystal are naturally aligned with a permanent electric dipole. The thin slab is cut and arranged such that the direction

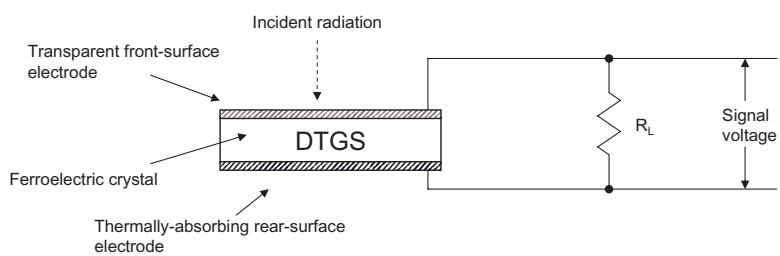


Figure 5.6 *The DTGS pyroelectric detector.*

of spontaneous polarization is normal to the large faces. Typically one surface of the crystal is blackened to enhance thermal absorption, and the entire assembly has very low thermal mass.

Changes in the temperature of the DTGS crystal, induced by incident broadband infrared radiation, cause the degree of polarization to change, and hence changes the amount of charge stored at the crystal surface. Electrodes arranged as in Figure 5.6 are driven by this difference in surface charge between the back and front faces of the crystal, and a current will flow generating a signal voltage across the load resistor R_L.

- Since it is a thermal and not a photon detector, there is no band gap, and no specific cut-off frequency. The device is a broadband detector capable of operation from the NIR to the long-wavelength mid-infrared.
- Because there is no band gap, and electron-hole pair generation is not involved in the detector response, there is no issue of bleaching (carrier depletion) under high levels of irradiance, so the detector can achieve good linearity performance.
- To generate an output signal the DTGS detector requires a modulated NIR signal. A constant irradiance will generate no change in voltage across the load. Even with a low thermal mass detector element, and appropriate values of R_L, the response time of a DTGS device will be slow, but can extend into the kHz range, which fits well with the amplitude modulation audio frequencies typical of an FT instrument.
- Because the response time of the detector depends on the thermal time constant of the detector element / electrode assembly, coupled with the electrical time constant of the device capacitance and load resistor – the response versus modulation frequency (f_m) shows a typical $1/f_m$ form.

The pyroelectric DTGS detector is a very useful low-cost, general purpose, wideband NIR detector well suited for use in FT-based analyzers. It is not normally used in scanning monochromators where higher sensitivity detectors are needed to match the lower optical throughput and discrete wavelength scanning requirements.

We now turn to photoconductive and photodiode detectors, both of which are semiconductor devices. The difference is that in the photoconductive detector there is simply a slab of semiconductor material, normally intrinsic to minimize the detector dark current, though impurity doped materials (such as B doped Ge) can be used for special applications, whereas by contrast, the photodiode detector uses a doped semiconductor fabricated into a p–n junction.

In an intrinsic semiconductor material there is an energy gap, known as the band-gap, between the valence band (populated with electrons) and the higher energy, nominally empty conduction band. An incident photon of sufficient energy can interact with a valence band electron, and push it to a higher energy state in the conduction band, where it is free to migrate, under the influence of an applied bias voltage, to the collection electrodes deposited on the lateral surfaces of the detector element. Any change in detector element conductivity will result in a flow of current around the circuit, driven by the bias voltage, and detected across a load resistor. Typical materials used in NIR photoconductive detectors[14] are PbS, PbSe, InSb and InAs (lead sulfide, lead selenide, indium antimonide and indium arsenide). The response curves of examples of these materials are shown in Figure 5.7.

Since photon energies decrease with increasing wavelength, there will be a maximum wavelength beyond which the photons lack sufficient energy to cause electron transitions into the conduction band, known as the cut-off wavelength, λ_g

$$\lambda_g = hc/\Delta E_g \tag{5.8}$$

There are a few complications to this simple picture. First, it has been assumed the conduction band is normally empty of electrons. This is never entirely true, and for many semiconductor materials used in

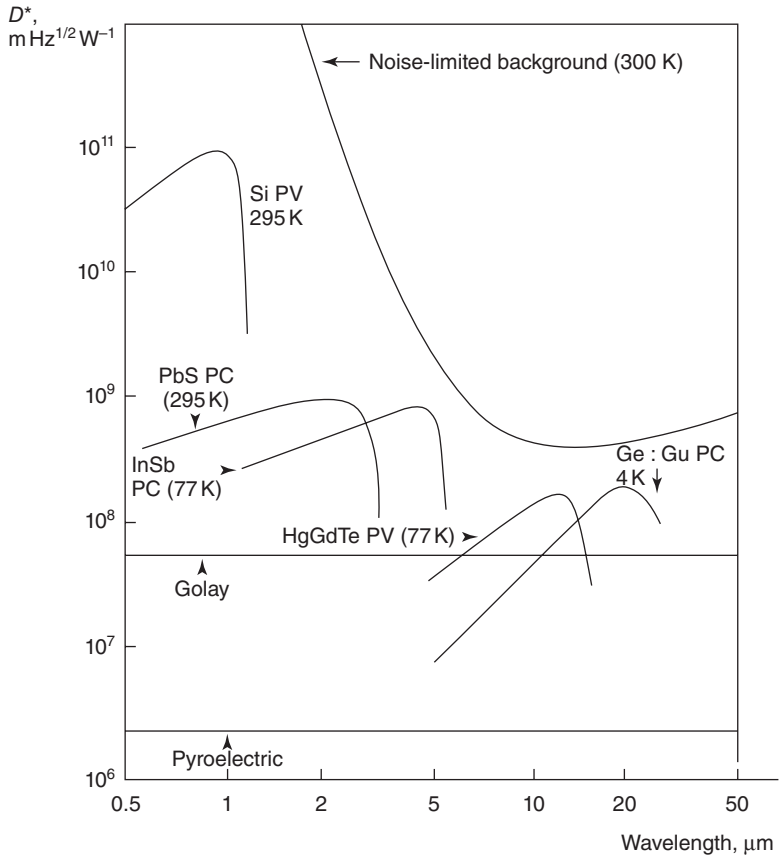


Figure 5.7 *Response curves of various detector materials (Reprinted with kind permission of Prentice Hall International).*

photoconductive devices with reasonably large values of λ_g, it is not even approximately true. Where kT is an appreciable fraction of ΔE_g there will be significant thermal promotion of electrons into the conduction band, and consequently a high detector dark current. For this reason, many photoconductive detectors are cooled, often using thermoelectric devices (for easy packaging) in order to reduce kT, and hence the dark current.

Second, not all electrons generated by incident photons are swept out into the detection circuit by the bias voltage. In fact there is a continuous dynamic process of electron-hole pair generation by incident photons, and electron-hole pair recombination via a variety of lattice relaxation mechanisms. The detector noise associated with this dynamic generation-recombination process (even assuming the detector is operated at a temperature where $kT << \Delta E_g$ so that thermally generated processes can be neglected) is known as generation–recombination noise, and is the dominant noise contribution in photoconductive devices.

The difference between a photoconductive detector and a photodiode detector lies in the presence of a thin p-doped layer at the surface of the detector element, above the bulk n-type semiconductor. Holes accumulate in the p-layer, and electrons in the n-type bulk, so between the two there is a region with a reduced number density of carriers, known as the depletion layer. The important effect of this is that electron-hole pairs, generated by photon absorption within this depletion layer, are subjected to an internal electric field (without the application of an external bias voltage) and are automatically swept to the p and n regions, and

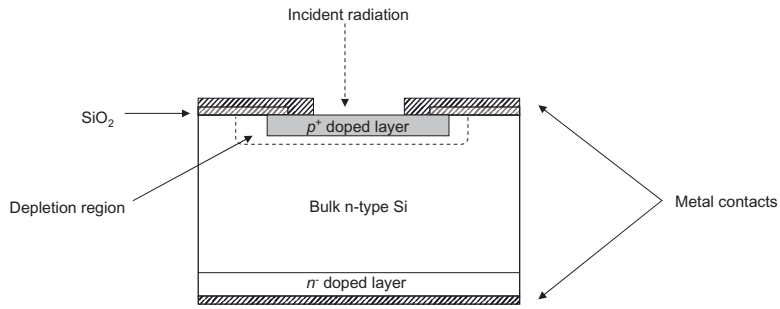


Figure 5.8 *Schematic of a photodiode p-n junction.*

a measurable external voltage will appear at the electrodes (Figure 5.8). The signal is measured as the voltage across an external load resistor in parallel. This situation represents a photodiode detector operated in photovoltaic mode.

The most common photodiode materials applied to NIRS are Si, Ge and indium gallium arsenide (InGaAs). The latter alloy semiconductor has the significant advantage that the band gap can be tuned for different wavelength cutoffs, and hence for different applications, by varying the alloy composition. For example, the basic InGaAs photodiode produced by epitaxial deposition on an InP substrate is constrained to an alloy composition of In (53%)/Ga (47%) in order to retain lattice-matching with the substrate. This has a cutoff around 1700 nm, and very high sensitivity even at ambient temperature. In order to achieve different alloy compositions, graded lattice structures must be deposited, absorbing the lattice stresses until the main active diode layer can be formed at, say, In (82%)/Ga (18%) giving a much longer wavelength cutoff at about 2600 nm. The narrower band gap requires this photodiode to be thermoelectrically (TE) cooled for optimum performance.

5.3.1.3 Detector Performance Characeristics

The fundamental performance parameter of any detector is its noise equivalent power (NEP). This is simply the input irradiance power necessary to achieve a detector output just equal to the noise. This NEP is dependent on a number of detector and signal variables such as modulation frequency and wavelength (the input signal is defined as sine wave modulated monochromatic light), and detector area, bandwidth and temperature.

In order to eliminate some of these variables in comparing detector performance characteristics, it is usual to use a parameter described as specific detectivity (D^*) and defined as:

$$D^* = (A\Delta f_{\mathrm{m}})^{1/2} / NEP \tag{5.9}$$

D^* depends on the test wavelength and the modulation frequency, and should ideally be quoted in some such form as $1 \times 10^{10}\,\mathrm{mHz^{1/2}W^{-1}}$ (1000 nm, 1 kHz).

Fortunately, for most photon device photodetectors, the dependence on modulation frequency is fairly flat across most of the region likely to be encountered with conventional NIR analyzer technology. Provided the D^* is satisfactory at the required measurement wavelengths, it is one useful parameter for detector selection.

5.3.2 The scanning grating monochromator and polychromator diode-array

5.3.2.1 Principles of Operation: the Diffraction Grating

The diffraction grating monochromator[15] is a specific example of multiple beam interference effects. Interference between multiple beams can be generated by both division of amplitude (as in the Fabry–Perot interferometer) or by division of wave front (as in the diffraction grating). (Figures 5.9 and 5.10)

The two examples shown demonstrate how an interference effect can be produced either by amplitude division of an incident beam, followed by retardation (achieved in this case by multiple reflection between the partially reflective parallel plates of a Fabry–Perot resonator) and recombination, or by division of the wave front at the multiple equally spaced slits of a diffraction grating, again followed by recombination.

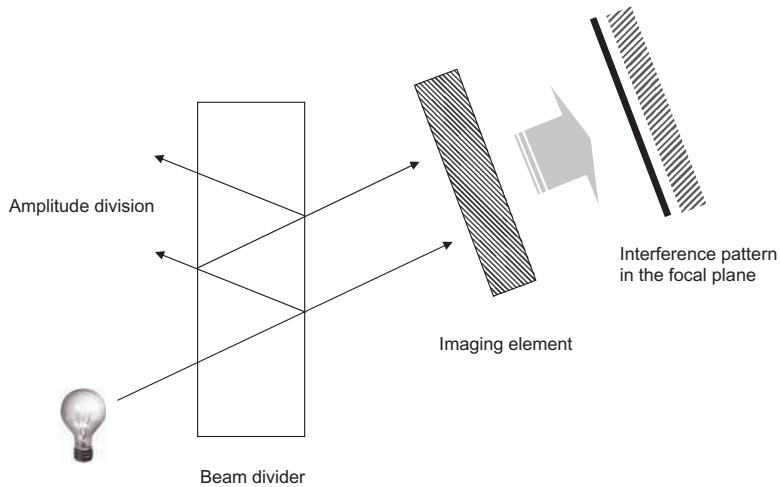


Figure 5.9 *Multiple-beam interference by division of amplitude.*

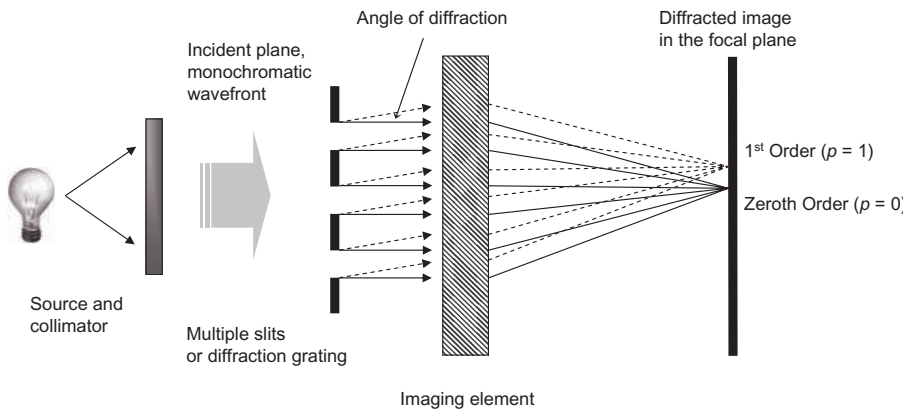


Figure 5.10 *Multiple-beam interference by division of wavefront.*

The condition for the formation of an interference maximum at the point of recombination is that the contributions to the total irradiance at that point from all the slits of the grating should be in phase. This condition is met when the classic diffraction equation for normal incidence is satisfied:

$$p\lambda = (a+b)\sin\theta = d\sin\theta \tag{5.10}$$

Where in this expression a and b are the width and spacing between the slits of the grating, and hence d (= a + b) is defined as the grating constant. Here d and λ are measured in the same units of length. The diffraction order is given by p, an integer value.

Note that for a specific wavelength, λ, interference maxima will occur at a series of integer values of p, the order of diffraction, including $p = 0$. As λ varies, so the position, defined by θ, of the interference maximum varies. So for an incident polychromatic beam, a series of interference maxima for each wavelength is formed at varying θ. If the range of wavelengths in the incident polychromatic beam is large enough, then for certain ranges of θ, there will be contributions from different diffraction orders superimposed. From this it follows that separation of the diffraction orders will be necessary to avoid excessive stray light contributions to the detected signal. Note also that all wavelengths are coincident at zero order, so this signal must also be isolated in some way to avoid a scattered light contribution to the detected signal.

5.3.2.2 Practical Aspects of Operation

The current practical realization of the diffraction grating monochromator for process analytical technology using NIRS takes the form of a concave holographic grating. In this the diffraction effect of a grating and the image-forming effect of a concave reflector are combined. Thus compared with a conventional plane-grating arrangement (e.g. a Czerny–Turner type monochromator) there is no requirement that the incident and diffracted beams are collimated. Holographic gratings are formed from a grating blank with a photoresist coating. The photoresist is subjected to an holographic image of interference fringes formed from a rigorously monochromatic source (a laser). The combination of holographic imaging and the curved surface of the photoresist layer on the grating blank are capable of exactly reproducing the hyperbolic grating pattern first described by Rowland in the late 19th century. Holographic gratings are effectively capable of producing a diffracted image free from optical aberration and very low in scattered (stray) light. Also the optical (nonmechanical) means of production lends itself very well to repeatable manufacturing techniques, so that the performance and characteristics of the gratings installed in a set of analyzers can be made reasonably reproducible.

The concave grating format (Figure 5.11) also allows the analyzer design more easily to be optimized for focal length and optical throughput. The longer focal length gives easier, more stable alignment, but raises f-number; however the larger grating diameter possible for holographic gratings decreases f-number for the same resolution and slit area and recovers throughput. Also, and significantly from the point of view of manufacturing design, the use of a concave grating reduces the number of optical components involved, and therefore the number of optical surfaces generating stray light and imaging errors, as well as manufacturing complexity and the risk of misalignment.

One further point to mention about the design and use of concave holographic gratings concerns the exit focal plane. For a scanning monochromator it is of no great concern that the focal plane is not necessarily flat since the focal point can be maintained in position by the rotation of the grating. However, this is an issue for the polychromator photodiode array based analyzer since the linear photodiode array is of necessity normally fabricated as a flat device, and all the exit wavelengths are intended to be imaged simultaneously across the detector plane. Therefore the polychromator should produce a more-or-less flat image plane at

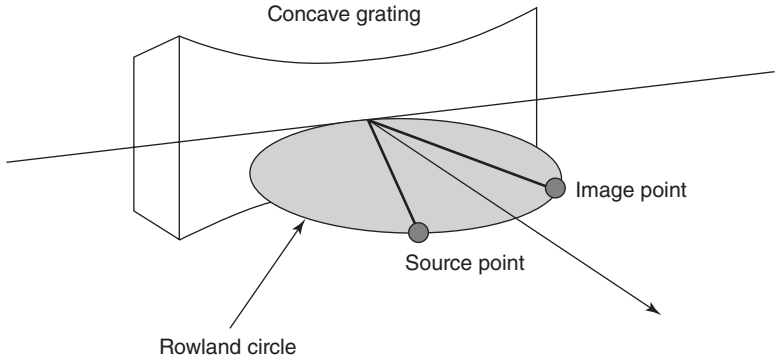


Figure 5.11 *The concave grating monochromator.*

the detector surface. In the case of holographically produced concave gratings the shape of the focal curve can be modified, and can be made to approach a flat focal plane.

In an analyzer based on scanning monochromator technology, one should be aware of a number of critical design issues:

- There is a direct compromise between spectroscopic resolution and optical throughput. In order to achieve wavelength separation at the detector, both input and output beams to and from the grating need to be imaged at entrance and exit slits, and the image formed at the exit slit refocused onto the NIR detector. In order to achieve an appropriate level of optical throughput the slits need to be sufficiently wide. In order to achieve adequate optical resolution (a narrow instrument line shape) the slits need to be sufficiently narrow. In practice the entrance and exits slits are normally set to the same size, which is selected depending on whether optical resolution or SNR is the more important parameter for a specific application.
- Another point of practical significance is the issue of rotation of the grating to achieve reproducible wavelength scanning and accurate wavelength registration. This has been the Achilles' heel of scanning monochromator design. Originally it depended upon purely mechanical scanning mechanisms. It has been greatly facilitated by the use of miniature industrial electric motor drives and optical encoding (position sensing) of the grating. This is particularly important in terms of the development of complex full-spectrum partial least squares (PLS) or similar calibration models for the more difficult applications, which require subnanometer wavelength stability over an extended period. For example, the calibration stability of applications involving refinery hydrocarbons, particularly middle distillates or heavier products, can be seen to require a frequency reproducibility between analyzers of at least $0.1\,cm^{-1}$ at $5000\,cm^{-1}$. This is equivalent to about 0.05 nm at 2000 nm – a demanding objective for a scanning device.

5.3.2.3 Order Sorting

The function of the diffraction grating monochromator is to physically separate wavelengths along the focal plane of the concave grating. As seen above, for a broadband source different wavelengths diffracted at different diffraction orders might be coincident at the detector. For example, if the scanning monochromator is required to scan a NIR spectrum between 2000 and 2500 nm, in the first order, then there will be coincident radiation from the source hitting the detector at the same time from the range 1000–1250 nm in the second order, which is certainly not negligible in terms of typical source output.

The effect of multiple-order diffraction, if not corrected, is to introduce a stray light signal at the detector, which badly affects the measured linearity of the required absorbance signal. In the example above, if the required measurement is at 2000 nm, and only 0.1% of stray light at 1000 nm reaches the detector, then appreciable absorbance nonlinearity is seen at absorbances above 2.0 AU.

In general the first order diffraction is the more intense, so is normally used for NIRS; occasionally the second order is used also to extend the available spectral range for a given physical grating rotation. This situation requires that order-sorting band-pass optical filters are introduced into the beam, normally just before the detector, to eliminate unwanted diffraction orders. A stepper motor or similar arrangement will introduce or remove specific band-pass filters for certain wavelength scan ranges.

Multiple diffraction orders are not the only potential source of stray light in a scanning monochromator. Internal reflections and scattering from the surfaces of optical components, including the grating, any additional focussing optics, order-sorting or blocking filters, detector windows and any 0th-order diffraction may all contribute.

5.3.2.4 The Polychromator Photodiode array (PDA) Analyzer

There is a great deal in common, in terms of fundamental optical principles, between a scanning monochromator and a photodiode array polychromator.[16,17] For example, both commonly use a concave grating as the dispersive element. Operationally they are quite distinct, however. The PDA polychromator exploits the semiconductor fabrication capabilities of photodiode manufacturing. By fabricating a linear array of photodiodes, an assembly can be produced which sits at the focal plane of the concave grating exit beam, and for which each individual photodiode element records a specific wave band in the spectrum. No exit slit is required, and no physical scanning mechanism is involved. Just the electronic accumulation and data logging from each detector element are required. There does, however, need to be a good match between the diode array size, the linear dispersion of the grating, the bandwidth of the detector and the correct order sorting filter arrangement. Since the whole PDA is subjected to a dispersed image of the entrance slit (or input fiber optic), it is even more important than in the scanning monochromator that a correct order sorting filter, working in conjunction with the spectral response range of the detector, is used to block stray light from irrelevant diffraction orders.

Given the high D^* of Si or thermo-electrically cooled InGaAs photodiodes, and the multiplex advantage of simultaneous multichannel detection, it is clear one can expect such devices to have high SNR with fast data acquisition times. One can anticipate a potential SNR enhancement of $n^{1/2}$ where n is the number of resolution elements measured by the PDA compared with a scanning monochromator over the same data acquisition time. However the linear array detector elements are often much smaller than a typical single detector, and there may be a significant throughput limitation.

There are in general two classes of noise associated with NIR analyzer systems: additive and multiplicative noise. Additive noise is mostly wavelength-independent, and does not scale with the signal strength. It is in effect the white noise from the detector and associated read-out electronics, which is added to the pure signal. In PDAs the major contribution is the generation / recombination noise associated with the photoproduction of carriers. Multiplicative noise on the other hand arises from almost anywhere else in the optical train and is mostly associated with instabilities of various sorts – source instability, mechanical vibrational effects on gratings, slits, apertures, fiber optics, mirrors, etc. This type of noise scales with the signal, and is often not white noise, but may have significant wavelength dependency.

In a polychromator PDA spectrometer there is no exit slit – all diffracted wavelengths (after order sorting) fall onto the diode array detector. In this case what defines the analyzer spectral line shape? Each individual detector element integrates the signal falling onto it, and allocates that signal to a specific wavelength. In reality, of course, it is seeing a band pass of wavelengths, depending on the angular dispersion of the grating,

and the flatness of the focal plane. For standard longer wavelength NIR applications – which tend to predominate in real-life process analytical technology situations – the PDA will be a thermo-electrically cooled InGaAs assembly. These are costly, especially if a large number of detector elements is required. Consequently most PDA analyzers working in this region will tend somewhat to compromise resolution and match the detector element size to a fairly wide entrance slit image. Depending on the exact layout of the detector pixels, this can produce a variety of instrument line shape functions, and for real spectral data recording can also result in line shape asymmetry. This brings to mind one of the key operational difficulties of a PDA spectrometer. Wavelength calibration is absolutely necessary, and the variation in grating linear dispersion across the whole range of measurement (and hence a changing overlap between slit image and pixel size) leads to some difficulty in this case. In principle, the PDA device can be built in a reasonably monolithic form so that the influence of the external environment, mechanical and thermal, can be reduced. But for effective transfer of applications and calibrations between analyzers, an absolute wavelength accuracy better than 0.05 nm (at 2000 nm) may be required, and that is currently a challenging target for PDA devices.

Due to the small size and applied bias voltage (when used in photoconductive mode) the response time of each diode element in the PDA can be very fast indeed, so total integration times can also be short. In principle, data rates of 200 spectra per second are achievable, but in reality spectral co-adding is a more sensible objective, to improve overall SNR.

5.3.3 The acousto-optic tunable filter (AOTF) analyzer

5.3.3.1 Principles of Operation

Both the diffraction grating devices described above make use of a physical diffraction grating, where the diffraction takes place at or from a simple plane or concave surface. Thus all incident wavelengths are diffracted at some angle or other, and either the exit slit, or the discrete elements of the photodiode array are needed to sort out which wavelength is which. In contrast to a surface grating monochromator, an AOTF device[18] operating in the Bragg regime diffracts only a single wavelength (or at least a very narrow band pass of frequencies) for a given tuning condition. This is therefore in principle a full aperture device – input and exit slits are not required. However there are throughput limitations of a different sort,[19] which will be discussed.

The heart of an AOTF analyzer is an NIR transmitting, optically anisotropic (birefringent) crystal lattice, normally TeO_2 (tellurium dioxide) or $LiNbO_3$ (lithium niobate). A piezotransducer is cold-metal bonded to one surface of this optically polished crystal. An RF driver is used to generate a high-frequency electric field that causes a coupled high-frequency vibration of the transducer. Since this is closely bonded to the AOTF crystal, this in turn is subjected to an acoustic wave, which propagates through the crystal, and is normally absorbed at the far surface of the crystal to avoid undue acoustic reflections. Another way to describe the impact of the RF driven piezotransducer on the AOTF crystal is as a strong source of lattice phonons which propagate through the crystal with a more or less fixed wavelength (Λ) determined by the RF frequency of the oscillator, and a flux dependent on the RF power.

Thus the incident NIR beam is a source of photons, and the energy from the piezo-transducer provides a source of lattice phonons that propagate through the crystal. The resulting acoustically scattered photons are Doppler shifted up or down in frequency. This is a minor effect. But there is a significant effect on the scattered (diffracted) beam direction. For a given input wavelength, and given K, there will be just one scattering angle given by:

$$\Phi_B = \sin^{-1}(K/2k_0) = \sin^{-1}(\lambda/2\Lambda) \tag{5.11}$$

where K and k_0 are the relevant wave vector magnitudes, λ is the wavelength of incident light, and Λ is the wavelength of the sound.

The alternate model of the Bragg diffraction regime is similar to the case of X-ray diffraction in a crystal lattice. The diffracted beam is extensively re-diffracted inside the crystal (Figure 5.12) due to its thick optical path length, and the diffraction takes place at a series of planes, rather than the lines of a surface grating, leading to the same Bragg angle equation as that derived above from photon-phonon scattering arguments.

$$\sin \Phi_B = (\lambda/2\Lambda) \tag{5.12}$$

where $\theta_I = \theta_d = \Phi_B$ for phase matching.

Note that this condition is extremely wavelength selecting. For a given angle of incidence between the input optical beam and the acoustic wave front, and a given value of the acoustic wavelength, only one value of λ is permitted.

The Bragg angle can be calculated for reasonable conditions. For TeO_2 the acoustic velocity is around $650\,m\,s^{-1}$, so at an RF frequency of 50 MHz, the phonon wavelength is approximately 1.3×10^{-5} m, and the Bragg angle of deflection for an incident beam of 1000 nm light is 40 milliradians (2°). Note that an increase in RF driver frequency to 100 MHz would halve the acoustic wavelength, and the Bragg angle for the same incident wavelength would increase to approximately 4.5°.

So far a modulator device has been described, not a filter. In order to understand how an AOTF device goes further than this, it is necessary to look at the effect of Bragg regime diffraction and the impact of the anisotropy of the AOTF crystal medium.

In an AOTF analyzer the input optical beam to the crystal is polychromatic from a broadband source. For a given acoustic frequency and angle of incidence, only a very narrow band pass of wavelengths from the

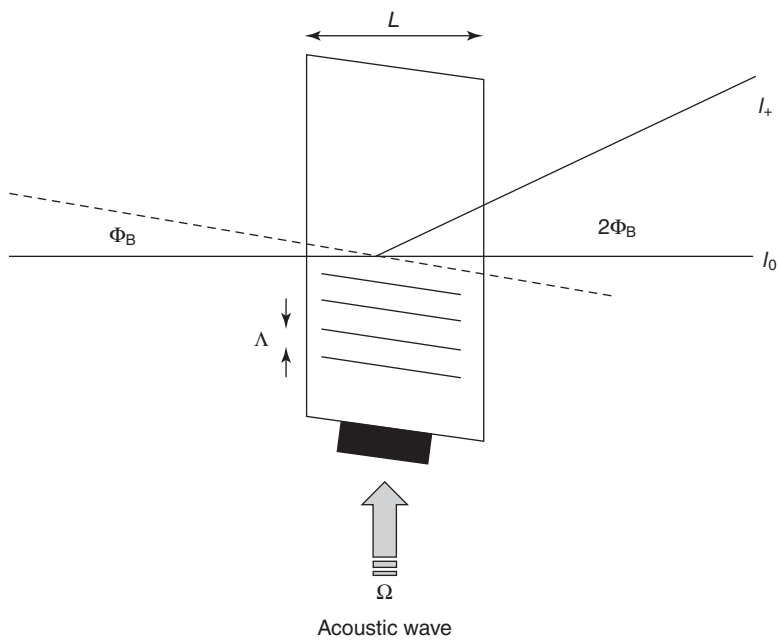


Figure 5.12 *Bragg-type AO diffraction.*

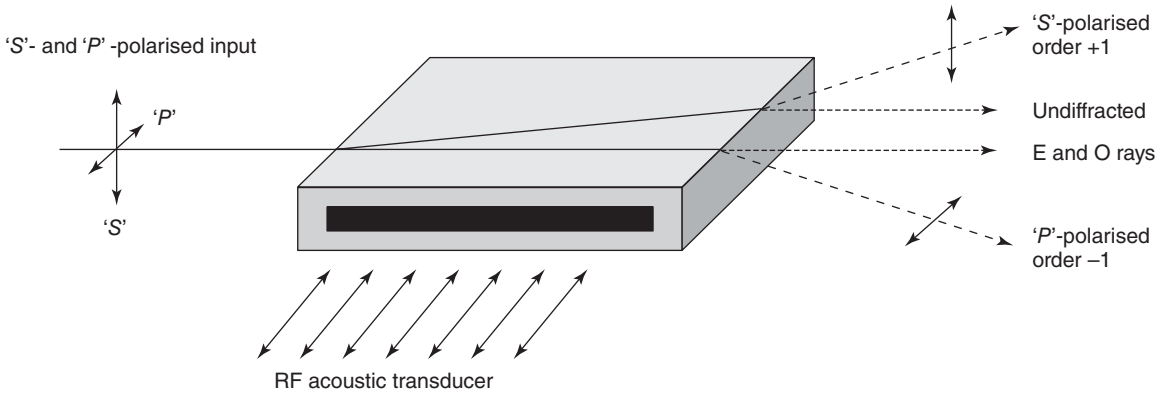


Figure 5.13 *AOTF with polarization rotation.*

polychromatic input beam will satisfy the condition for anisotropic Bragg diffraction. That band pass is angularly deflected *and* its polarization is rotated through 90 degrees. The remainder of the polychromatic input beam carries straight on through the crystal, and needs to be separated from the monochromatic diffracted and deflected beam. Above, one saw that the values of the angular deflection in Bragg regime acousto-optic diffraction are not particularly large. Consequently the polarization rotation of the diffracted beam is a very useful feature in AOTF devices, because it allows for a greater chance to separate the monochromatic, diffracted beam from the polychromatic nondiffracted beam using a combination of physical beam blockers and crossed polarizers (Figure 5.13).

The raw band pass of an AOTF has a sinc squared function lineshape with sidebands, which if ignored, may amount to10% of the pass optical energy in off-centre wavelengths. This apodization issue is normally addressed by careful control of the transducer coupling to the crystal.

5.3.3.2 Practical Aspects of Operation

The very specific nature of the acousto-optical effect requires one to consider some particular issues regarding device performance. These include:

- thermal conductivity of the AOTF crystal
- the maximum input aperture width to avoid band-pass compromise
- the full optical acceptance angle
- diffraction efficiency
- wavelength stability and the temperature-tuning coefficient
- blocking efficiency
- RF driver tuning range.

Quite a large amount of input optical power is being channeled through the AOTF device, as well as the acoustic wave power being coupled in via the piezotransducer. The AOTF crystal must have a high thermal conductivity to allow this power to dissipate without heating up the crystal. This is the main reason that TeO_2 has become the standard material for NIR AOTF, since it has a wide transparency from approximately 370 nm up to 3500 nm. However, the useable wavelength range for any particular device setup is normally

more limited, both by the detector range (typically 900–2600 nm for a thermo-electrically cooled InGaAs detector), and more importantly by the frequency drive range of the RF oscillator. At the level of stability required this can rarely exceed one octave. So a typical AOTF device might have a tuning bandwidth from 900–1800 nm, 1000–2000 nm, or 1200–2400 nm.

Although there are no slits associated with an AOTF device, there are some necessary compromises regarding throughput. First, there are physical restrictions on the size of available crystals, and in the case of the normal noncollinear arrangement where the input beam direction is traversed by the acoustic wave front, then the open beam aperture needs to be limited so that the transit time of the acoustic wave front across the crystal is limited, in order that a narrow optical band pass can be maintained. Typically this is of the order of a few millimeters.

It has also been shown that the diffraction condition requires control of the incident beam angle. Normally a collimated beam is incident on the AOTF crystal, but there will be some beam divergence from a real non-point source. The full acceptance angle is often in the range of a few degrees only; so input beam divergence must be limited. It is for these reasons that AOTF analyzers are very often designed as fiber-optic coupled devices. The small input beam diameter is very well matched with a fiber optic's input, although the relatively high NA of a fiber optic makes beam collimation within the full acceptance angle of the crystal difficult.

The undiffracted beam, which retains all the other wavelengths which have not been coupled into the narrow band pass of the diffracted beam, is still present, and only separated from the emerging diffracted beam by a small angle and its retained polarization, which will now be at 90 degrees to that of the diffracted beam. An important practical consideration is how well a combination of physical beam blocking and crossed polarizers will eliminate the untuned, out-of-band wavelengths, which represent stray light. Again the use of fiber-optic coupling is one way to help separate these small diameter exit beams. There are some design complications here, because multimode optical fibers do not normally preserve input polarization, so any polarization selection must be done after the input fiber and before the output fiber connections. The typical spectral resolution for an AOTF device is >5 nm at 2000 nm (roughly 12 cm^{-1}), so is comparable to a scanning grating analyzer.

Finally, and possibly the most important practical issues for an AOTF device are those of temperature control and wavelength stability. As in the case of the grating polychromator, there is no inherent calibration for wavelength in the device. For any particular device for a known angle of incidence and RF frequency, the Bragg wavelength is determined in theory, but is dependent in fact on manufacturing tolerances and operational environment. We saw earlier that for demanding NIR applications, a wavelength accuracy of at worst 0.05 nm at 2000 nm is typically required.

The temperature tuning coefficient of TeO_2 is 0.025 nm/°C, so it is clear that in order to combat the effects of thermal heating of the crystal due to incident optical and acoustic power, as well as environmental considerations, a good level of thermal control is needed. This can be in the form of a thermo-electrically cooled enclosure. This will deal to a certain extent with the need to achieve analyzer wavelength repeatability. However, it does not address the issue of wavelength reproducibility between AOTF modules, which impacts the transportability of calibrations and datasets between analyzers.

A key operational advantage of an AOTF device lies in the very rapid and random access wavelength selection possible by programming the RF driver frequency. For example, once an application is defined, an isophotonic scanning regime can be defined, so that data acquisition is extended in spectral regions with high sample absorptivity, helping to give a uniform SNR across the spectrum. Similarly, combinations of discrete wavelengths can be accessed with a switching time below 10 μs. This, along with the low voltage power requirements of the RF oscillator and tungsten lamp source, lend the unit to miniaturization and remote location. Thus AOTF systems have been used very effectively in 100% whole-tablet analysis applications and oil seed content discrimination applications on production lines.

5.3.4 Fourier transform near-infrared analyzers

5.3.4.1 Principles of Operation: Modulation and Wavelength Encoding

So far, all devices for NIRS that have been considered depend on some form of wavelength selection. This ranges from the physical selection by slits in a scanning grating analyzer, through the selection by discrete detector elements in a PDA polychromator, to the electronic tuning selection of an AOTF device. For all of these devices, wavelength identification, and therefore spectral repeatability and reproducibility, is a design issue, albeit one for which many elegant and effective solutions[20] have been found.

Fourier transform near-infrared analyzers do not depend on any form of wavelength selection, but instead provide a method of wavelength encoding, so that all transmitted wavelengths may be measured simultaneously at a single detector. Moreover the mechanism by which the wavelength-dependent modulation is introduced into the DC NIR source output is itself controlled by a very precise wavelength-registered mechanism (a reference laser channel) – so that inherent frequency calibration of the NIR spectrum is achieved. The situation is not, of course, completely plain-sailing. To achieve this objective quite demanding issues of optical alignment and integrity need to be maintained, but again elegant and effective solutions have been found,[21–23] and will be described below.

Two-beam interference requires the bringing together of two at least partially coherent wave trains. In practice, the most efficient way to do this is to start with a single light source, generate two beams from this by division of amplitude, and then introduce an optical retardation into one of the beams, relative to the other, and then recombine the beams. This means that the two interfering beams are generated from the same light source, and the optical retardation introduced is small enough to be within the coherence length of the source (Figure 5.14).

For a wavelength λ, and an optical retardation δ, one expects that the amplitude of the resultant beam after recombination (the interference signal) will be

$$I(\delta) = A_0[1 + m\cos(2\pi\delta/\lambda)] \tag{5.13}$$

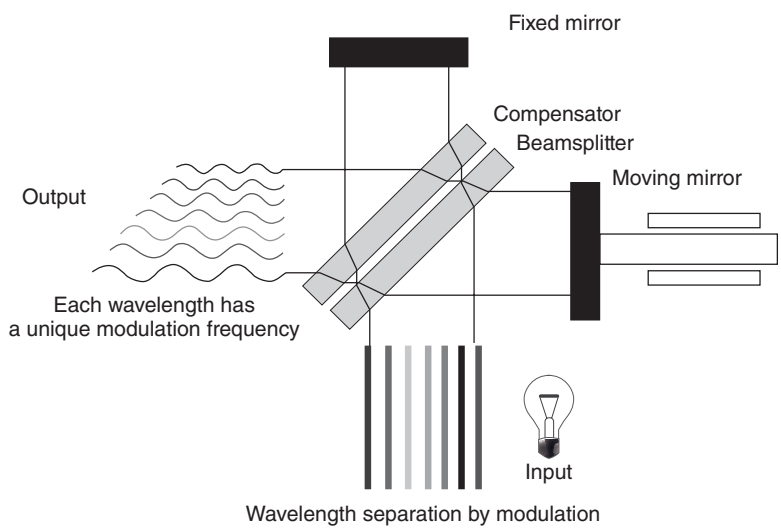


Figure 5.14 *The generation of a Fourier-modulated NIR beam.*

where A_0 is effectively the total intensity of the input beam, factored by the efficiency of the division / recombination process, and m is the modulation efficiency of the recombining beams. In the ideal case this is 1, but one will see the effect of changes in this parameter later.

Thus the AC part of this signal is

$$I'(\delta) = A_0 m \cos(2\pi\delta/\lambda) \tag{5.14}$$

Remembering that the wavenumber $\nu = 1/\lambda$, one has

$$I'(\delta) = A_0 m \cos(2\pi\delta\nu) \tag{5.15}$$

Therefore for fixed ν and a linearly changing δ, the output of the ideal two-beam interferometer is a cosine function. In other words it is a cosine modulator of the original DC light source. The modulation frequency observed in the AC output of the FTIR spectrometer detector is dependent on the rate at which δ is increased or decreased, and is given by

$$f_{\mathrm{m}} = \delta'\nu \tag{5.16}$$

where δ' is the rate of change of the optical retardation, in units of $\mathrm{cm\,s^{-1}}$.

For a typical broadband NIR source (for example the tungsten halogen lamp described previously) ν will cover a range from around 3500–12 000 $\mathrm{cm^{-1}}$, and δ' might be 1 $\mathrm{cm\,s^{-1}}$ for a typical fast-scanning interferometer, so the observed range of frequencies in the AC detector output could be 3.5–12 kHz. In other words, signals typically in the audiofrequency range.

Assuming that the rate of change of the optical retardation introduced by the interferometer is the same for all of input radiation frequencies (which is normally the case) then each individual value of ν in the broadband source output contributes a different value of f_m in the AC component of the detector output. These contributions are, of course, summed by the detector. The combined detector output, arising from the simultaneous measurement of all modulated input signals, is a sum of cosine functions (Figure 5.15). Such a sum is a Fourier series, and the amplitudes of the individual input modulations at each Fourier frequency can be recovered using Fourier decomposition. These amplitudes are in proportion to the optical intensity throughput of the spectrometer at each wavelength. This combination of modulation, summation, and Fourier decomposition constitutes the essential and basic principle of FTIR, and explains how it allows measurement

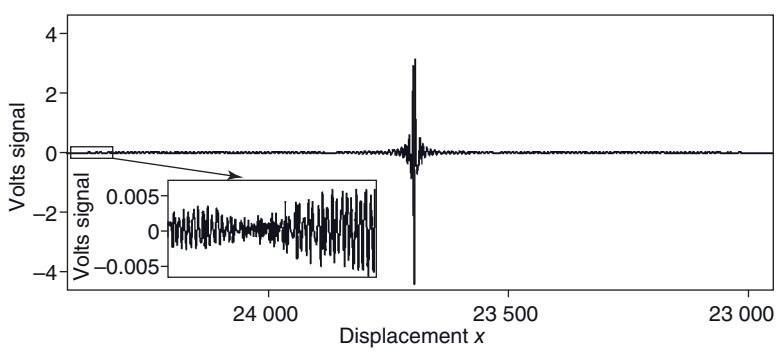


Figure 5.15 *The interferogram of a broadband NIR source.*

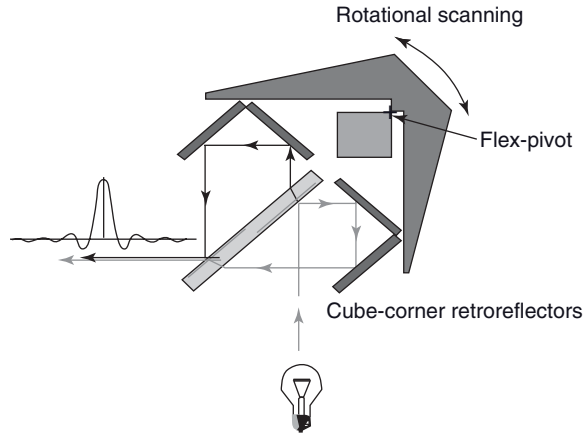


Figure 5.16 *The retroreflector rotational modulator.*

of the intensity of all source wavelengths simultaneously. The combined detector signal for all incident radiation frequencies is thus

$$I_{\mathrm{TOT}}(\delta) = \Sigma\, A_{\nu 0} m_{\nu} \cos(2\pi\delta\nu) \tag{5.17}$$

where the summation applies in principle over all wavelengths from zero to infinity, but in practice covers the throughput range of the overall source–interferometer–detector combination.

There are a number of ways in which the necessary optical retardation can be introduced within the FTIR modulator. Historically, when FTIR spectrometers were developed initially for mid-infrared operation, the classical Michelson interferometer design predominated, with one fixed plane mirror, and a linearly-scanned moving plane mirror (to provide the variation in δ). Although functional in the mid-infrared, and capable of generating very high spectral resolution, such designs are not at all suitable for industrial and process analytical chemistry applications in the near-infrared. The majority of FTIR analyzers used for industrial process analytical applications use interferometer designs based on retroreflector mirrors (cube-corner mirrors, or variants thereof) and rotational motion to achieve the variation in δ (Figure 5.16). There are also designs based on the use of a linearly mechanically scanned refractive optical wedge driven into one beam of the interferometer to achieve the necessary optical retardation.

The typical NIR source employed is the tungsten halogen lamp, as described previously, although high-power silicon carbide (SiC) sources can also be used in combination with suitable beamsplitter materials, for example zinc selenide (ZnSe) or calcium fluoride (CaF_2) for combined mid-infrared and near-infrared operation. Typical detectors are pyroelectric DTGS for general-purpose applications, and cooled photoconductive detectors such as InAs and InGaAs for low throughput or fiber-optic based applications.

5.4.3.2 Practical Aspects of Operation: the Advantages of FTIR-based Analyzers

It is a commonplace that FTIR-based analyzers are the predominant technology for mid-infrared applications. This arises from a unique tie-in between the inherent advantages of the FTIR method and serious limitations in the mid-infrared range. The most serious problem for mid-infrared spectroscopy is the very low emissivity of mid-infrared sources combined with the low detectivity of mid-infrared thermal detectors.

This causes a direct and severe conflict between the desire for good SNR and optical resolution, when any form of wavelength-selective (low throughput) analyzer design is employed.

The classical FTIR advantages are based on: (1) geometry (the Jacquinot, or throughput advantage), (2) statistics (the Fellgett, or multiplex advantage) and (3) physics (the Connes, or wavelength advantage).

1. The Jacquinot advantage arises because of the lack of significant optical throughput restrictions in the optical layout of an FTIR analyzer. There are no slits or other wavelength selecting devices, and an essentially cylindrical collimated beam from the source travels unhindered through the beam paths to the detector. There is one proviso to this, in that a circular aperture Jacquinot-stop (J-stop) must be included at an intermediate focal plane to allow proper definition of achievable optical resolution, but the effect on throughput is minor. In fact, for many FT-NIR analyzers the source, the sample device itself or the detector element act as the J-stop and no additional throughput restriction occurs.
2. The Fellgett advantage is based on the simultaneous modulation of all incident wavelengths, which allows the FTIR analyzer to measure all wavelengths all of the time. For a total data acquisition time of T, a wavelength selective device measures each wavelength for only T/m, where m is the number of resolution elements (determined by the slit width), whereas in the FTIR analyzer each wavelength is measured for the whole of time T. If the detector noise is the dominant noise contribution, this allows a SNR advantage of $m^{1/2}$.
3. The Connes advantage arises from the fact that the AC modulation frequency introduced into each wavelength of the source output is in effect measured using a very high wavelength-precision and wavelength-accuracy device (a HeNe laser). The single-frequency interferogram signal from a HeNe laser is used as an internal clock to both control the rate of change of δ, and to trigger the data acquisition of the resulting modulated NIR detector signal.
4. There is in addition a fourth, unnamed FTIR advantage, in that FTIR analyzers are immune to stray-light effects. The detector is acting in effect as a lock-in amplifier, looking only at modulated signal in the audio-frequency range (those signals having been generated by the FTIR modulator). Thus incident external DC light components have no impact.

It is easy to see that for mid-infrared operation the advantages listed (1) and (2) above have an absolutely overriding significance. They produce by themselves a decisive advantage for the FTIR technique in mid-infrared applications. The situation is not so clear at first sight in the near-infrared, but careful consideration also shows here that FT-NIR-based analyzers can demonstrate significant advantages over alternative technologies.

The argument for the applicability (or lack of it) of the above factors in FTIR-based near-infrared applications centres on the issues of optical throughput and detector responsivity. Near-infrared sources are comparatively very bright, and detector responsivities may be many orders of magnitude better than in the mid-infrared. Hence in order to avoid detector saturation when operating in open-beam mode with no additional attenuation due to sample or sampling devices, an FTIR-based analyzer using a photoconductive detector will need the near-infrared source to be significantly attenuated. The issue of SNR then becomes limited by the digital resolution of the fast analog-to-digital circuit (ADC) used to capture the detector signal, with the open-beam optical throughput and detector gain adjusted so that the detector dark noise is just providing noise to the least significant ADC bits for co-adding. However, this situation can be manipulated to advantage.

First, the availability of high optical throughput in an FT-NIR analyzer means that lower-cost, more robust and also more linear DTGS detectors can be used routinely. These in turn allow (through the choice of suitable beamsplitter materials and source characteristics) a wider range of operation than the conventional

pure near-infrared, including for example combined near-infrared and mid-infrared operation in the same analyzer in the same scan.

Second, many real-life situations in process analytical technology applications are in fact severely throughput limited. This might include multiplexed applications using fiber-optic interfaces, highly scattering samples (tablets, powders, slurries), highly absorbing samples (crude oil, bitumen), etc. In these circumstances the application can quickly become detector noise limited, and the full power of the Jacquinot and Fellgett advantages apply.

Finally, however, it is the advantages (3) and (4) that really have a marked effect on NIR applications. The wavelength precision and reproducibility achievable with a good FTIR analyzer design translates into long-term spectroscopic stability both within one analyzer, and across multiple analyzers, so that measurement errors between analyzers are in the region of a single milliabsorbance unit. For the untroubled development of near-infrared calibration models, and their maintainability over time, this of itself is critically important. Probably the best example of how this advantage has played out in a widespread application of FT-NIR technology for process optimization lies in the area of refinery hydrocarbon process stream monitoring, for which FT-NIR has emerged as the most consistently successful technology. A more detailed explanation is given in Section 5.5.

5.4.3.3 The Implementation of FTIR Technology for Near-infrared Process Analytical Applications

The achievement in practize of the theoretical advantages of FTIR technology as described above, especially for near-infrared process analytical applications, is a demanding task. Although essentially a simple device, the FTIR modulator is very unforgiving in certain areas. For example, failure exactly to control the relative alignment of beamsplitters and mirrors during the scanning of optical retardation, and the divergence of the collimated input beam from the NIR source both lead to degradation in achievable optical resolution and error in absorbance band shapes and wavelengths. These are important errors, especially in quantitative applications, and their presence would severely limit the usefulness of FT-NIR, were no ways available to control them.

An approximate calculation for the differences in optical retardation for the central and the extreme ray of a beam-path which is diverging as it travels through the FTIR modulator, separated by a half-angle (in radians) of α, shows that the maximum allowable value of α to achieve a required resolution $\Delta\nu$ at a wavenumber of ν_{max} is:

$$\alpha_{max} = (\Delta\nu/\nu_{max})^{1/2} \tag{5.18}$$

More seriously, the effect of the above beam divergence also impacts wavenumber accuracy, and the deviation in measured wavenumber for the extreme ray versus the central ray is approximated by:

$$\nu' = \nu(1 - \tfrac{1}{2}\alpha^2) \tag{5.19}$$

Taking an example of a measurement at $5000\,cm^{-1}$ with a required resolution of $2\,cm^{-1}$ but a wavenumber accuracy requirement of $0.04\,cm^{-1}$ (which are entirely typical of the requirements for a demanding near-infrared quantitative application) – the limit on α_{max} is about 0.8° for the resolution, but around 0.2° for the wavenumber accuracy.

However, the most significant practical issue for the implementation of an FT-NIR device suitable for quantitative process analytical applications arises from the need to maintain a very high degree of alignment at the point of recombination of the divided interferometer beam paths –at the exit of the interferometer.

In Equation 5.17 the output signal from an FTIR modulator was described as:

$$I'(\delta) = A_0 m \cos(2\pi\delta\nu) \tag{5.20}$$

where m was described as the modulation efficiency, and should ideally have a value of 1.

Figure 5.17 shows how the modulation efficiency of the FTIR device is affected when the recombining exit beams are no longer parallel and coincident. The effect is that the theoretically uniform irradiance across the circular exit aperture, at any given value of optical retardation, is broken up into a series of fringes,[24] and the detector signal is then an average over these fringes. This amounts to a loss of contrast in the interferogram signal, introducing the effects described above for beam divergence. This misalignment of the exit beams is precisely what will happen in case there is any tilt misalignment of plane mirrors used in a standard Michelson arrangement. For mid-infrared FTIR devices, especially those required to operate at very high optical resolution (and therefore requiring large amounts of optical retardation) successful solutions have been found using dynamic alignment systems in which the scanning plane mirror is dynamically tracked and realigned during the scanning process. This type of arrangement is, however, less suitable for a process analytical environment, and for that type of FT-NIR application the more universally adopted and successful strategy has been to use optical layouts with inherent tilt-compensation. The most successful designs typically use a combination of cube-corner retroreflectors, and a rotational motion for the mirror scanning mechanism around a fixed balance point. This avoids the need for a linear mechanical scanning mechanism, and allows for minimal individual mirror movement, since both mirrors are moving in opposition, and the effective optical retardation is doubled.

There are of course other causes for exit beam misalignment other than simply the process of optical retardation by mirror scanning. The overall alignment integrity of the cube-corner/beamsplitter/cube-corner module as well as its optical relationship to the collimated input beam from the NIR source, and the colinear HeNe laser are all critical. The thermal and mechanical instability of this whole assembly is roughly a function of its volume. Small compact FTIR modulators will perform better in process analytical applications than those less compactly designed. Fortunately, the cube-corner modulator layout in Figure 5.17 allows for both a compact layout and tilt compensation. The cube-corner retroreflector works perfectly only when the three plane mirror sections of which it is composed are set exactly orthogonal at 90° to each other. Under that condition it also displays near-perfect optical stability with respect to temperature changes. For near-

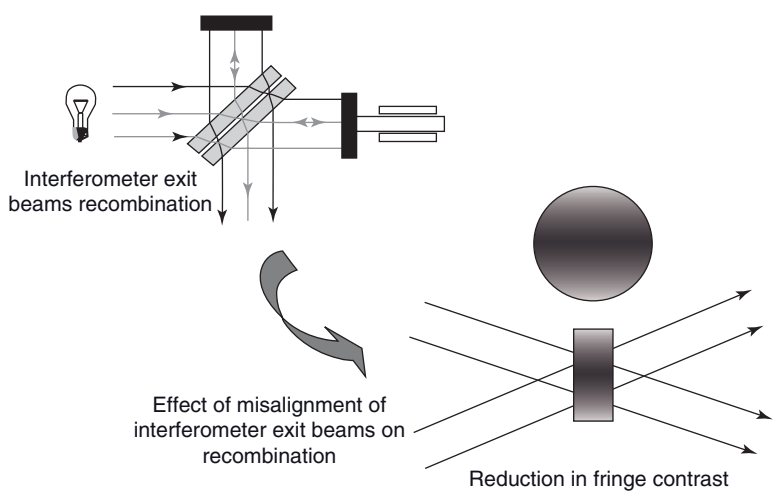


Figure 5.17 *Recombination at the exit of the interferometer.*

infrared applications the accuracy of orthogonality should be around one second of arc, which is now an achievable manufacturing specification on a routine basis.

The optimum FTIR modulator design, based around the above principles, will depend on factors such as the wavelength range of operation, the required optical resolution, and also issues such as the packaging requirements and the need to deal with field-mounting or hostile environments. This is an area of active current development with emerging possibilities and variations around the central theme of rotating retro-reflector modulator designs.

What is quite clear from current experience is that the optical and mechanical simplicity of FT-NIR analyzers, coupled with their compact implementation and appropriate packaging allow the exploitation of the advantages of FTIR devices in near-infrared process analytical applications. The current generation of industrial FTIR devices have a robustness and longevity entirely in keeping with the needs of quantitative industrial process monitoring applications, despite the very real engineering challenges described above.

5.3.4.4 Recent Developments in rugged FTIR Modules for Process Applications

We have seen that one of the key aspects of Fourier transform NIR analyzers is their control of frequency accuracy and long-term reproducibility of instrument line shape through the use of an internal optical reference, normally provided as a HeNe gas laser. Such lasers are reasonably compact, and have acceptable lifetimes of around 3 to 4 years before requiring replacement. However, we also saw how the reduction in overall interferometer dimensions can be one of the main drivers towards achieving the improved mechanical and thermal stability which allows FT-NIR devices to be deployed routinely and reliably in process applications.

The latest generation of FT-NIR interferometers are making use of developments in solid-state laser technology to implement very long-lived, stable, narrow bandwidth and extremely compact reference lasers based on quantum-well devices. These so-called VCSEL (vertical cavity surface-emitting laser) devices allow the interferometer, reference laser and NIR source to be integrated together into a metrology module with much more compact dimensions than previous FT-NIR designs (Figure 5.18). These implementations are expected to form the basis of future even more rugged platforms for industrial process FT-NIR applications.

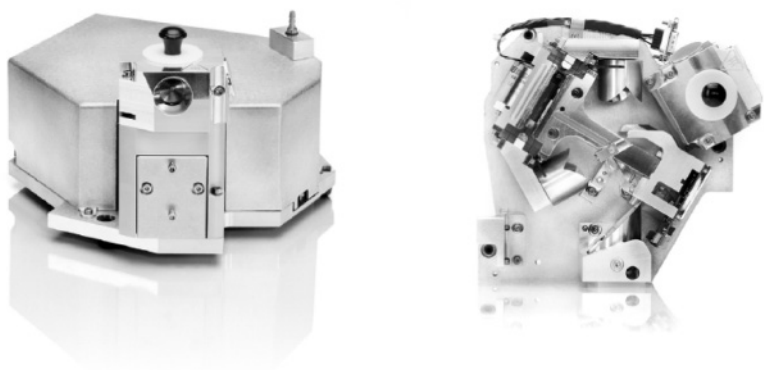


Figure 5.18 *Compact FT-NIR device with integrated VECSL laser metrology.*

5.3.5 Emerging technologies in process NIR analyzers

In Sections 5.3.1 to 5.3.4 we have reviewed the main NIR technologies in successful current widespread use for process analytical technology applications. NIR is an evolving technology, however, and there are a number of recent developments which may in time prove significant. This is not the place for a detailed review of such technologies, but a brief overview for orientation will be provided.

The common enabling technology lying behind the majority of such recent development is the use of MEMS (microelectromechnical systems) methods to manufacture extremely compact[25,26] and integrated devices. These have sometimes been referred to as 'the spectrometer on a chip'. MEMS devices are very familiar, if not very visible, in everyday life as accelerometers in passenger car air bags, as image stabilization devices in digital cameras, or as the microphones in mobile telephones. Due to the compact sizing, a variety of design possibilities open up, one of which is the use of alternative NIR light sources, compared with the common tungsten lamp source described above for the more conventional devices. Indeed smaller, brighter NIR sources would be a key ingredient in the design of any MEMS based spectrometer device in order not to sacrifice more than is necessary overall optical throughput, given the reduced surface area of the other optical elements.[27]

Examples of such new NIR sources include super-luminescent LEDs (SLEDs), in which an LED is fabricated with a MEMS waveguide to allow efficient optical coupling to the next (miniaturized) component in the analyzer optical sequence. High brightness is obtained, first by the spatial constraint of the waveguide coupling, but also by running the device at high current to achieve stimulated emission. Note this is not a semiconductor laser. In fact every effort is made in the design to avoid laser amplification in order to achieve as broadband an emission profile as possible. Even so the emission bandwidths are only in the order of 100 nm, so full-range spectroscopy does require multiple SLED devices to be used in tandem.

One elegant use of the SLED emission characteristics is to link it in one monolithic device with a tuneable filter. A Fabry–Perot filter consists of two reflective surfaces forming a resonant optical cavity, and by changing the mirror separation physically the pass band of the device can be changed. The MEMS development is to manufacture such a device on a small enough scale to be coupled to a bright SLED source. Its low mass allows the reflective surfaces to be modulated at frequencies around 100 kHz, yielding a very compact fast-scanning and mechanically and environmentally robust device. Somewhat similar methods can be used to develop miniaturized Fourier transform devices for NIR operation, but achieving throughput and resolution both present a challenge. Generally, the devices are designed with micromechanically flexed mirror assemblies, operating at their mechanically-resonant harmonic frequencies. This raises an interesting complication, because the velocity profile of the scanning mirror will then be sinusoidal, instead of the rigorously linear profile of conventional FT devices.

Although not strictly dependent on MEMS technologies, modern manufacturing methods such as lithography have brought analyzers based on Hadamard and similar discrete transform methods[28] back into the picture. A typical Hadamard device uses a physical encoding mask, coupled with a two-dimensional detector array to achieve wavelength encoding (Figure 5.19). A discrete transform, based on the characteristics of the mask (usually binary; 1 and 0 for light and dark) then recovers the spectrum from the detector signal. This approach has been developed in two interesting ways, one method using lithographic mask encoding, and the other using a MEMS active diffraction grating as (amongst other things) a programmable Hadamard mask.

The lithographic approach encodes a rotating disk mask with a circumferential sinusoidal pattern, of increasing wavelength across the radius of the disk. The (conventional) broadband NIR source beam is dispersed using a grating across the surface of the rotating mask, so that each wavelength is encoded, but this time with a sinusoidal amplitude modulation. The spectrum can then obviously be recovered with a

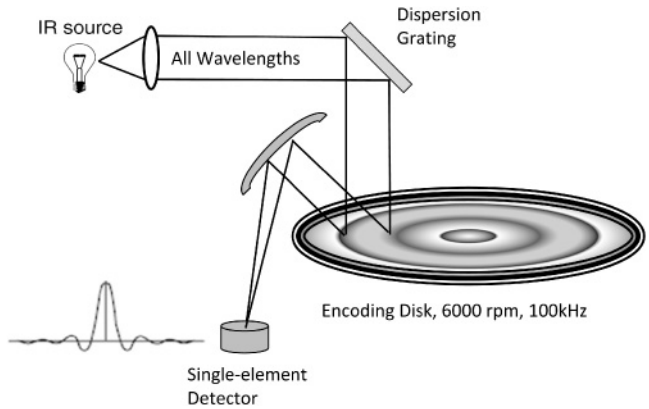


Figure 5.19 *Rotating disk NIR encoding device.*

conventional Fourier transform. A slight drawback is the rather limited number of encoded wavelength elements across the rotating mask, but the technology has the nice feature that the only moving part is rotary and not reciprocating as in conventional interferometric modulators.

The alternative MEMS active grating approach allows for very flexible operation. The technology here is somewhat analogous to a very familiar everyday technology in computer light projectors. These devices use a DLP (digital light projector) which is a MEMS device consisting of thousands of micro-mirrors which can be tilted to deflect a light signal and in effect act as a pixel-sized switch. The MEMS active grating has a similar approach, but the mirrors form one-dimensional tracks, which can be very rapidly lifted or depressed above and below the plane of the surface, thus forming an addressable diffraction grating. Diffraction along any one track can be switched on or off by raising or lowering the adjacent tracks by ¼ wavelength of the incident beam. In the MEMS active grating device, again the basic layout disperses the NIR source beam across the MEMS device, but this time instead of being encoded by a rotating disk, the wavelengths in the NIR beam are selected by tuning the characteristics of the grating (Figure 5.20).

The device can be used in a number of different modes. By sequentially tuning adjacent tracks across the MEMS grating it can act as a fast scanning monochromator. Alternatively by programming a sequence of grating pattern changes (50% on, 50% off) it can act as an Hadamard transform device. This latter mode of operation obviously has a partial multiplex advantage compared with the scanning mode.

Where will these new technologies lead? In principle, they may offer significant manufacturing and operational advantages in terms of reproducibility, robustness, ease of packaging and certification for field mounting in hazardous process environments. But to be useful they will, at least for a good number of current NIR process analytical applications, need to demonstrate demanding levels of spectroscopic performance and long-term stability. This is yet to become clear.

One can see from the comparison in Figure 5.21 that taken overall, looking specifically at spectroscopic performance, and particularly long-term spectral reproducibility – an absolute requirement for useful NIR process analyzers which must be able to maintain the integrity of their installed predictive models over the long term – that FT-NIR is probably on balance the most effective current technology. But it has characteristics, even in its most robust and process-hardened current form, of environmental sensitivity which require careful handling and packaging for process installation. But we may expect to see the development of the new technologies and the optimization of FT-NIR interferometer design converging on the target of a new generation of robust process NIR analyzers.

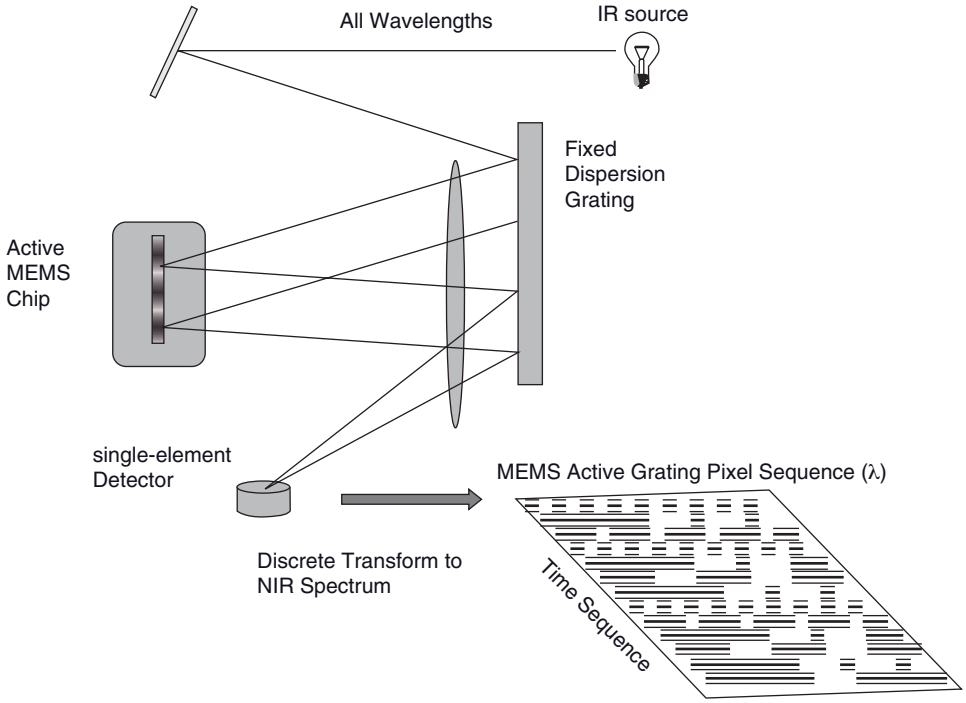


Figure 5.20 *Active-grating MEMS NIR device in Hadamard encoding mode.*

5.4 The Sampling Interface

5.4.1 Introduction

Section 5.3 described a number of alternative design and implementation strategies for near-infrared analyzers, suitable for operation in a process analytical environment. However, none of these analyzers can operate without a robust, maintainable and repeatable sampling interface with the process sample under consideration. In addition to this question of the optical interface to the sample, there is a whole wider area of concern, which is how far the particular sample interface is representative of the sample in the process as a whole. This complex issue is not addressed here, and is dealt with separately in Chapter 3.

In terms of practicality the optical interface to the sample is certainly the most demanding aspect of all. It requires a number of issues to be addressed and clearly understood for each specific application to be attempted. These include:

- The physical nature of the process stream. Is it single-phase or two-phase? Is it liquid, solid, vapor or slurry? What is its temperature and pressure at the sampling point, and how far can these be allowed to change during sampling? What is its viscosity at the appropriate sample measurement temperature?
- The chemical nature of the process stream. Is it at equilibrium (a final product) or is it to be measured mid-reaction? Is sample transport possible, or must the sample be measured in situ? Is it corrosive, and what material and metallurgical constraints exist?
- The optical nature of the process stream. Is it a clear fluid or scattering, highly absorbing, diffusely transmitting or reflecting?

Summary of Quality Features in Emerging and Established NIR Technologies							
Technology / Quality	Scanning Grating and Polychromator Diode Array	Grating MEMS, Hadamard	MEMS Grating Rotating Disk Encoding	AOTF	Tunable SLED/ Fabry-Perot	MEMS FTS	Process FT-NIR
spectral range	900-1700nm low cost, 1100-2100nm medium cost, 1400-2400nm, poor quality	900-2400nm with 3 different gratings	900-4500nm with up to 5 different gratings	800-2300nm	1100-1850nm	vis to 2300nm limited by FO coupling	700-20000nm
resolution	5 to 20 nm	10 to 20 nm, 128 channels	10 to 20 nm, 256 channels	20 to 50 nm	5 nm	40 cm-1	1-32 cm-1
repeatability	excellent	excellent	good	very good	good	short term good	excellent
stability	can be very good	good	good	limited	limited	good	excellent, with permanent alignment
reproducibility	poor, needs mathematical treatment	poor, needs mathematical treatment	poor, needs mathematical treatment	poor	good	good	excellent, guaranteed
measurement speed	simultaneous, good for transients, typical up to 50 scans/s.	1 scan/s.	up to 100 scans/s.	100 scans/s.	10 scans/s.	400 scans/s. resonant scan	up to 5 scans/s., high frequency modulation
photometric linearity	stray light	good	good	stray light	good	good	very good
wavelength scale	non-linear	non-linear	non-linear	temperture sensitive	good	good with metrology	excellent
lineshape	poor	poor	poor	poor	good	good	excellent
size	compact	compact	medium	medium	medium	very small	medium
cost	medium	low	low	medium	medium	low	high
vib robust	very	very	very	good	good	very	medium
temperature	little effect	little effect	little effect	sensitive	good	unkown	sensitive

Figure 5.21 *Technology comparison for NIR process analytical spectroscopy.*

- What issues concerning sample system or probe fouling and process stream filtration or coalescing need to be addressed?
- Where is it possible or desirable to locate the near-infrared analyzer module? What hazardous area certification requirements apply to both analyzer and sampling system?
- Are there multiple process streams to be sampled? If so, can they be physically stream switched, or does each stream require a dedicated sampling interface?
- If fiber optics are going to be used, what physical distances are involved, and what civil engineering costs are implied (for gantry, cable tray, conduit, tunneling etc)?
- What are the quantitative spectroscopic demands of the calibration? How much impact will variation in the sampling interface (as opposed to analyzer instability) have on the likely calibration success and maintainability? What issues concerning transport of laboratory-developed calibration models to the process analyzer need to be addressed?
- What near-infrared spectral region will be targeted for the calibration development, for example, combination region, first combination/overtone, or second combination/overtone? Each region will imply a different transmission path length, and hence cell or probe design.

Depending on the answers to these questions, various sampling interface strategies may be developed which fall into the main categories described in the following sections.

5.4.1.1 Extractive Fast Loop Sample Conditioning Systems

This is based on a sample fast loop with a sample take-off probe in the main process line, and a return to either a pump-suction or ambient pressure sample recovery system (Figure 5.22). The fast loop provides a filtered slip-stream to the analyzer, along with flow control and monitoring (low-flow alarm), and frequently some form of sample autograb facility for sample capture. The slipstream flow to the internal analyzer sample flow cell is relatively low volume and allows for a very high degree of sample temperature control (typically

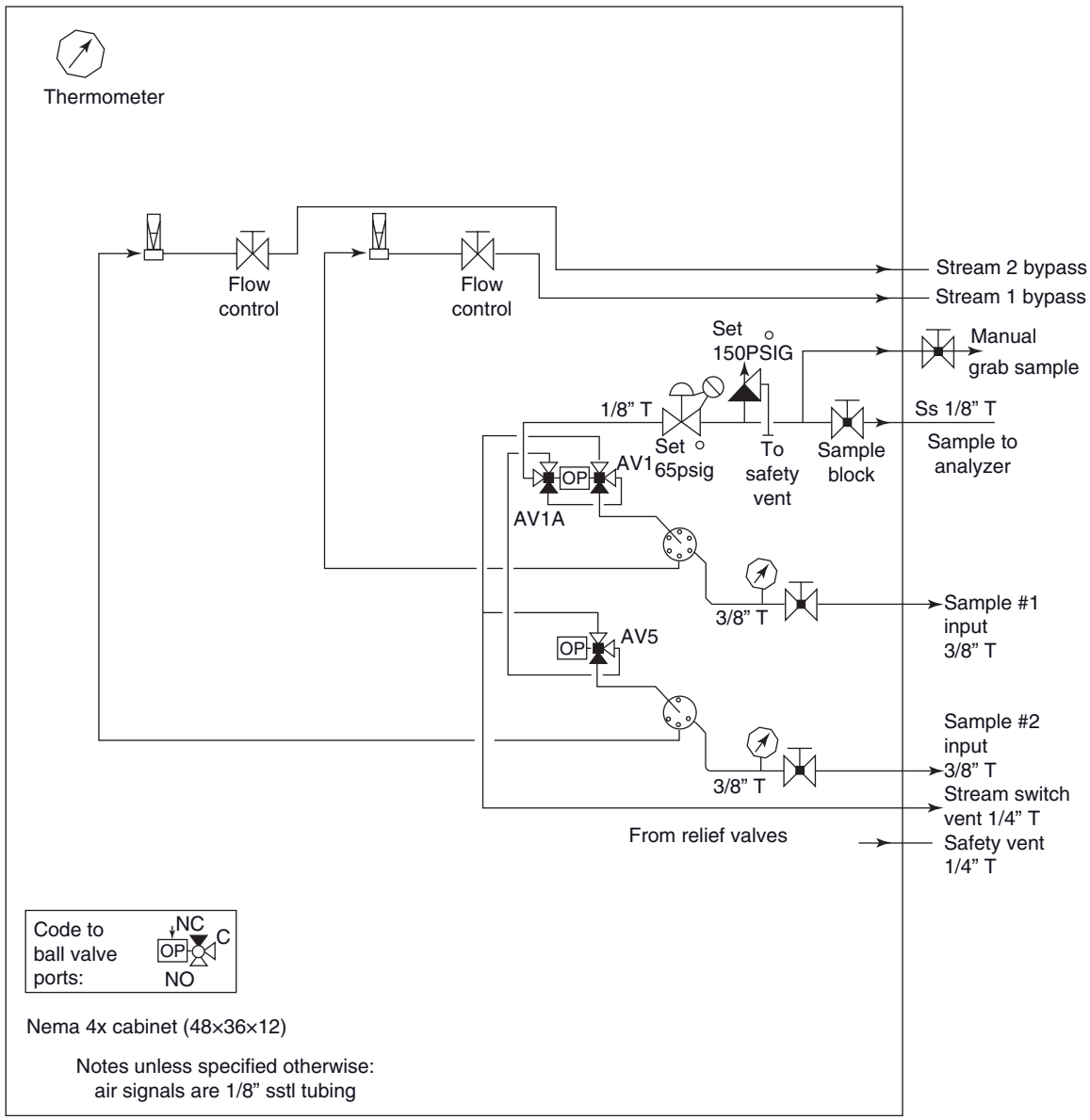


Figure 5.22 *Extractive sample-conditioning fast-loop.*

0.1 °C) within the analyzer. This option is mainly applicable to single-phase liquid equilibrium process streams – and is the *de facto* standard for refinery hydrocarbon applications where the equilibrium (non time-dependant) sample condition is met and where the quantitative stability and reproducibility aspects of the application demand the highest level of sampling control (Figure 5.23).

Where physical stream switching is acceptable, this arrangement allows for multiple sampling fast loops to feed a single near-infrared analyzer, with stream switching controlled by the analyzer. The analyzer will most likely need to be hazardous area compliant, but may be either located in a dedicated analyzer shelter, or may be field-mounted.

5.4.1.2 Local Extractive Fiber-Optic Flow Cell Sample Systems

This is a development of the above where a fiber-optic linked liquid sample transmission cell is integrated with the sample fast loop cabinet (Figures 5.24 and 5.25). There can be multiple sample streams, take-offs and fast loops, each with its own separate fiber-optic transmission cell. The analyzer can either be local with short fiber-optic runs to the sampling cabinet(s), or remote, where a safe area location for the analyzer module may be feasible, but at the cost of longer, potentially less stable fiber-optic runs. This system avoids physical stream switching.

The other clear opportunity presented by this arrangement is the chance to monitor multiple process streams with one analyzer, where the streams require different measurement temperatures. In this case each cell or group of cells requiring a particular temperature regime can be mounted in separate heated cabinets. In typical refinery hydrocarbon applications, the former, extractive fast loop arrangement with physical stream switching is ideal for lighter hydrocarbon streams (naphtha through to light gas oil) and the fiber-optic flow cell arrangement is preferred for heavier streams such as heavy gasoil and similar high viscosity crude distillation unit side-streams.

5.4.1.3 Remote In Situ Fiber-Optic Based Sample Systems

This option comes into play when there are either multiple remote and widely distributed sampling points targeted to be analyzed using one NIR analyzer, or, where the nature of the process stream or reactor makes it impossible to use any significant run of sample transport line. There can be a variety of formats depending on the nature of the sample. In the case of an equilibrium sample, but where the process temperature is high, this can allow for the location of a dedicated high-temperature fiber-optic sample cell cabinet very close to the sample take-off point, minimizing the need for heated sample transport lines. A number of such cabinets can sample multiple process streams and feed back to a centrally-located hazardous area certified fiber-optic multiplexed analyzer module, or at the cost and risk of longer fiber runs, out to a safe area.

Alternatively, for nonequilibrium process streams, where a pumped reactor sample recycle line is available, in-line fiber-optic transmission cells or probes (Figures 5.26 and 5.27) can be used to minimize sample transport. It is highly desirable that some form of pumped sample bypass loop is available for installation of the cell or probe, so that isolation and cleaning can take place periodically for background reference measurement.

In the worse case, where either sample temperature, pressure or reactor integrity issues make it impossible to do otherwise, it may be necessary to consider a direct *in situ* fiber-optic transmission or diffuse reflectance probe. However, this should be considered the position of last resort. Probe retraction devices are expensive, and an *in situ* probe is both vulnerable to fouling and allows for no effective sample temperature control. Having said that, the process chemical applications that normally require this configuration often have rather simple chemometric modeling development requirements, and the configuration has been used with success.

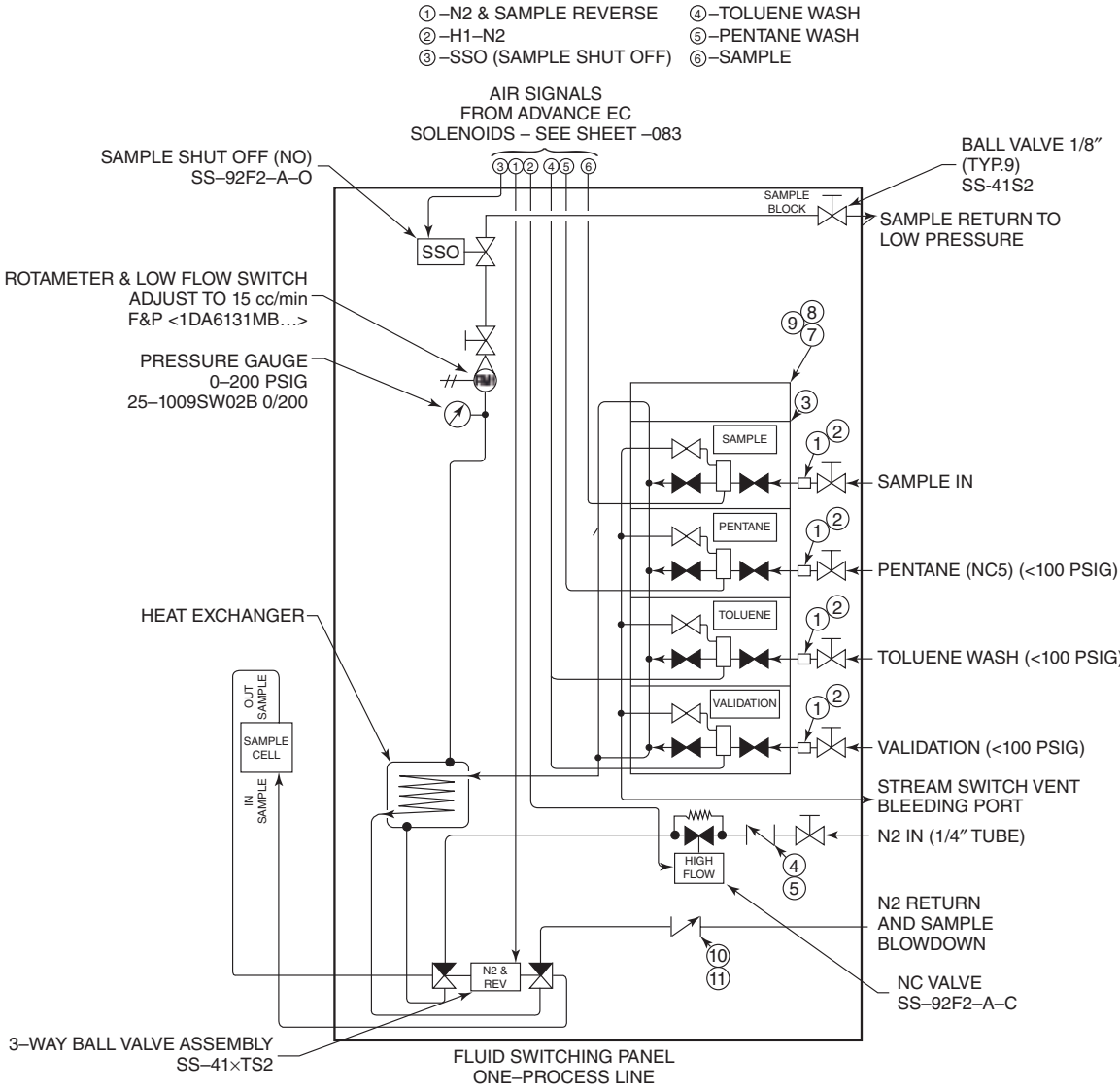


Figure 5.23 *Internal analyser sample flow system.*

Figure 5.24 *Example fiber-optic-coupled flow cell sample system.*

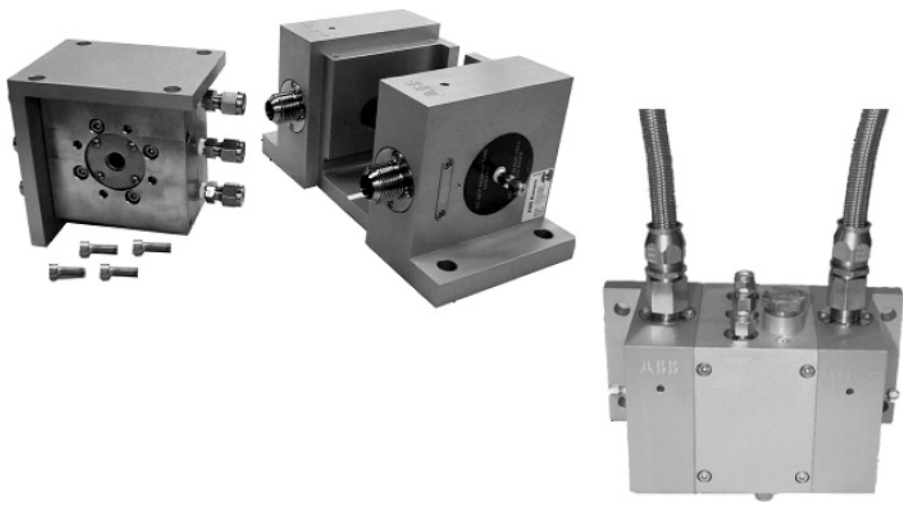


Figure 5.25 *Example fiber-optic-coupled flow cell detail.*

Figure 5.26 *Example fibre-optic-coupled in-line transmission cell* (Reproduced courtesy of Specac Limited, www.specac.co.uk).

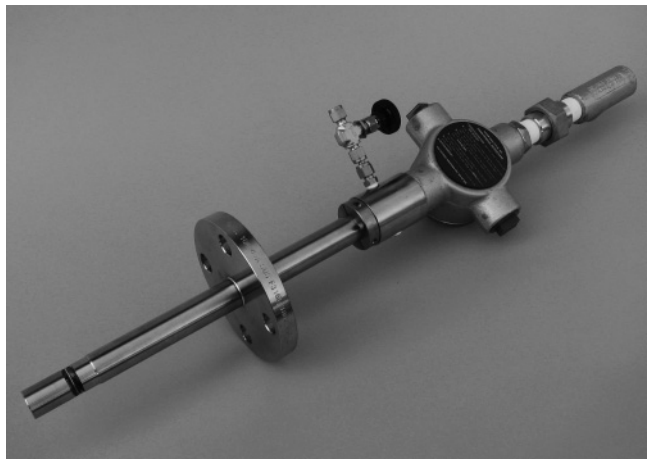


Figure 5.27 *Fibre-optic-coupled insertion transmittance probe* (Reproduced courtesy of Axiom Analytical Inc, www.goaxiom.com).

Examples of field analyzer system configurations for extractive and fiber-optic based sampling are shown in Figures 5.28 and 5.29.

5.4.2 Problem samples: liquids, slurries and solids

These three main classes of process sample streams are in increasing order of difficulty for near-infrared process analysis. In general, liquid streams are best measured in a transmission sampling mode, solids (powders) in diffuse reflectance mode, and slurries in either diffuse reflectance or diffuse transmission according to whether the liquid phase or the suspended phase is of greater analytical significance. If the

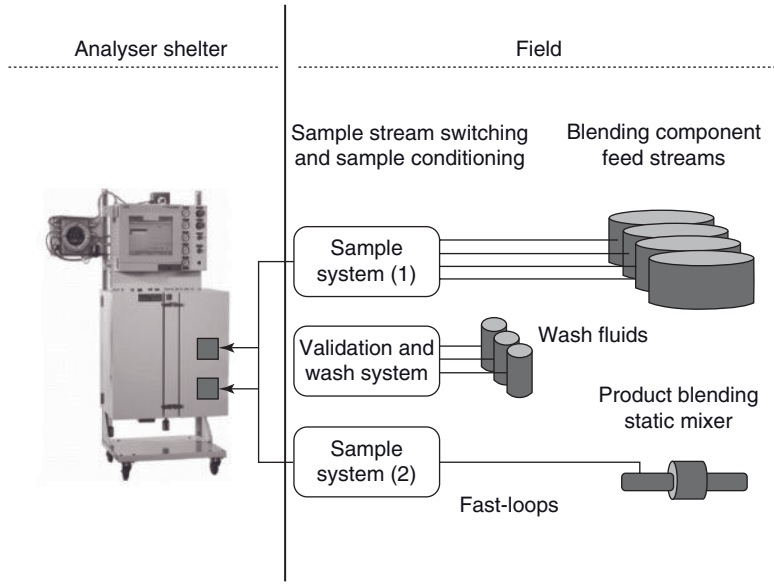


Figure 5.28 *Typical field configuration for extractive on-line process analyzer.*

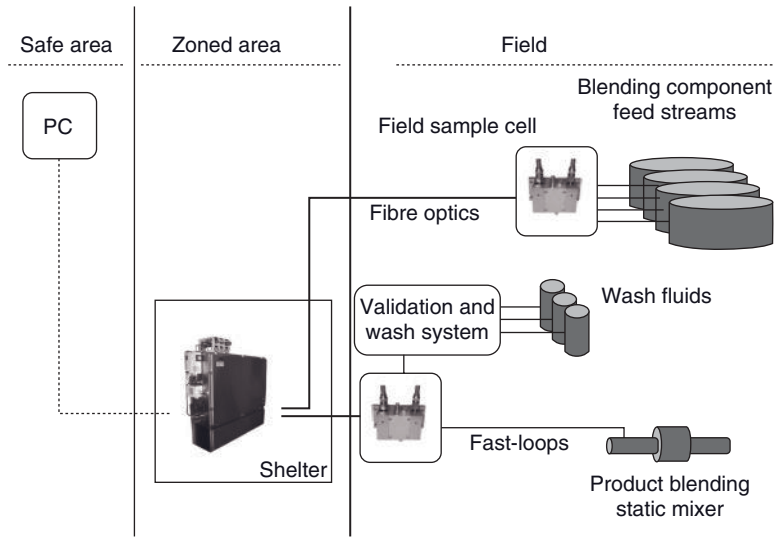


Figure 5.29 *Typical field configuration for fiber-optic on-line process analyzer.*

latter, and diffuse transmission is used, a specialized setup using an NIR analyzer, a fiber-optic bundle for illumination and a local dedicated high sensitivity large area detector in close proximity to the sample (in order to maximize light collection) will very likely be needed.

When considering liquid sampling, issues of sample path length, sample temperature and pressure, and sample viscosity, along with the overall process engineering environment, will influence the final decision

concerning analyzer and sample interface format. Many of these issues are interconnected. Determination of the optimum spectral region for calibration model development may suggest use of the combination region (4000–5000 cm^{-1}), which will imply the use of relatively short path lengths (around 0.5 mm or less). This option is both feasible and desirable for light, clean, hydrocarbon streams. This setup is often the chosen standard for refinery and petrochemical applications, where the convenience and additional sample control provided by a filtered bypass loop, sample back pressure and precise sample temperature control all pay large dividends relative to the rather trivial extra effort of sample system maintenance. However, for both higher temperature or higher viscosity sample streams, this degree of sample transport and manipulation may not be possible, and selection of a wavelength region requiring relatively short path lengths may not be viable. In those cases, in increasing order of potential difficulty, remote fiber-optic coupled sample transmission cell cabinets, in-line fiber-optic transmission cells installed in a pumped bypass loop, in-line fiber-optic transmission probes and finally in-reactor fiber-optic transmission probes may be used.

Although the use of a sample fast loop sample conditioning system is not always possible, it is important to realize the benefits which are sacrificed if the decision is made to use alternative interface formats. First and foremost, the sample conditioning system allows for easy sample cell cleaning, maintenance and the use of reference or calibration fluids. Also it allows incorporation of a local process sample take-off or autograb facility, to ensure that recorded spectra and captured samples are correctly associated, for calibration development or maintenance. In addition many process streams operate at elevated pressure, and although this has a negligible effect on the near-infrared spectrum, uncontrolled pressure drops can cause cavitation and bubbling. A sample system allows for either sample back-pressure and flow control, or a sample shut-off valve to stop the flow against a back pressure during measurement to eliminate this risk. Some of these benefits are also realizable in a pumped bypass loop for in-line cell or probe installation. Here it is normally possible to arrange isolation valves to allow for cell or probe removal between process shutdowns. Also wash or calibration fluids can be introduced, and to some extent sample flow and pressure can be controlled. The control of sample temperature is however somewhat compromised, and calibration modeling strategies will have to be adopted to address that. It is for these reasons, rather than any fundamental issues relating to near-infrared technology or spectroscopy, that those applications where a reliable bypass or fast loop cannot realistically be envisaged (for example hot polymerization or polycondensation processes) have proved the most demanding for near-infrared process analyzes. Where (as in these cases) direct in-reactor insertion transmission probe installation is the only physically practical option, then issues of probe fouling, reference measurement, sample capture and probe resistance to process conditions must be addressed.

As mentioned above, the sampling of solids (powders) is also a challenging objective,[29,30] but one for which quite satisfactory solutions can be found. The typical NIR process analysis applications for solids are for powder drying (either moisture or solvents) and for blend content uniformity (end point) determinations. Though the energy throughput of a fiber-optic coupled diffuse reflectance probe in contact with a suitably scattering powder sample may be orders of magnitude less than the throughput of a simple clear liquid transmission measurement, this is an area where the available power output of an FT-NIR analyzer or use of more sensitive detectors can be beneficial, and allow the use of smaller fiber-optic bundles, multiplexed, multichannel sample outputs, and hence longer fiber-optic runs at less cost. The main difficulty in developing quantitative calibration models for the measurement of solid process streams is that of sample presentation and homogeneity. Unlike liquid process streams, process powders are often transported or manipulated in ways which make them difficult to sample reproducibly. Powders flowing down pipes may not offer a uniform density to a sample probe head. Powders in dryers or blenders may either cake the probe head, or simply sit around it, unrefreshed by new sample. Thus sample probe location and environment are critical considerations when setting up solids applications. To make this decision appropriately, knowledge of the process dynamics is needed.

5.4.3 The use of fiber optics

5.4.3.1 Principles of Operation

Simply stated, fiber-optic cable allows for the transfer of light from one point to another without going in a straight line, and as such their use in process analytical applications is highly seductive. In a likely near-infrared process application there may be multiple sample stream take-offs, each with different sample characteristics, located in positions spread around a process unit, with the only apparently convenient analyzer location some distance away.

Fiber-optic cables[31,32] work on the principle of total internal reflection (Figure 5.30), when an incident light beam falls on an interface between two media of differing refractive indices, going in the direction of higher to lower refractive index. In a fiber-optic cable the higher refractive index is provided by the fiber core and the lower by a cladding. This can either be a totally different material (e.g. a polymeric coating) in close optical contact with the core, or a layer formed around the core made from the same material, but with added doping to change the refractive index. In this latter case, the change in refractive index between the core and the cladding can either be abrupt (a step index fiber) or gradual (a graded index fiber).

The maximum input acceptance angle for a fiber-optic cable is determined by the critical angle for total internal reflection, which is given by Snell's law:

$$\theta_c = \sin^{-1}(\eta_1/\eta_2) \tag{5.21}$$

From which the acceptance angle is:

$$\theta_a = \sin^{-1}\left(\eta_1^2 - \eta_2^2\right)^{1/2} \tag{5.22}$$

and the fiber-optic numerical aperture (NA):

$$\mathrm{NA} = \sin\theta_a = \left(\eta_1^2 - \eta_2^2\right)^{1/2} \tag{5.23}$$

Typical values for low-hydroxyl silica fibers are in the range 0.22 to 0.3.

It must be noted that the wave propagating through the fiber-optic core can potentially do so in a number of geometric forms (depending on the electric field amplitude distribution across the plane of the core) referred to as modes. Even for a monochromatic input, modes with a more direct path through the core will travel down the fiber with a larger linear velocity than those with a more transverse component. This leads to the effect of mode dispersion, where a short input light pulse is stretched as it travels down the fiber. Hence, for digital and telecommunications purposes very narrow step -index fibers, which will typically

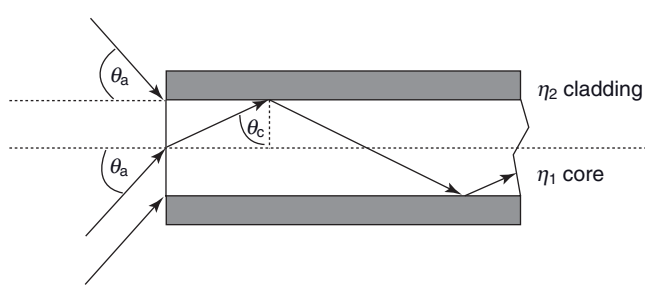


Figure 5.30 *Internal reflection and acceptance angle in a fiber-optic cable.*

transmit only a single mode, in a limited band pass of wavelengths, are more suitable. By contrast, for process analytical spectroscopy, where continuous broadband sources are used, and the dominating requirement is the overall transmission efficiency of the fiber, then larger diameter multimode fibers are completely acceptable.

5.4.3.2 Losses in Fibers

The throughput of an optical fiber depends on a number of factors including:

- the input fiber diameter
- the input numerical aperture
- the length of the fiber
- the number of impurity centres causing scattering
- the concentration of impurities causing absorption
- the absorption edge of the core material itself
- the installed configuration of the fiber, in particular the minimum bending radius to which it is subjected.

All of these factors need to be considered against the cost, the robustness, and the amount of light available to couple into the fiber from the near-infrared process analyzer. In general the physical dimensions of the source image, as it is focussed onto the fiber input, even when the numerical apertures are matched, will be larger than the fiber diameter, and there will be consequent injection losses for smaller diameter fibers. However, these may be compensated for by the fact that the spread of input angles, when the source image is stopped down in this way, is reduced, and the overall fiber transmission, once the light is inside, is improved compared with a larger diameter fiber.

Also, for parallel multichannel devices, the source image can be divided, and injected simultaneously into a number (for example up to eight) of output fibers. Alternatively for more demanding diffuse reflectance or diffuse transmission applications with much lower optical throughput, the whole source image dimension can be used, but injected into micro- or macrobundle fiber optics.

A typical low hydroxyl silica fiber used for near-infrared liquid sampling applications should be a single-fiber, approximately 250–300 μm diameter. This provides a good compromise between injection losses, transmission losses, cost and robustness. In general larger diameter single fibers are more fragile and more expensive (typically in the ratio of the fiber cross-sectional area). Alternatively for lower throughput applications microbundles of $7 \times 200\,\mu$m fibers (effectively a 600 μm fiber overall) can be used, or minimacrobundles up to a total of 2 mm diameter. These increase in cost proportionally, but, being made up of multiple small fibers, are far more robust than a large single fiber.

For near-infrared applications the overwhelmingly most common fiber-optic material is low-hydroxyl silica. For realistic run lengths this allows for use down to about 2000 to 2200 nm before the silica absorption band edge cuts in too strongly, depending on the exact fiber length. Their successful use is made possible by careful control of the number of –OH groups left as impurities in the silica core. There are significant absorption contributions from –OH centers, and these must be kept to the minimum level possible. This is because the other fiber loss mechanisms such as scattering or bending tend in general to produce fairly benign and easily handled distortions in the resultant sample absorbance spectrum (for example baseline offsets, tilts or curvature), which can be dealt with by standard mathematical correction procedures (baseline correction or multiplicative scatter correction). However, the –OH absorption bands are chemical in origin, and they are typically temperature sensitive as regards intensity, position and band shape. Unless these features are minimized, the risk of nonreproducible absorption-like band distortions appearing in the sample spectrum is high, which will be difficult to deal with in the calibration model development.

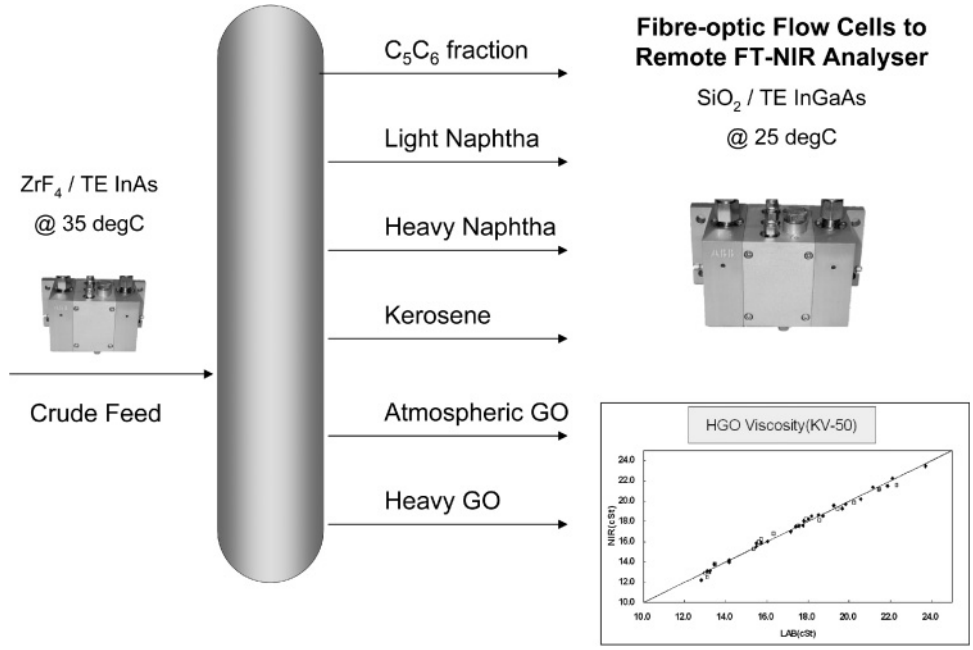


Figure 5.31 *Crude distillation unit (CDU) application of PAT using ZrF4 and silica fiber optics.*

Alternative fiber-optic materials exist, but the only serious candidate for PAT applications is ZrF_4, which has the property of transmitting down to below 2000 cm^{-1}, in other words well across the 4000–5000 cm^{-1} combination band region when used in combination with an indium arsenide (InAs) detector. Although it is fragile and somewhat expensive compared with silica fiber, it is a good possibility for specialized refinery hydrocarbon applications which require not only the combination band region, but also physical separation of streams (for example the crude oil feed and multiple side-draw streams of a crude distillation unit).

Note how the long wavelength transmission capability of the ZrF_4 fiber optics can be used to link the NIR analyzer to a short path length sample flow cell dedicated to the crude feed stream, where the intensely black, asphaltene-loaded sample stream requires to be measured in the 4000–5000 cm^{-1} combination band region (shorter wavelengths are too scattering), whereas the lighter side-draw streams can be measured more conventionally in the overtone region with low-OH silica fibers (Figure 5.31).

5.5 Practical Examples of Near-Infrared Analytical Applications

The principal functional groups contributing to NIRS are C–H, N–H and O–H, with additional contributions from C=O and C=C groups influencing particularly the shape and location of combination bands especially at lower frequencies (4000–5500 cm^{-1}). However, though one is dealing with only three major functional groups, all containing the H atom, it does not follow that the information content is limited. In real polyatomic molecules the normal modes of vibration can involve multiple X–H units in different environments, and the individual X–H bond strengths are significantly affected by the local bonding environment. In Sections 5.5.1 and 5.5.2 we shall review two important applications of NIR analysis which demonstrate the

power of the technique to develop predictive calibration models for key compositional and bulk physical properties through the NIR spectroscopic data[33,34].

5.5.1 Refinery hydrocarbon streams

For saturated hydrocarbons the functional groups of concern are methyl, methylene and methyne ($–CH_3$, $–CH_2–$, >CH–). The spectra of typical hydrocarbon mixtures (for example as in gasoil and gasoline) are dominated by two pairs of strong bands in the first overtone and combination regions (5900–5500 cm^{-1} and 4350–4250 cm^{-1}) which are predominantly methylene $–CH_2$. The methyl end groups typically show up as a weaker higher frequency shoulder to these methylene doublets.

Olefinic, unsaturated hydrocarbons have CH groups adjacent to C=C double bonds. The C=C vibrations do not show up directly except as weaker contributions in the combination region, but the nucleophilic effect of the C=C bond tends to shift nearby CH groups to significantly higher wavenumber, and this is indeed seen in hydrocarbon spectra (Figure 5.32).

Not only are there clearly distinguishable saturated hydrocarbon functionalities, but the spectra possess a set of absorption band groups between 4760–4450 cm^{-1} and 4090–4000 cm^{-1}. These features are due to aromatic C–H groups appearing as combination bands. Similar, though less clearly separated and defined aromatic C–H features are also observable in the first and second overtone C–H stretch regions (approx 6250–5550 cm^{-1} and 9100–8000 cm^{-1}, respectively). The band group in the combination region at approximately 4600 cm^{-1} is particularly informative because it arises from aromatic C–H and ring C–C modes in combination.

The outcome of this is that although not easily interpretable by inspection (as for example in mid-infrared absorbance spectra) the NIR spectra of hydrocarbon streams possess a very high information content. Use of suitable chemometric methods allows for the reduction of this information content into extremely powerful predictive models for not only chemical compositional properties, but also bulk physical and fuel properties of hydrocarbon process streams and products.[35,36]

Figure 5.33 shows an example dataset of mixed hydrocarbons used as a petrochemical feedstock. These are straight run naphthas, which consist of a wide range of alkane, isoalkane, aromatic and naphthenic

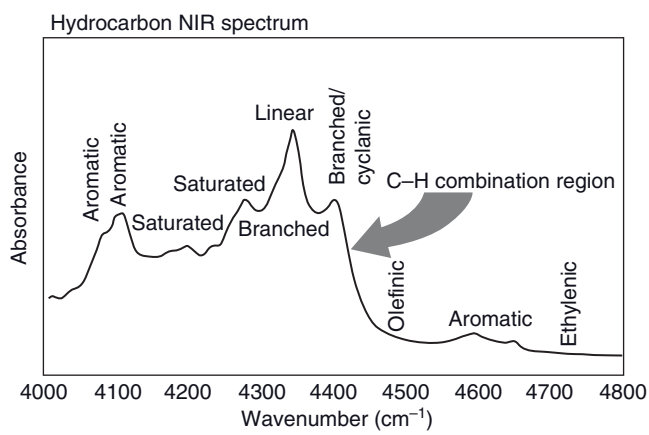


Figure 5.32 *Typical hydrocarbon spectrum with assignments.*

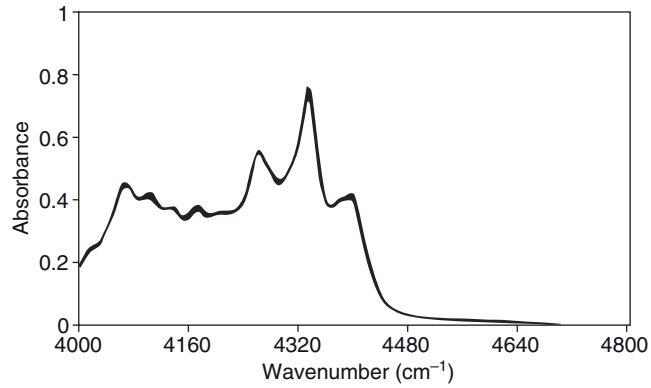


Figure 5.33 *Naphtha data set (100 samples).*

hydrocarbons, mainly in the range C_4 to C_9. The conventional analytical method for naphtha analyzis is temperature programmed GC, which can provide a full analysis including C-number breakdown, but which is rather slow for process optimization purposes. The process analytical objective in this case is to provide feed-forward compositional data to an on-line cracking furnace optimizer (such as SPYRO) to allow rapid adjustment of cracking conditions in the event of naphtha feed stream quality transitions.

Figure 5.34 shows the FT-NIR predictive models for total vol% paraffins, isoparaffins, naphthenes and aromatics and carbon number distribution for a typical naphtha dataset. Since NIRS can be demonstrated to have sufficient hydrocarbon speciation capability to reproduce the analyzis with the same precision as the GC method, but in a fraction of the time, then a useful process analytical goal has been achieved.

Further examples of the capability of NIR analyzers to yield detailed predictions of hydrocarbon sample stream properties can be seen (Figure 5.35) throughout the range of refining operations including distillation, hydrotreating, cracking, alkylation, reforming, isomerization and final product blending.

5.5.2 Polyols, ethoxylated derivatives, ethylene oxide/propylene oxide polyether polyols

Alcohols and organic acids both contain the O–H functionality, which is the next most important NIR signature after C–H. However it behaves very differently. The dominant feature of the –OH group is its capacity for both intra- and inter-molecular hydrogen bonding. The essential –OH group contributions in the NIR spectrum are combination and overtone stretching modes that occur around 5250–4550 cm^{-1} and 7200–6000 cm^{-1} (Figure 5.36).

However, the position, shape and intensity of the –OH absorptions that appear in these regions are very strongly influenced by hydrogen bonding effects, which generally tend to randomize the electronic environment of the –OH group, spreading out the range of absorption frequencies, and reducing absorbance intensity. Moreover, where these are intermolecular effects, they are strongly influenced by sample temperature, changes of state and sample dilution. Nonetheless, there is a very wide range of industrially significant –OH containing materials, often used as building blocks in polymer and oleochemical processes. Amongst these materials, such as glycols, ethoxylates, polyester polyols, polyurethane prepolymers, etc, applications such as hydroxyl number or acid value determination are highly dependent upon the NIR spectroscopy of the –OH and –COOH group. An example of a NIR calibration for hydroxyl value and acid number, developed as a quicker alternative to conventional analysis by wet-chemical titration is shown in Figure 5.37.

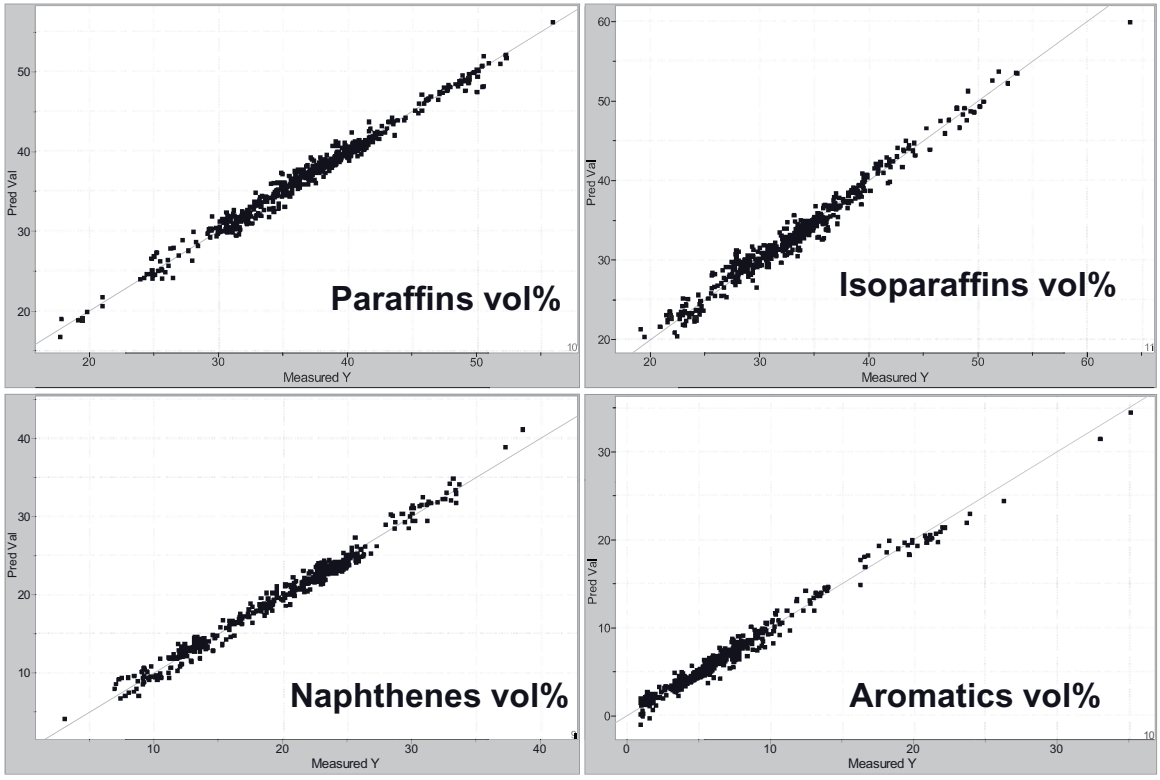


Figure 5.34 *Predictive models for naphtha compositional analysis (PINA).*

Refining

Blending Operations
- Gasoline Blending
- Diesel Blending
- Jet/ Kerosene Blending

Conversion Operations
- Catalytic Reforming
- Isomerisation
- HF Alkylation
- Hydrotreating
- Hydrodesulphurisation
- Cat Cracking
- Hydrocracking
- Residue Upgrading

Distillation Operations
- Crude desalting and stabilisation
- Crude Blending
- Distillation
- Rundown
- Vacuum Distillation
- Heavy Residues
- Base Oils

Primary Petrochemical and Downstream

Petrochemical Processes
- Naphtha Cracking
- Ethylene / Propylene
- Aromatics / BTX
- Butanes, Butadienes
- Syngas, Methanol
- Ammonia

Downstream Operations
- Butadiene / Isoprene
- Caprolactam / Nylon 6
- Adipic Acid / Nylon 6.6
- Acrylonitrile
- Phenol / Cumene / Acetone
- Phthalic anhydride
- Terephthalic Acid
- Maleic Anhydride
- Acetic Anhydride
- Urea
- Polyethylene / Polypropylene
- Vinyl Acetate
- Acetic Acid
- Polyvinyl chloride
- ABS resins

Figure 5.35 *Refinery hydrocarbon process analytical applications of FT-NIR.*

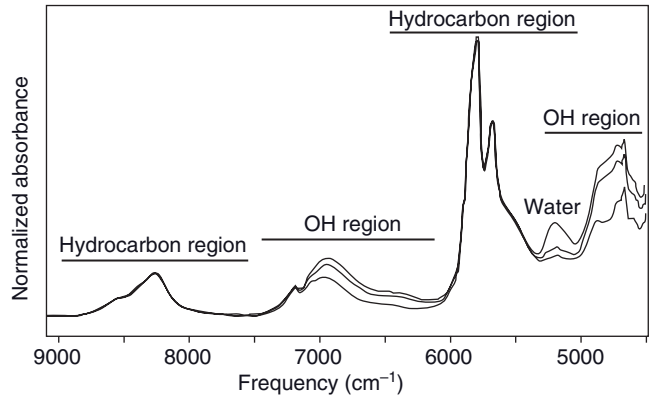


Figure 5.36 *NIR spectra of polyols.*

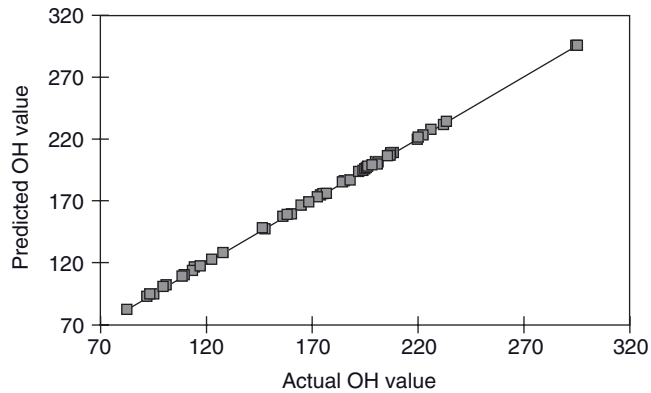


Figure 5.37 *Calibration curve for hydroxyl number.*

5.5.3 Oleochemicals, fatty acids, fatty amines and biodiesel

A further example of process quality monitoring and reactor batch profiling using NIRS comes with oleochemical and biodiesel production. An established use of FT-NIR analysis (AOCS Method Cd 1e_01) is the determination of the key vegetable oil processing parameters – iodine value (IV) and percentage trans fat content (%Trans) (Figure 5.38).

Vegetable oil feedstocks are used in the biodiesel production market, which is in a phase of rapid growth, driven largely by increasingly detailed economic incentives to achieve biodiesel substitution in mineral-diesel fuel blends. The base process is the conversion of raw vegetable oils (trigycerides) to their fatty acid methyl esters through a process called transesterification. This can be a batch, continuous or semi-continuous process. Transesterification is based on the chemical reaction of triglycerides with methanol to form methylesters and glycerine in the presence of an alkaline catalyst. This reaction is effected in a two-stage mixer-settler unit (Figure 5.39).

The standard analytical methods for biodiesel methyl ester are time-consuming and often require multiple physical analyzes per sample. NIRS allows for the simple set up of calibrations for multiple important biodiesel qualities such as mono-, di- and triglycerides, residual methanol, glycerol and moisture as well as

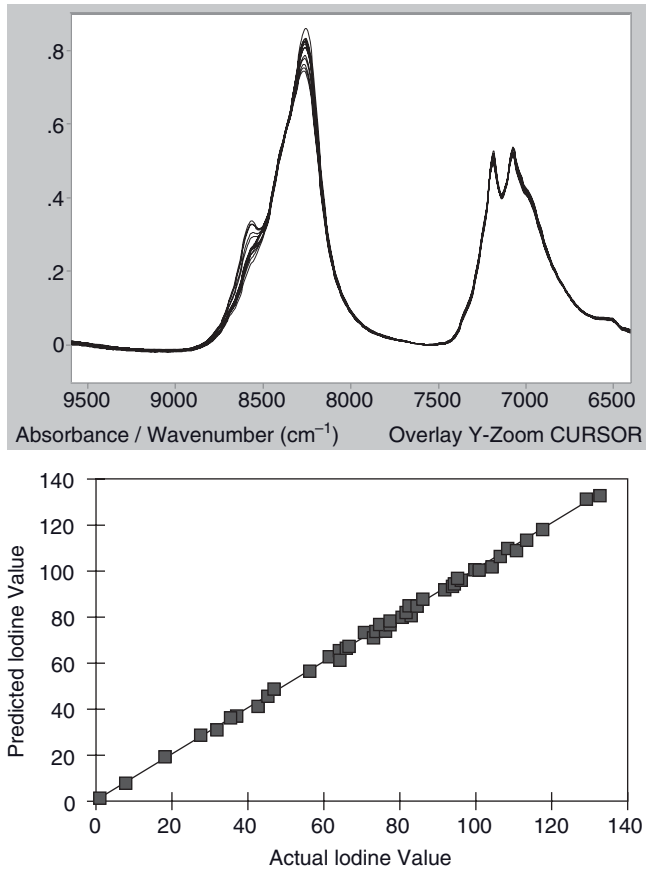


Figure 5.38 *NIR spectra of vegetable oils and iodine value (IV) calibration.*

conventional fuel properties such as cloud point, cetane index, viscosity and distillation properties. The rapid at-line prediction of these key biodiesel properties allows for effective biodiesel reactor profiling and batch end-point determination, thus increasing yield, throughput and final product quality compliance and helping to reduce costly re-work or blending.

5.6 Conclusion

In surveying the main available technologies applicable for near-infrared process analytical technology applications, it is clear that a very wide range of analyzer and sampling interface technologies can be brought to bear on any specific proposed process analytical application. This spread of techniques allows for a remarkable flexibility. The careful use of this flexibility to select the optimum analyzer and sampling configuration for the target application is one key to success. The other keys to success in a NIR PAT application lie outside the scope of this chapter, but are no less important and are covered elsewhere in this book. They include key issues such as calibration modeling development techniques, project management, application support and on-going performance audit, validation and project maintenance. However, with all these issues

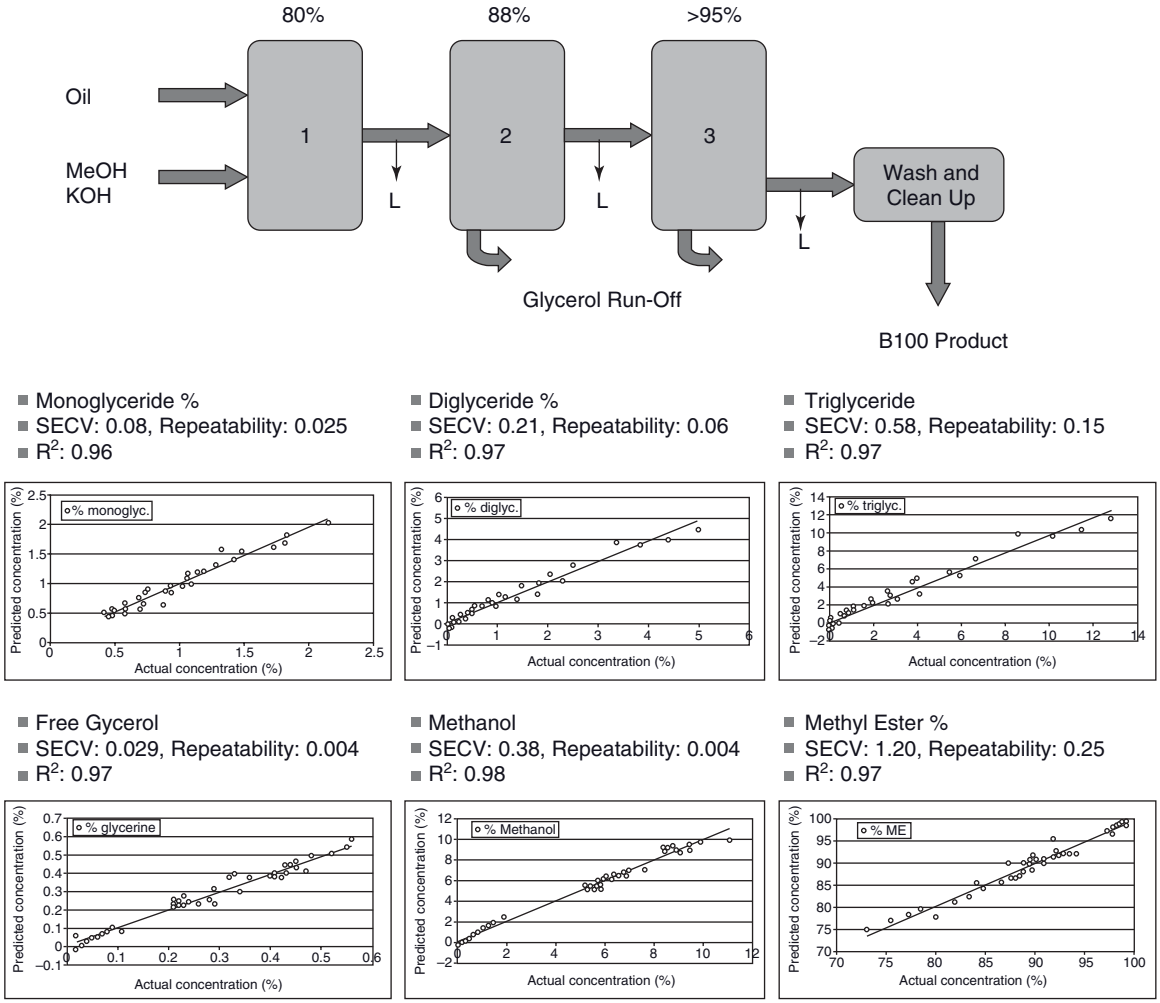


Figure 5.39 *Biodiesel process-monitoring and FT-NIR calibrations.*

correctly addressed, near-infrared technologies allow a unique opportunity to provide the highest quality real-time process analytical information ideally suited to process control and optimization techniques. It is a hallmark of modern near-infrared process analyzers, with their improved reliability, sampling methods, and suitable process-oriented software and communication capabilities, that they are more and more a normal and indispensable part of process unit control strategies.

References

1. K.H. Koch, *Process Analytical Chemistry: Control, Optimisation, Quality and Economy*, Springer-Verlag, Berlin, 1999.
2. J. Chalmers (Ed), *Spectroscopy in Process Analysis*, Sheffield Academic Press, Sheffield, 2000.

3. F. McClellan and B.R. Kowalski (eds), *Process Analytical Chemistry*, Blackie Academic and Professional, London, 1995.
4. P. Williams and K. Norris (eds), *Near-Infrared Technology in the Agricultural and Food Industries*, 2nd edn, American Association of Cereal Chemists, Inc., St. Paul, MN, 2001.
5. D.L. Wetzel, Contemporary near-infrared instrumentation, in *Near-Infrared Technology in the Agricultural and Food Industries*, P. Williams and K. Norris (eds), 2nd edn, American Association of Cereal Chemists, Inc., St. Paul, MN, 2001.
6. P.J. Brimmer, F.A. DeThomas and J.W. Hall, Method development and implementation of near-infrared spectroscopy in industrial manufacturing processes, in *Near-Infrared Technology in the Agricultural and Food Industries*, P. Williams and K. Norris (eds), 2nd edn, American Association of Cereal Chemists, Inc., St. Paul, MN, 2001.
7. C.E. Miller, Chemical principles of near-infrared technology, in *Near-Infrared Technology in the Agricultural and Food Industries*, P. Williams and K. Norris (eds), 2nd edn, American Association of Cereal Chemists, Inc., St. Paul, MN, 2001.
8. W.F. McClure, Near-infrared instrumentation, in *Near-Infrared Technology in the Agricultural and Food Industries*, 2nd edn, P. Williams and K. Norris (eds), American Association of Cereal Chemists, Inc., St. Paul, MN, 2001.
9. D.A. Burns and E.W. Ciurczak (eds), *Handbook of Near-Infrared Analysis*, 2nd edn, Marcel Dekker, New York, 2001.
10. E.W. Ciurczak, Principles of near-infrared spectroscopy, in *Handbook of Near-Infrared Analysis*, 2nd edn, D.A. Burns and E.W. Ciurczak (eds), Marcel Dekker, New York, 2001.
11. W.J. McCarthy and G.J. Kemeny, Fourier-transform spectrophotometers in the near-infrared, in *Handbook of Near-Infrared Analysis*, D.A. Burns and E.W. Ciurczak (eds), 2nd edn, Marcel Dekker, New York, 2001.
12. J.J. Workman and D.A. Burns, Commercial NIR Instrumentation, in *Handbook of Near-Infrared Analysis*, 2nd edn, D.A. Burns and E.W. Ciurczak (eds), Marcel Dekker, New York, 2001.
13. H.W. Siesler, Y. Ozaki, S. Kawata and H.M. Heise (eds), *Near-Infrared Spectroscopy: Principles, Instruments, Applications*, Wiley-VCH, Weinheim (Germany), 2002.
14. J. Wilson and J.F.B. Hawkes, *Optoelectronics: An Introduction*, Prentice-Hall International, London, 1983.
15. F.A. DeThomas and P.J. Brimmer, Monochromators for near-infrared spectroscopy, in *Handbook of Vibrational Spectroscopy*, J.M. Chalmers and P.R. Griffiths (eds), vol **1**, John Wiley & Sons, New York, 2002.
16. E.W. Stark, Near-infrared array spectrometers, in *Handbook of Vibrational Spectroscopy*, J.M. Chalmers and P.R. Griffiths (eds), vol **1**, John Wiley & Sons, New York, 2002.
17. J.M. Chalmers and P.R. Griffiths (eds), *Handbook of Vibrational Spectroscopy*, John Wiley & Sons, New York, 2002.
18. D.L. Wetzel, A.J. Eilert and J.A. Sweat, Tunable filter and discrete filter near-infrared spectrometers, in *Handbook of Vibrational Spectroscopy*, J.M. Chalmers and P.R. Griffiths (eds), vol **1**, John Wiley & Sons, New York, 2002.
19. D. Gies and T.-C. Poon, Measurement of acoustic radiation pattern in an acousto-optic modulator, *Proceedings of IEEE SoutheastCon*, 441–445 (2002).
20. D.S. Goldman, Near-Infrared spectroscopy in process analysis, in *Encyclopedia of Analytical Chemistry*, J.W. Peterson (ed.), vol **9**, Process Instrumental Methods, John Wiley & Sons, New York, 2000.
21. P.R. Griffiths and J. de Haseth, *Fourier Transform Infrared Spectrometry*, Wiley-Interscience, New York, 1986.
22. P. Hendra, C. Jones and G. Warnes, *Fourier Transform Raman Spectroscopy*, Ellis Horwood, Chichester, 1991.
23. J.P. Coates and P.H. Shelley, Infrared spectroscopy in process analysis, in *Encyclopedia of Analytical Chemistry*, J.W. Peterson (ed.), vol **9** Process Instrumental Methods, John Wiley & Sons, New York, 2000.
24. J. Kauppinen, J. Heinonen and I. Kauppinen, Interferometers based on rotational motion, *Appl. Spectrosc. Rev.*, **39**, 99–129 (2004).
25. M. Kraft, Compact MEMS high-speed Fourier transform spectrometer, *4th International Conference on Vibrational Spectroscopy*, Corfu (2007).
26. R.A. Crocombe, Miniature optical spectrometers (Part IV), *Spectroscopy Magazine* at www.spectroscopyonline.com (2008), accessed 13 November 2009.
27. R.V. Kruzelecky and A.K. Ghosh, Miniature spectrometers, in *Handbook of Vibrational Spectroscopy*, J.M. Chalmers and P.R. Griffiths (eds), vol **1**, John Wiley & Sons, New York, 2002.
28. M. Harwir and N.J.A. Sloane, *Hadamard Transform Optics*, Academic Press, New York, 1979.
29. N.M. Doyle, Tech Note AN-922, *International Foundation Process Analytical Chemistry (IFPAC)*, Conference Paper 1-020 (2001).
30. D.L. Wetzel and J.A. Eilert, Optics and sample handling for near-infrared diffuse reflection in *Handbook of Vibrational Spectroscopy*, J.M. Chalmers and P.R. Griffiths (eds), vol **1**, John Wiley & Sons, New York, 2002.

31. F.M. Mirabella (ed.), *Modern Techniques in Applied Molecular Spectroscopy*, John Wiley & Sons, New York, 1998.
32. C.W. Brown, Fibre optics in molecular spectroscopy, in *Modern Techniques in Applied Molecular Spectroscopy*, F.M. Mirabella (ed), John Wiley & Sons, New York, 1998.
33. C.E. Miller, Near-infrared spectroscopy of synthetic and industrial samples in *Handbook of Vibrational Spectroscopy*, J.M. Chalmers and P.R. Griffiths (eds), vol **1**, John Wiley & Sons, New York, 2002.
34. P. Dallin, NIR analysis in polymer reactions, *Process Control and Quality*, **9**, 167–172 (1997).
35. H. Chung, Applications of NIR in refineries and important issues to address, *App. Spectrosc. Rev.*, **42**, 251–285 (2007).
36. U. Hoffman and N. Zanier-Szydlowski, Portability of near-infrared spectroscopic calibrations for petrochemical parameters, *J. Near Infrared Spectrosc.*, **7**, 33–45 (1999).

6

Infrared Spectroscopy for Process Analytical Applications

John P. Coates

Coates Consulting, Newtown, CT, USA

6.1 Introduction

Infrared (IR) spectroscopy offers many unique advantages for measurements within an industrial environment, whether they are for environmental or for production-based applications. Historically, the technique has been used for a broad range of applications ranging from the composition of gas and/or liquid mixtures to the analysis of trace components for gas purity or environmental analysis. The instrumentation used ranges in complexity from simple filter-based photometers to optomechanically complicated devices, such as Fourier transform infrared (FTIR) spectrometers. Simple nondispersive infrared (NDIR) instruments are in common use for measurements that feature well-defined methods of analysis, such as the analysis of combustion gases for carbon oxides and hydrocarbons. For more complex measurements it is normally necessary to obtain a greater amount of spectral information, and so either full-spectrum or multiple wavelength analyzers are required.

Of the analytical techniques available for process analytical measurements, IR is one of the most versatile, where all physical forms of a sample may be considered – gases, liquids, solids and even mixed phase materials. A wide range of sample interfaces (sampling accessories) have been developed for infrared spectroscopy over the past 20 to 30 years and many of these can be adapted for either near-line/at-line production control or on-line process monitoring applications. For continuous on-line measurements applications may be limited to liquids and gases. However, for applications that have human interaction, such as near-line measurements, then all material types can be considered. For continuous measurements sample condition, as it exists within the process, may be an issue and factors such as temperature, pressure, chemical interferants (such as solvents), and particulate matter may need to be addressed. In off-line applications this may be addressed by the way that the sample is handled, but for continuous on-line process applications this has to be accommodated by a sampling system.

Process Analytical Technology 2e Edited by Katherine Bakeev
© 2010 John Wiley & Sons, Ltd.

IR is one of three forms of vibrational spectroscopy that is in common use for process analytical measurements; the other two being near-IR (NIR) and Raman. Each one of these techniques has its pros and cons and the ultimate selection is based on a number of factors ranging from sample type, information required, cost and ease of implementation. The sample matrix is often a key deciding factor. NIR has been the method of choice for many years within the pharmaceutical industry, and sample handling has been the issue, especially where solid products are involved. IR is not particularly easy to implement for the continuous monitoring of solid substrates. However, often there is no one correct answer, but often when the full application is taken into account the selection becomes more obvious. In some cases very obvious, such as the selection of IR for trace gas analysis – neither NIR nor Raman is appropriate for such applications.

Instrumentation used within an industrial environment normally requires special packaging to protect the instrument from the environment, and to protect the operating environment from the instrument, in regard to fire and electrical safety. This places constraints on the way that an instrument, or more specifically the analyzer, is designed and the way that it performs. For continuous monitoring applications this places important demands on the mechanical and electrical design, providing the reliability and stability necessary to accommodate 24/7 operational performance. Traditional infrared instrumentation often requires significant adaptation to meet these needs, and in most cases an analyzer, which is normally a single- function system, must be developed from the ground up to ensure that the performance and reliability demands are met.

In recent years a number of new technologies have become available, many as a result of the telecommunications developments of the late 1990s. In the first edition of this book some of the newer technologies were mentioned. Since that time, some have developed further into useful tools for IR instrumentation, and others have failed to meet expectations, and in some cases they are no longer available. Two issues that accompany these technology spin-offs are miniaturization, which tends to lead to lower cost and more reliable components, and also the inherent built-in reliability of products made for the telecommunications industry. Both are attractive attributes for process analyzers. Two important trends that have emerged are the development of handheld analyzers, including handheld FTIR spectrometers, and the increased use of lasers, with special reference to broadly tunable lasers for the mid-IR spectral region (quantum cascade lasers, QCL).

Analytically, IR (FTIR) spectroscopy is unquestionably one of the most versatile techniques available for the measurement of molecular species in the laboratory today, and also for applications beyond the laboratory. A major benefit of the technique is that it may be used to study materials in almost any form, and usually without any modification; all three physical states are addressed: solids, liquids and gases. Also, it is a fundamental molecular property, and as such the information content can be considered to be absolute in terms of information content, and as such can be very diagnostic in terms of material purity and composition. Traces of impurities can be both uniquely detected and in most cases characterized. This is a very important attribute in a process analytical environment.

Before going further it is important to define IR spectroscopy in terms of its measurement range and in context of other methods of spectral measurement. It has been suggested, based on the information content, which is assigned to molecular vibrations, that the operational wavelength range for 'infrared' spectroscopy is from 800 nm to 200 000 nm (0.8–200 μm), or 12 000 cm^{-1} to 50 cm^{-1}, in frequency units. This can be separated into functionally what is defined as fundamental vibrations and overtones of the fundamentals. This leads to the classic view of IR spectroscopy, which is based on the terms mid-infrared (mid-IR) and NIR. Where mid-IR (or in this context, just IR) features both the fundamentals and in some cases overtones and combination bands, and NIR is a mix of overtones and combination bands. This delineation can be used to define the relative roles of these two areas of measurement. NIR, as traditionally applied to process monitoring applications, is discussed in Chapter 5. However, NIR applications over the years have tended to be used in well-defined areas: notably agricultural and food, chemical and petrochemical, and pharmaceutical. Applications that were once considered to be classic mid-IR applications, such as gas monitoring applica-

tions, are also being considered, in some cases (normal-high concentrations) by NIR, primarily because of laser-based technologies (use of TDLs and cavity ring down spectroscopy). Such applications of NIR will be mentioned later in this chapter.

The mid-IR region covers the fundamental vibrations of most of the common chemical bonds featuring the light- to medium-weight elements. In particular, most gases (with the exception of diatomic molecular gases such as oxygen and nitrogen) and organic compounds are well represented in this spectral region. Today, the mid-IR region is normally defined in terms of the normal frequency operating range of laboratory instruments, which is 5000 cm^{-1} to 400 cm^{-1} (2–25 μm).[1] The upper limit is more or less arbitrary, and is chosen as a practical limit based on the performance characteristics of laboratory instruments where all fundamental absorptions occur below this limit. The lower limit, in many cases, is defined by materials used for specific optical components, such as potassium bromide (KBr) with an optical transmission cutoff of 400 cm^{-1}, and zinc selenide (ZnSe) with a cutoff of around 650 cm^{-1}. An understanding of this spectral range is important because it dictates the type of optics, optical materials and optical functional components, such as sources and detectors that have to be used. Note that the spectral range in wavelength is absolute in IR, which is an absorption-based technique. Raman, on the other hand provides similar spectral information relative to the wavenumber (cm^{-1}) shift, which is correlated to molecular vibrations, but the optical components are typically made of glass or quartz, and are not constrained in the same manner as IR optics. For Raman, the spectral range is typically defined by the efficiency of the laser blocking filters (how close one can get to the laser line) and the laser wavelength combined with the type of detector that is used.

Infrared spectroscopy is one of the oldest spectroscopic measurements used to identify and quantify materials in on-line or near-line industrial and environmental applications. Traditionally, for analyses in the mid-infrared, the technologies used for the measurement have been limited to fixed wavelength NDIR filter-based methods, and scanning methods based on either grating or dispersive spectrophotometers or interferometer-based FTIR instruments. Originally, the scanning instruments were used more for laboratory-based applications, whereas filter instruments were used mainly for process, field-based and specialist applications, such as combustion gas monitoring. Today, scanning instruments are widely used, as well as filter analyzers, for process and field-based applications.

The use of classical scanning dispersive instruments in the mid-IR range is today virtually nonexistent, and so no further commentary in regard to their use outside of the laboratory will be made. Note that this differs from NIR where scanning dispersive instruments are in use for process applications. There are a few exceptions relative to dispersive instruments that should be noted; there are at least three technologies that are effectively scanning dispersive instruments that can be applied to IR measurements that are based on nontraditional approaches. These will be covered later in the instrument technologies section later in this chapter. Originally, FTIR instruments were considered to be 'too fragile' to be used outside of the laboratory. Starting in the 1980s, two companies, Bomem (now ABB) and Analect (now Hamilton Sundstrand, AIT division) changed this perception by introducing FTIR instruments designed and packaged for process monitoring applications. Since that time, several other manufacturers have introduced process hardened or dedicated FTIR analyzers, including products tailored for chemical reaction monitoring and portable analyzers.

Practical FTIR solutions have been developed by paying attention to the fundamental design of the instrument. Moving an FTIR instrument out of the benign environment of a laboratory to the more alien environment of either a process line or that of a portable device is not straightforward. A major emphasis on the instrument design in terms of both ruggedness and fundamental reliability of components is critical. Furthermore, issues such as environmental contamination, humidity, vibration and temperature are factors

[1] In the mid-IR, unlike NIR, it is customary to use microns or micrometers (μm) for the wavelength scale, not nanometers (nm). Also, for filter-based instruments the wavelength scale is traditionally used instead of the wavenumber (cm^{-1}) frequency based scale; this tends to be historical based on thin film interference filter technology.

that influence the performance of FTIR instruments. As a consequence, FTIR instrumentation intended for process and field-based applications must be designed to accommodate these factors from the start. Some of the main issues are packaging and making the system immune to the effects of vibration and temperature fluctuation, the need to protect normally sensitive optical components, such as beam splitters, and the overhead associated with making a high-precision interferometric measurement. Today, companies such as A2Technologies (Danbury, CT, USA), Block Engineering (Marlboro, MA, USA) and D&P Instruments (Simsbury, CT, USA) focus on instrument designs to meet these needs.

Vibrational spectroscopy, in the form of mid-IR, NIR and Raman spectroscopy has been featured extensively in industrial analyses, both quality control (QC), process monitoring applications and field-portable applications [1–6]. The latter has been aided by the need for advanced instrumentation for homeland security and related HazMat applications. Next to chromatography, it is the most widely purchased classification of instrumentation for these measurements and analyses. Spectroscopic methods in general are favored because they are relatively straightforward to apply and to implement, are rapid in terms of providing results, and are often more economical in terms of service, support and maintenance. Furthermore, a single spectrometer or spectral analyzer, in a near-line application, may serve many functions, whereas chromatographs (gas and liquid) tend to be dedicated to only a few methods at best.

Today, the value, functionality and benefits offered by mid-IR, NIR and Raman spectroscopy techniques are largely uncontested. Taking into account all of the techniques, virtually any form of sample can be handled, including gases, aerosols, liquids, emulsions, solids, semisolids, foams, slurries, etc. This is also a major factor when comparing chromatographic methods, which are traditional for process analysis, with spectroscopic methods. The nature of the sample plays an important role in the decision process, and in many ways it is the most demanding element for consideration when assembling an analyzer for a process application. This can also be the deciding factor between IR, NIR and Raman. The costs associated with implementing these instrument technologies is broadly comparable; typically greater than $20 000 and hopefully less than $100 000, except for the most demanding or challenging applications, and so taking cost out of the equation it often comes down to the ease of implementation and the sample. Here are some simple relevant sample and/or sampling-related guidelines.

- If used for continuous monitoring applications and if the sample is a powder then the technique of choice is most likely NIR or Raman. NIR works well with diffuse reflectance, and Raman is comparable in its method of measurement.
- If the sample is a liquid, solid or paste and is amenable to physical sampling, then an FTIR-based analyzer equipped with an ATR probe or accessory will provide a good method of measurement.
- If the sample is in the gas or vapor phase and is measured either at high or moderate concentrations then either IR or NIR can be used (Raman is generally not applicable). However, if a trace gas application is involved then IR is traditionally more applicable. Gas cells with extended path lengths up to 20 m are readily available for measurements in the ppb levels.
- If the sample is dissolved in water all three methods can be used, but as a general rule Raman has a benefit because water has a relatively weak Raman signal. If NIR or IR is used then attention has to be paid to temperature effects and impurities.
- If spectral diagnostics are required the IR spectrum (like the Raman spectrum) offers a host of well-documented diagnostic spectral data that can be used to identify and characterize materials from basic principles. This is ideal for confirming material identity and for determining the presence of contaminants, even at low concentrations.
- NIR and Raman instruments typically do not require purging for removal of water vapor; however, sealing and purging may still be a requirement for plant safety, depending on the relevant fire codes applied.

- In order to get usable spectral intensities from liquids mid-IR requires a short path length (often 0.1 mm or less) or an ATR method of sample handling. In the NIR the optical path lengths tend to be much longer ranging from 1 mm up to 50 or 100 mm (SW-NIR, 700–1100 nm).

IR spectroscopy, both in the forms of NIR and mid-IR spectroscopy, was once hailed as a major growth area for process analytical instruments [7–9]. This view assumes that traditional laboratory instruments are outside of those covered by the areas concerned. Overall, it is difficult to define the boundaries between the laboratory and the process; where they start and end. Some of the confusion arises from the term 'process analysis' itself. In the most liberal form, it can be considered to be any analysis that is made within an industrial environment that qualifies or validates a material or a product. This ranges from the QC of incoming raw materials, to the control of the manufacturing process through its various intermediate stages, and on to the analysis of the product. How this is viewed tends to be industry dependent. Today, portable instruments readily bridge the gap and can fulfill multiple roles from the laboratory to the plant floor to external locations around a plant. The role of process analytical technologies (PAT) has been associated mostly with the pharmaceutical industry. This industry is unique in its definition of process and the usage of process instrumentation. Traditionally, process instruments have been discussed in the context of continuous processes as encountered in the petroleum and chemical industries. Analyzers designed for continuous processes often have more stringent requirements than instruments for batch operations, and this is primarily because of the need to operate 24/7 and with little or no down time, other than scheduled maintenance. In a pharmaceutical operation process analyzers can cover the range from systems on a cart used for incoming raw material quality control, to mobile analyzers used for reaction monitoring, to handheld instruments for vessel cleanliness checking, to hardened analyzers that are strapped to blenders to on-line monitoring systems for dryers and solvent recovery systems. All of these systems can be considered to be part of 'the process', and as such can be considered to be process analyzers.

The ideal role for a process analyzer is to provide feedback to the process control system, providing real-time quality information and helping to optimize production efficiency. For some industries product quality is the most important controlling parameter, and this requires analytical controls throughout critical stages of production. This is a key component of the PAT initiative that is being endorsed by the FDA and the pharmaceutical industry. Other examples include high-tech manufacturing, such as in the semiconductor industry, and specialty or value-added chemical production. It is important to review the physical implementation of the process analyzer, which gives rise to terms such as on-line, off-line, near-line, at-line, and even remote and portable methods of analysis. In general, these descriptions are self-explanatory and the only question tends to be; which is the best way to implement an analysis? The important deciding factors are the turn-around time required for the analysis and the cost and ease of implementation. Whatever approach is used, it is important to address the instrumentation as a complete package or a system. Depending on the requirements of the analysis and the technique selected, a system consists of several key components; the sampling system (how the sample is extracted from the process, and how it is conditioned), the sample interface, the basic spectrometer (including how it is packaged) and the way that the system communicates back to the process. Relative to this last point, if the measurement system is not permanently on-line, in a continuous mode of operation, then the issue may be addressed from human interface or via a networked communication (wired or wireless).

6.2 Practical Aspects of IR Spectroscopy

As noted in the previous section, the mid-IR the spectral region is generally characterized by the range 5000–400 cm^{-1}, where the region below 400 cm^{-1}, is sometimes classified as the *far infrared*, and is assigned

to low frequency vibrations. Note that today, the lower frequency vibrations fall into the domain of terahertz spectroscopy, and instrumentation specifically designed for those measurements is now commercially available. Note that relative to NIR there is some overlap here because the classical definition of NIR is the region between the infrared and the visible part of the spectrum, which is nominally from 700 nm to 3300 nm (0.75–3.3 μm, or approximately 13 300–3000 cm^{-1}). On older IR instruments, and with some of today's instruments, the default range may start from 4000 cm^{-1}. Having said that, and for the record, there is one fundamental vibration that occurs above 4000 cm^{-1}, and that is the H-F stretching frequency of hydrogen fluoride gas which extends to around 4200 cm^{-1}.

There are pros and cons to consider when considering mid-IR versus NIR for process monitoring applications. The biggest advantage of mid-IR is its broad applicability to a multitude of applications based on the information content of the mid-IR spectrum [9–12]. Analytically, the fundamental information content of the NIR is limited and varies throughout the spectral region, with different regions used for different applications. In all cases, the measurements utilize absorptions associated with overtones and combinations from the fundamental vibrations that occur in the mid-IR region [13]. For the most part, these regions are dominated by absorptions from the 'hydride' functionalities – CH, OH, NH, etc. Three groups of information are observed, and these are nominally assigned to first, second and third overtones of the hydride fundamentals, combination bands, plus some coupling from vibrational frequencies of other functional groups. Note that most functional groups are measured directly in the mid-IR region, but are only measured by inference in the NIR.

The issue of intensity is a relevant and practical issue. Moving through the overtones in the NIR the intensity is reduced by orders of magnitude (typically between one and three orders). This has an important impact on the way that one views the sample and the method used for the sample interface. Sample thickness or path length is a critical part of a quantitative measurement. For condensed phase IR measurements the path length is generally measured in micrometers, or tenths of millimeters (typical range of 0.01–0.2 mm), whereas millimeter and centimeter path lengths are used in the NIR to measure comparable levels of light absorption. The longer path length simplifies sampling requirements for all condensed-phase materials, and in particular for powdered or granular solids. It also favors more representative sampling. On the down side, the presence of particulate matter, which leads to light scattering and loss of energy, has a larger impact with shorter wavelengths and with larger path lengths. Liquid flow applications in IR can be difficult to implement if a transmission measurement is made, however today this limitation has been overcome by the use of liquid ATR (see below) measurements. The latter has benefits in terms of implementation because the impact of particulates and bubbles can be minimized in a flowing system. For low level gas or vapor phase measurements the mid-IR region is generally preferred.

Optical construction materials, as used in cell windows, optical accessories (such as ATR) and also for light conduits (including optical fibers) is another important practical area for consideration. Common silicate materials such as quartz and glass are transparent throughout the NIR, in contrast to most of the mid-IR region (note that anhydrous quartz is transparent to around 2500 cm^{-1}/4.0 μm). Often hygroscopic or exotic materials are used for sampling and instrument optics within the mid-IR region.

A majority of traditional NIR measurements are made on solid materials and these involve reflectance measurements, notably via diffuse reflectance. Likewise, in the mid-IR not all spectral measurements involve the transmission of radiation. Such measurements include internal reflectance (also known as attenuated total reflectance, ATR), external reflectance (front surface, 'mirror'-style or specular reflectance), bulk diffuse reflectance (less common in the mid-IR compared to NIR), and photoacoustic determinations. Photoacoustic detection has been applied to trace-level gas measurements and commercial instruments are available based on this mode of detection. It is important to note that the photoacoustic spectrum is a direct measurement of infrared absorption. While most infrared spectra are either directly or indirectly correlated

to absorption, it is also possible to acquire infrared spectra from emission measurements. This is particularly useful for remote spectral measurements and in an industrial environment this is especially important for hot fugitive emissions and stack gas emissions from plants.

6.3 Instrumentation Design and Technology

In its most basic form, infrared spectroscopy is a technique for the measurement of radiation absorption. By convention, the measurement is made on a relative basis, where the absorption is derived indirectly by comparing the radiation transmitted through the sample with the unattenuated radiation obtained from the same energy source. For most measurements, the resultant signal is expressed either as a percentage of the radiation transmitted through the sample or reflected from the sample. For many process measurements, featuring simple filter-based NDIR analyzers the measurement may be made at a nominal single wavelength, or a set of fixed wavelengths. However, in common with most laboratory instruments, many modern IR process analyzers utilize multi-wavelength information, often acquired from a scanning spectrometer. Irrespective of the actual style of instrument used, the main components of an IR analyzer can be summarized in the form of a block diagram (Figure 6.1). Fundamentally, this is a generic format that features an energy source, an energy analyzer, a sampling point or interface, a detection device, and electronics for control, signal processing, and data presentation.

Although not indicated in Figure 6.1, the sample conditioning and the nature of the optical interface used to interact with the sample are as important, if not at times more important, to the success of the measurement than the actual instrument technology itself. Note that the generic format used in the illustration also applies to emission measurements, where the sample is in essence the source. As illustrated, the configuration implies that the radiation passes through the sample, and this only applies to transmission measurements. For reflection applications, the surface of the sample is illuminated, and the reflected radiation is collected from the same side. Another variant on the theme (as illustrated) is the remote sampling approach where the sample is located away from the main spectrometer and the radiation to and from the sample is carried

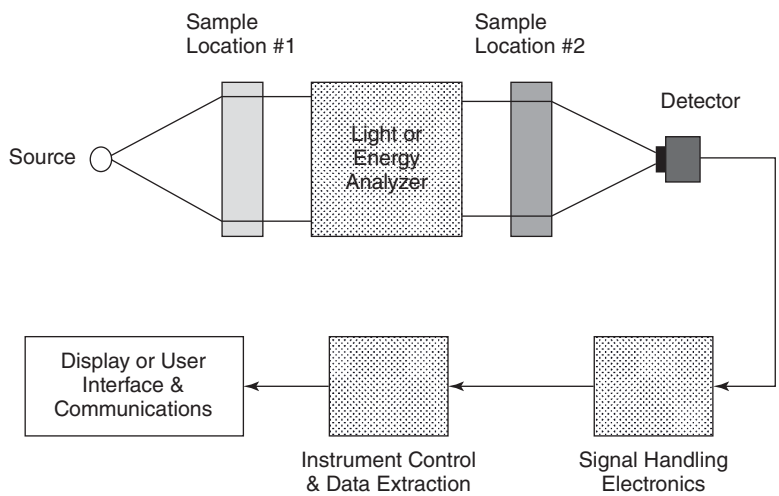


Figure 6.1 *Generic block diagram for a spectral analysis system.*

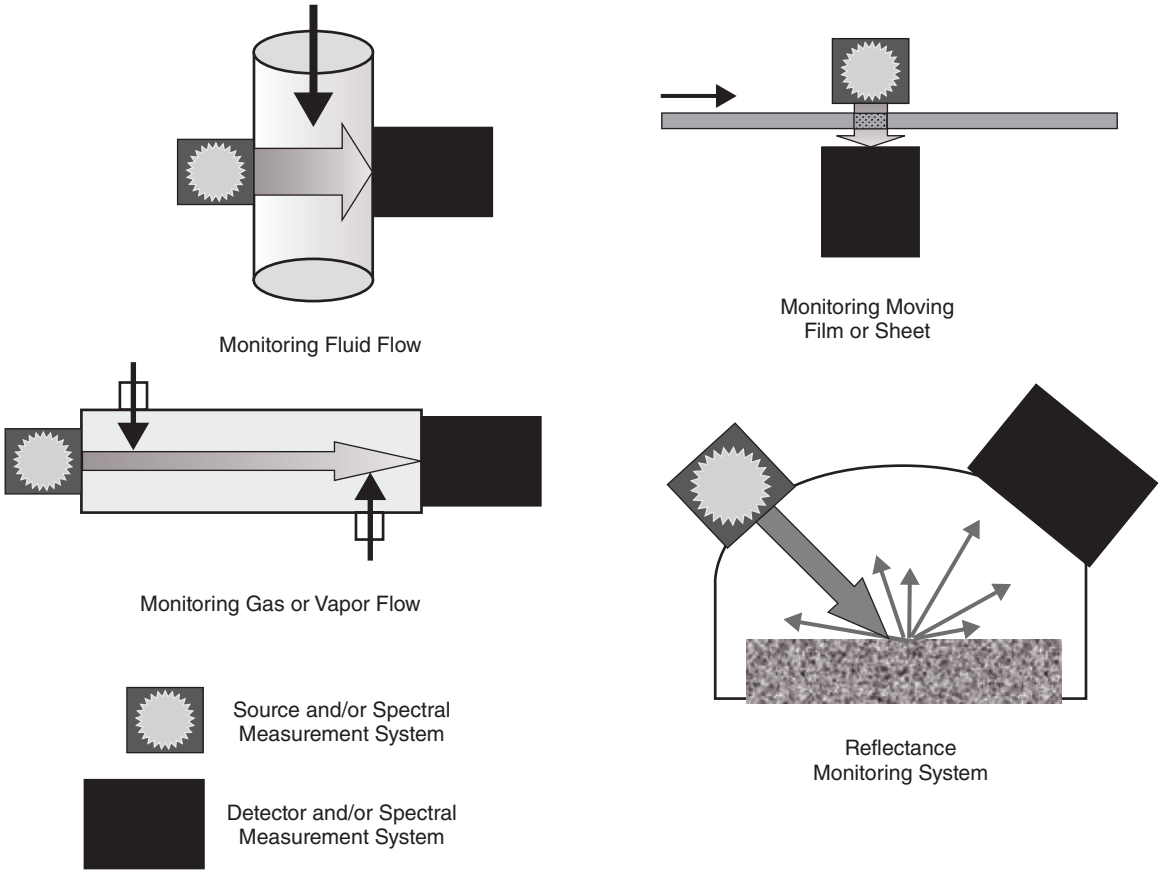


Figure 6.2 *Common sampling configurations for spectral process analyzer systems.*

via optical fibers or some form of light conduit. Figure 6.2 illustrates some of the possible component-sample configurations that are used for infrared spectral measurements.

There are several measurement techniques that can be considered for use in IR-based instrumentation and these are summarized in Table 6.1, with cross reference to NIR measurement methods. Some of these techniques are classical, as in the case of optical filter-based instruments, scanning monochromators and interferometers. Detector arrays, based on MEMS fabrication are now becoming available in the IR region, with a relatively flat response throughout most of the spectrum. Tunable lasers are becoming available, and narrowly tunable lasers have been used for some time for gas measurement applications. Broadly tunable lasers, known as quantum cascade lasers (QCLs) [14–18] covering the spectral range from 6 to 12 μm, in increments of up to 2 μm, are also available but at the time of writing the devices are still expensive. In the future, with larger scale production, there is an opportunity for these devices to become available at a relatively low cost and with broader tuning ranges.

There are two ways to view instrumentation; in terms of the method used to separate the infrared radiation into its component parts as used for the analysis, or in terms of the enabling technologies used to selectively capture the relevant radiation required to perform the analysis. The selection of the approach used for the measurement, in line with the techniques outlined in Table 6.1, tends to dictate which view is the most relevant. To explain this we can consider the following examples:

Table 6.1 *Measurement technologies available for NIR and mid-IR instrumentation*

Measurement technology	NIR	Mid-IR
Fixed and variable optical filters	X	X
Correlation filters (example gas correlation filters)		X
Scanning slit monochromators	X	X
Fixed polychromators	X	X
Detector arrays (photonic and thermal)	X	X
Broadband tunable filters (examples AOTF, LCTF)	X	
Narrowly tunable diode lasers, TDL	X	X
Broadly tunable lasers (example; QCL)		X
Interferometers (FTIR)	X	X

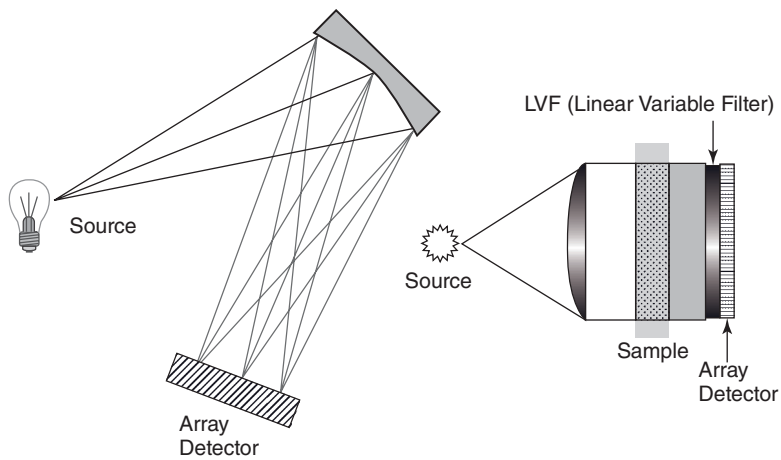


Figure 6.3 *Typical configurations for array-based spectral analyzers.*

- *An FTIR instrument:* The three critical components (excluding the sample) are the source, the detector and the interferometer. In terms of enabling technology it is the interferometer that is critical to the measurement.
- *A detector array spectrograph:* The three critical components (excluding the sample) are the source, the light separation technique that can be a variable filter or a standard diffraction grating-based spectrograph, and the detector array. In terms of the enabling technology it is the detector array that is critical in the way the instrument operates. Figure 6.3 provides typical configurations for array-based spectrometers. Typically, they are simple in construction, and they feature no moving parts.
- *A laser spectrometer:* In this case, assuming that the laser is tunable, there are only two critical components, the tunable laser and the detector. Typically, the enabling technology is the laser, which in this mode acts as the light source and the wavelength selection device. Note that in a broadly tunable laser there can be optical subsystems that perform wavelength scanning and selection.

Note that the examples above to some extent are oversimplifications, and other critical elements, such as optics, the associated electronics and the sampling interfaces are also important for a final implementation.

The latter is particularly the case for laser instruments; where attention to optical interfacing is essential, and often features specialized optical fiber based optical coupling.

Next, when instrumentation is implemented in an industrial environment there are many other parameters associated with the final instrument packaging and the electrical and fire hazard safety that need to be addressed. Often these issues are not trivial and can have a profound impact on both the design and final performance of an instrument. In the following sections instrumentation and component technology will be reviewed, and where appropriate the impact of operation in an industrial environment will be addressed.

6.4 Process IR Instrumentation

It takes a lot more than just connecting an infrared instrument to a pipe to make the instrument into a process analyzer. The system not only has to be constructed correctly and be capable of flawless operation 24-hours a day, but it also has to be packaged correctly, it has to be able to access the stream correctly and without any interference from process related disturbances, such as bubbles, water, particulates, etc. Finally, the data acquired from the instrument has to be collected and converted into a meaningful form and then transmitted on to the process computer, where the information is used to assess the quality of the production and to report any deviations in the production. Beyond that point, the process computer, typically not the process analyzer, handles any decisions and control issues. There are exceptions where robotics-based devices are integrated with the analyzer, and good versus bad product are assessed at the point of measurement. The use of FTIR analyzers for semiconductor wafer quality testing is an example, as mentioned below. Today, references to process analytical instrumentation, and the term PAI relate to the totally integrated system, implemented on-line, and made up of some or all of the components discussed in the following sections.

The term process analyzer sometimes conjures up the image of plumbing and a highly ruggedized instrument in a heavy-duty industrial enclosure. Not all analyzers used for process applications have to follow that format. For example, in the semiconductor industry, the traditional format of analyzers are used as on-line gas and water sensors and are considered as process monitors, but product monitoring analyzers, which are configured more as automated laboratory analyzers are viewed differently. The semiconductor wafer, the product, is monitored at various stages of the process, effectively *in-line* or *off-line* by FTIR analyzers. It should be pointed out that FTIR is the most commonly accepted method for the QC of the product, Examples being the ASTM Standard methods F1188/F1619 for interstitial oxygen measurement, and F1391 for substitutional carbon content. In a process that can involve up to 32 critical steps, and where a high yield at each step is essential, it is unreasonable not to consider FTIR as a true process analysis tool, in spite of its laboratory-style analyzer format. The application is made practical by the use of high-precision wafer-handling robotics, which are integrated into the analyzer.

IR photometers have also been considered as a standard instrument format for infrared measurements. These are used as on-line systems and are mainly used in the environmental market (combustion gases) and for the measurements of standard process gases, such as hydrocarbons. Today, there are a relatively large number of on-line process stream analyzers for liquids, solids and gases (mid- and NIR) that feature either filter (discrete) or full-spectrum analyzers. In many industries applications featuring 'near-line' or 'at-line' are now considered to be appropriate especially when the timescale does not require real-time measurements. Many of these analyses are legitimately viewed as process measurements compared to the direct on-line analogs. Such instruments are, however, often easier and much less expensive to implement than an on-line instrument. The final implementation and format are usually based on the nature of the process, the needs of the end user, the overall economics, and the required speed of analysis. In many cases this comes down to whether the process itself is operated in a batch mode or is continuous. A good example is the of a laboratory platform adapted for near-line and even mobile applications is the VFA Spectrometer (Wilks Enterprise,

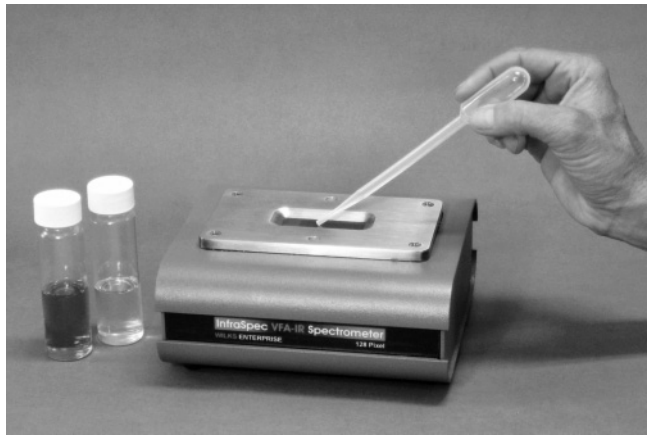


Figure 6.4 *A commercial IR Spectrometer for near-line, at-line and mobile IR analysis; example shown is the Wilks Enterprise VFA-IR Spectrometer.*

South Norwalk, CT, USA) as illustrated in Figure 6.4. This system is fully integrated, cost effective and can be programmed for pre-calibrated applications.

6.4.1 Commercially available IR instruments

As noted, IR instruments fall into several categories that range from simple photometers (single or multiple filter devices) to relatively complex full-spectrum devices, such as FTIR instruments. Today's process engineers often prefer optical methods of measurement rather than the traditional chromatographs because of perceived lower maintenance and ease of implementation. However, the final selection is often based on the overall economics and the practicality of the application.

6.4.1.1 A Historical Perspective

For the measurement of a single chemical entity, analyzers in the form of NDIR analyzers have been in use for decades for both industrial and environmental monitoring applications. These feature one or more wavelength specific devices, usually optical filters, customized for the specific analyte, such as CO, CO_2, hydrocarbons, moisture, etc. This type of instrument is normally inexpensive (on average around $5000) and relatively easy to service, dependent on its design. This class of instrument is by far the most popular if one reviews the instrumentation market in terms of usage and the number of analyzers installed.

Filter-based instruments are often limited to applications where there is simple chemistry, and where the analytes can be differentiated clearly from other species or components that are present. Today, we may consider such analyzers more as sensors or even meters, and the analytical instrument community does typically not view them as true instruments. Since the late 1980s a new focus on instrumentation has emerged based on the use of advanced measurement technologies, and as such is considered to be more of the consequence of an evolution from laboratory instruments. Some of the first work on full-spectrum analyzers started with an initial interest in NIR instruments. The nature of the spectral information obtained in the NIR spectral region is such that an analyzer capable of measuring multiple wavelengths or preferably a full spectrum is normally required.

In parallel to the development of full-spectrum NIR analyzers, work started on the use of mid-IR spectrometers. By the mid-1980s, FTIR became the *de facto* method for measuring the mid-IR spectrum in the laboratory. For many applications, FTIR was viewed as being too complex and too 'fragile' for operation on the factory floor or for implementation on a process line. Some pioneering FTIR companies, notably Bomem, Nicolet and Analect, proved that the concept would work with adaptations to the design of laboratory instruments. These companies, starting around the mid-1980s, and later others, made the decision to focus on industrial applications, and designed instruments for operation in more demanding environments.

Today, most of the major FTIR companies offer some form of instrumentation that is used in process monitoring applications. Some are application specific; others are industry specific, the pharmaceutical industry being a good example. Companies such as Hamilton Sundstrand (AIT Division, Pomona, CA, USA), Midac (Costa Mesa, CA, USA) and ABB (Quebec City, Quebec, Canada) tend to be 100% focussed on process applications, whereas a company such as Brüker (Billerica, MA, USA) supports both laboratory and process applications. All of these companies offer on-line and off-line implementations of their products. One other company worth mentioning here is Mettler-Toledo AutoChem (Columbia, MD, USA) for its ReactIR kinetics and reaction monitoring system. This is a highly targeted product that has played an important role in a number of applications, including Pharmaceutical R&D. Other FTIR companies, notably PerkinElmer (Shelton, CT, USA), Shimadzu (Columbia, MD, USA), Jasco (Easton, MD, USA) and Thermo (Madison, WI, USA) tend to lean towards the off-line applications, and largely in the consumer and pharmaceutical or the 'high-tech' application areas. The companies mentioned are the ones that have formed the core of the traditional instrument business. There are many less well-known or nontraditional companies that provide FTIR-based analyzers for specialized applications. Two interesting examples are Block Engineering (original designer of the Digilab instruments) and D&P Instruments. Over the years these two companies have produced instruments that have been developed for government programs, many of which involve external monitoring applications. Although not applied to process applications at this time, the ruggedized constructions on the instruments would readily lend themselves to plant applications. Another spin-off from government based programs has lead to the development of handheld systems, and Smiths Detection, A2Technologies and Ahura (Wilmington, MA, USA) are good examples of companies that have adapted the products to applications beyond Homeland Security roles.

In the ensuing sections the basic instrument technologies will be discussed as used in commercial instrumentation. A section will be included that reviews some of the emerging technologies that are used either as replacements or substitutes for the standard technologies, often for very specific applications.

6.4.1.2 Photometers

One of the oldest forms of process analyzers are the filter photometers, sometimes also known as filtometers [19]. They have simple construction, and in the simplest form they consist of a source, an optical filter (or filters), a sample cell and a detector (Figure 6.5 provides example presentations). For many applications they are single-channel devices, and consequently operate as a single beam instrument. Miniaturization has lead to the production of smaller more integrated devices that essentially behave as single species sensors. In each case, the selected filter, the source and the detector define the analytical spectral region. For the simplest versions, the filters can be integrated with the detectors, where the filter forms the window for the detector (as indicated in Figure 6.5B). Assemblies featuring two- (dual) and four- (quad) detector channels are available in a single package as indicated in Figure 6.6A, B). In such a package, a complete two- to three-component analyzer can be constructed with just a source, sample cell and a dual or quad detector. One other unique system worth mentioning is the species specific MEMS-based source/detector combina-

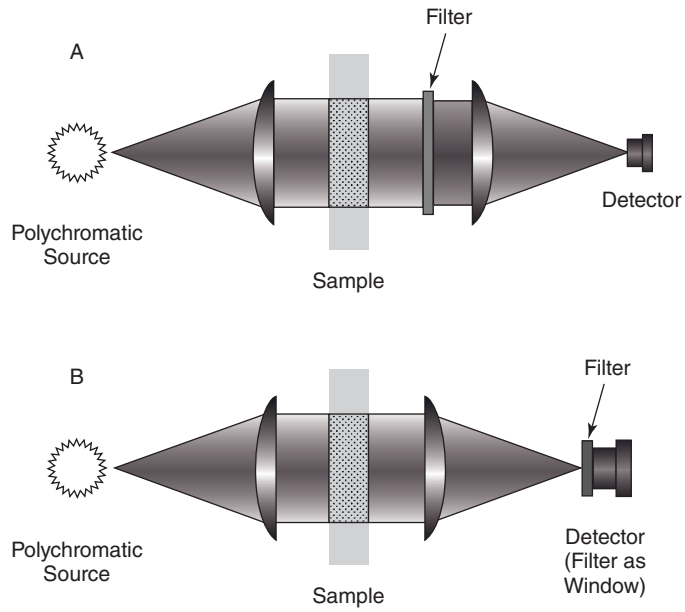


Figure 6.5 *Block diagram for simple filter photometer analyzers.*

tions offered by ICx (originally Ion Optics). The current systems offer carbon dioxide detection, and other species such as methane are also being considered.

With a traditional photometer a baseline is established with a zero sample (contains none of the analyte), and the span (range) is established with a standard sample containing the anticipated maximum concentration of the analyte. Although the electronics are simple, and designed for stability and low noise, it is usually necessary to establish zero and span settings on a regular basis, often daily, and sometimes more frequently. This can result in significant calibration overhead costs for even the simplest photometer. Note that often the lower instrument cost is negated by higher operational costs.

Conventional wisdom leans towards a simple measurement system featuring one or two optical elements and with a simple construction. In practice there are many forms of filter assemblies that can be used and these are not limited to the use of a single filter and implementations featuring multiple filters on a wheel are often implemented in multichannel analyzers. Such assemblies (Figure 6.6C) can be quite small (25 mm diameter as illustrated) and can accommodate a relatively large number of filters. Other forms of filter assemblies include clusters or arrays (multiple wavelengths), correlation filters (customized to the analyte), continuously variable filters, such as the circular variable filter (CVF) and the linear variable filter (LVF), and tunable filters (both broad and narrow band).

Variable filters, in the form of CVFs and LVFs are not new, and their use in scanning instruments featuring single element detectors has been reported for least 30 years. Note that commercial instruments based on these filters were available in the 1970s, the best examples being the Miran analyzers, originally offered by Wilks. The Miran gas analyzer was fully integrated and featured a 20 m path length gas cell, and was made fully portable with an add-on battery pack. The analyzer was a mechanical device and internally was relatively complex. Today, detector arrays are becoming available, and linear variable filters (LVFs) can be integrated with these arrays to provide a multiwavelength spectral measurement system. Such a combination leads to compact instrument for simple transmission, ATR, or diffuse reflectance measurements, and

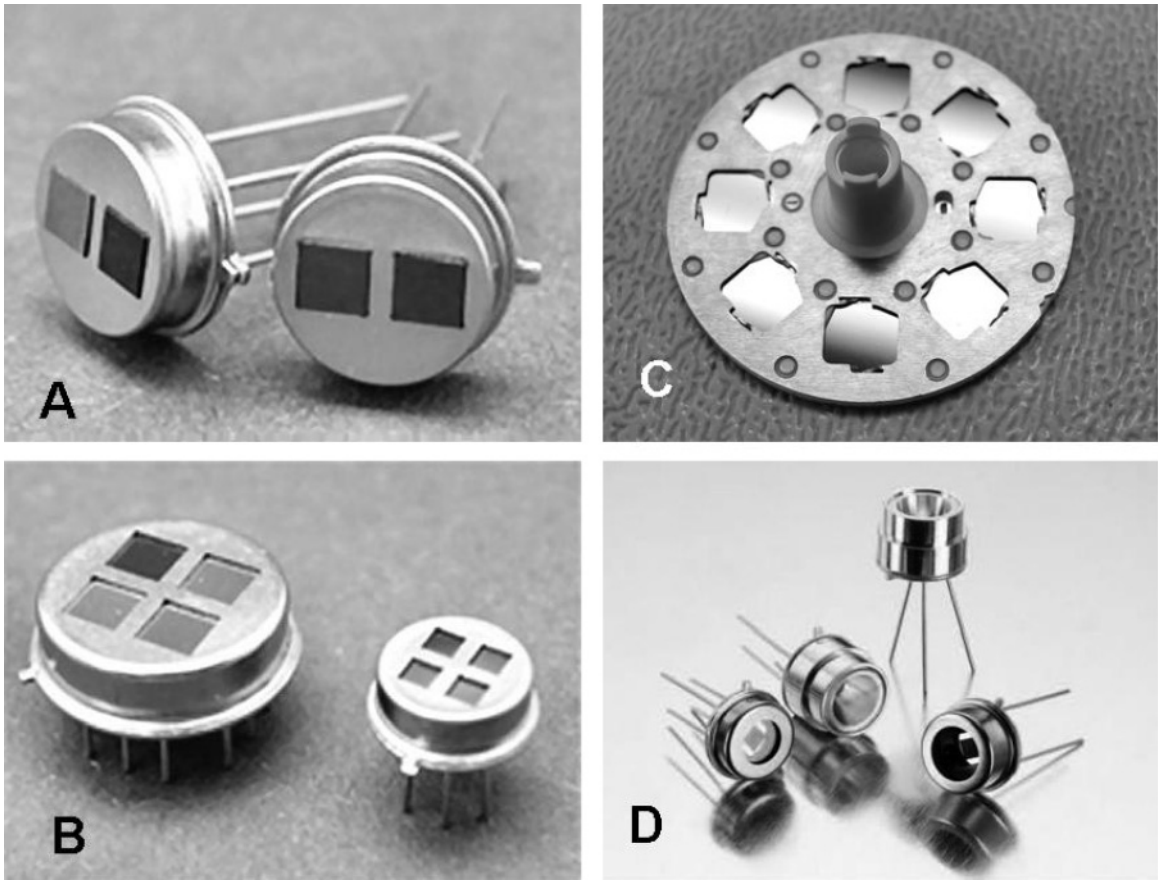


Figure 6.6 *Example components available for simple photometers: (A) a dual channel detector, (B) a quad, four-channel detector, (C) an example filter wheel with 8-filter channels, and (D) example MEMS-based IR sources.*

removes all issues associated with moving parts. An example of such an analyzer integrated with a horizontal ATR sampling accessory was featured earlier (Figure 6.4). A significant factor favoring of this type of design is its compact size, which leads to an environmentally stable system that has no requirement for purging, both big advantages in an industrial environment.

Tunable filters provide the benefit of rapid spectral tuning, providing narrow range spectral acquisition in about 100 ms, sometimes less. Most of these tunable devices, usually MEMS-based Fabry–Perot etalons, can be step scanned where the device is tuned to specific wavelengths. This is an extremely fast process, and emulates a multiwavelength filter system, with near-instantaneous output at each wavelength. Example devices have emerged from the telecommunications industry and have the potential to lead to the development of highly reliable tunable spectral measurement devices, albeit with a narrow spectral range.

6.4.1.3 Monochromators and Polychromators

Traditional optical spectrometers for both mid-IR and NIR were based on a scanning monochromator. This design features a single source and detector, and a mechanically scanned dispersion element in combination

with optical slits for spectral line separation. The dispersion element, either a refracting prism or a diffraction grating, and the slits were driven mechanically by relatively complex mechanisms. These instruments are now referred to as dispersive instruments. Although this arrangement worked adequately in a laboratory environment for years, it is now considered to be inappropriate for most applications, particularly the case for a heavy industrial setting. Reliability, calibration, and overall performance (or lack of) have been factors that limit the use of a monochromator. Today, a few commercial NIR instruments still feature monochromators, but without exception interferometer-based FTIR instruments have replaced dispersive mid-IR instruments.

Monochromators, theoretically, provide energy of one wavelength (or practically a few wavelengths) at a time; the device scans the spectrum sequentially. In contrast, the polychromator allows multiple wavelengths to be measured simultaneously. A polychromator is used in instruments that feature detector arrays or other arrangements that can handle multiple wavelengths at a time via unique methods of wavelength encoding. In a simplest form of polychromator, a concave diffraction grating spatially separates the source energy into the analytical wavelengths and provides the necessary imaging on the array. The advantages of array-based instruments are obvious based on the fact there are no moving parts, they are mechanically simple and good wavelength stability is usually achieved leading to highly reproducible measurements in a factory or plant environment. An instrumentation technique that utilizes the output from a monochromator/polychromator, and that provides some of the benefits of multiplexed data acquisition, is known as a Hadamard transform spectrometer. This class of instrument can feature either a monochromator or a polychromator dependent on the design and the detection system. A digitally encoded spectrometer system, very similar in concept to a Hadamard transform spectrometer is commercially available from Polychromix (Wilmington, MA, USA) operating in the NIR spectral range. This system features what is effectively a digitally tunable grating that feeds the encoded optical signal to a single element detector. At this time the company has focused on NIR spectral measurements, however, a change in the MEMS grating component and a change in detector is all that is required to convert the system into an IR spectrometer.

6.4.1.4 Interferometers

During the 1980s most mid-IR instruments moved away from dispersive methods of measurement (monochromator-based) to interferometric measurements based on the widespread introduction of FTIR instrumentation [20,21]. These instruments provided the performance and flexibility required for modern day mid-IR applications. At the heart of an FTIR instrument is a Michelson-style of interferometer. The critical elements in this style of instrument are the beam splitter and two mirrors: one fixed and the other moving (as illustrated in Figure 6.7). This type of measurement requires the production of an optical interferogram, which is generated at high optical precision from the mirror–beam splitter combination. The mechanical and optical tolerances on the interferometer design are extremely demanding, a factor that leads to many challenges for instruments placed in an industrial environment.

The drive mechanism of the moving mirror assembly, which eventually yields the spectral information, has to be critically controlled, and must be made immune to external interferences, such as temperature and vibration. Both can have significant impact on the performance of an interferometric measurement system. In the mid-IR region, the beam splitter is an environmentally fragile optical element, which is normally constructed from a hygroscopic substrate such as potassium bromide (KBr). This substrate has to be maintained to high optical flatness, and is coated with multiple layers of high index materials, such as germanium. This critical element must be environmentally protected, in particular from dust and humidity. At a minimum the instrument needs to be hermetically sealed, dry gas purged, or protected by desiccant. In an industrial environment, purging can be difficult, and so sealing and desiccation are normally required. For demanding industrial applications, nonhygroscopic materials, such as zinc selenide, are used for beam

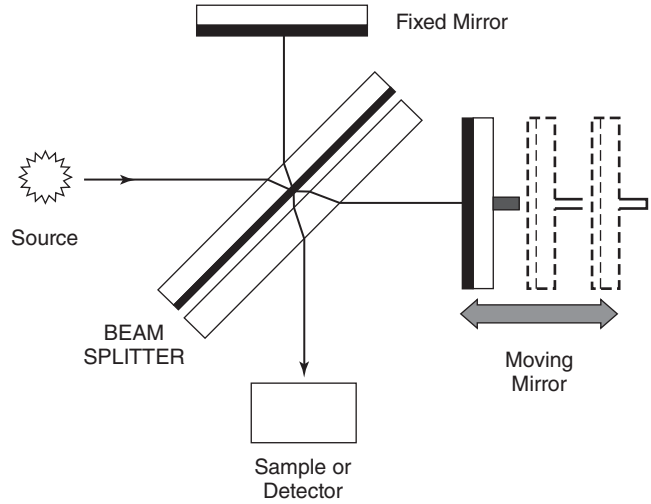


Figure 6.7 *Basic components for the Michelson interferometer.*

splitter construction. Temperature is another important factor, and unless well designed and/or protected the interferometer will be inherently sensitive to changes in ambient temperature. In acute situations the interferometer can lose optical alignment, with consequent losses in stability and energy throughput. Instruments designed for process applications often feature thermal compensation in the beam splitter design, and will also feature some form of controlled mirror alignment, to compensate for thermally induced changes in the interferometer geometry. One unique design, relative to the drive system was developed by D&P Technologies where the optical path difference is generated by a rotating refractive optic. This system is intended for high speed data acquisition, where the high speed of rotation effectively removes or reduces the mechanical drive issues of a conventional interferometer.

It is possible to construct a rugged interferometer suitable for use in a factory or production environment with careful attention to design details. Manufacturers specializing in process and portable applications have focussed on these issues and a range of ruggedized systems is available, suitable for process monitoring applications. Operational issues can arise, mainly because of the need to maintain good optical alignments, and the fact that an FTIR instrument still features critical moving parts. However, earlier new portable FTIR spectrometer systems were mentioned, produced by Ahura and A2Technologies (also used by Smith Detection), and in both cases the systems are highly ruggedized and appear to be immune to the normal environmental issues that plague FTIR.

6.4.2 Important IR component technologies

This section is focused primarily on source and detector technologies. For some applications the source or the detector actually defines the entire measurement technology, for example tunable lasers (source) and array spectrographs (detector). There are other important technologies to consider, especially in the area of data acquisition, control, computer, and communication technologies. These are rapidly changing areas, and if viewed generically, service all forms of instrumentation. Where practical, companies tend to use standard platforms, but for certain applications, where performance is critical, there is still a case for proprietary solutions.

Beyond performance optimization, issues relative to packaging and the need for compliance with certain safety and electronics regulatory codes are cited as reasons for a customized solution. In the latter case, a systems approach is required, especially when attempting to meet the code or performance requirements for compliance with European Certification (CE) mark or electrical and fire safety codes such as National Fire Prevention Association (NFPA) and CENELEC (European Committee for Electrotechnical Standardization). Off-the-shelf electronics may provide the necessary performance characteristics for generic applications, and their use eliminates large expenses related to product development, plus the associated time delays. Photonics-related components are solely addressed in this section because they are used to customize instruments for application-specific systems.

Source and detector selection are interrelated, where the output of the source is matched to the sensitivity range of the detector. However, the exact nature of the source is also dependent on the type of sample(s) under consideration, the intended optical geometry, the type of measurement technique, and the final desired performance. The bottom line is that adequate source energy must reach the detector to provide a signal-to-noise performance consistent with the required precision and reproducibility of the measurement. This assumes that the detector is well matched, optically and performance-wise, and is also capable of providing the desired performance.

6.4.2.1 Broadband Sources

Most optical spectral measurements, where the measurement of multiple wavelengths is required, will feature some type of polychromatic or broadband light source. There are a few exceptions here, such as tunable laser sources and source arrays. In such instances, the source is effectively '*monochromatic*' at a given point in time. These sources are covered separately under monochromatic sources.

Most sources intended for IR measurements are thermal sources, and feature some form of electrically heated element, providing a characteristic blackbody or near blackbody emission. There are essentially two types of broad-band sources used in IR instrumentation: the open element source and the enclosed element source. The latter is effectively a light bulb, with tungsten lamp with a quartz envelope being one of the most important. Note that this source does not transmit energy much below 5.5 to 5.0 μm (2000 cm^{-1}). A variant on the '*light bulb*' is an enclosed electrically heated element within a standard electronics package, such as a TO-5 style can (i.e. packaging style). The unit is sealed with a low-pressure atmosphere of a protective inert gas, such as nitrogen, by an IR transmitting window, such as potassium bromide (KBr), zinc selenide (ZnSe), sapphire or germanium (Ge). The thin filaments used in bulbs and bulb-like sources have a low thermal mass, and this provides the opportunity to pulse the source. A variant of this source, provided in similar packaging, is MEMS fabricated and has a diamond-like surface coating (Figure 6.6D), which provides exceptionally good thermal and oxidative stability. All of the above sources are particularly useful for process related applications; including handheld monitoring systems because they have low power, high efficiency and can be electrically modulated.[2] The modulation depth of such sources is often greater than the traditional incandescent light bulb and modulation rates of as high as 10 to 20 Hz, comparable to mechanical chopping systems used on traditional dispersive instruments.

The other forms of blackbody sources are adaptations of those used widely in conventional laboratory-style IR instruments, which feature exposed electrically heated elements. Various designs have been used, with metal filaments, made from Kanthral and Nichrome, being simple solutions in lower cost laboratory

[2] Most detector systems require that the IR beam be modulated, where the source energy is adequately differentiated in the measured signal from the ambient background. One of the traditional approaches is to use some form of mechanical 'chopper', usually in the form of a rotating sector wheel, which modulates the beam by blocking the radiation in one or more sectors during a rotation. Note that this is not a requirement for FTIR systems where the interferometer naturally modulates the beam.

instruments. They tend to be mechanically fragile, especially when subject to vibrations while hot. Electrically heated ceramic sources are currently popular and these are more robust and reliable than simple wire sources. These include sources made from silicon carbide, with the Globar and the Norton sources being common examples. The Norton source is used extensively on modern low-to-medium cost FTIR instruments and for most instruments used for process applications. It is a simple, rugged source, which was originally used as a furnace igniter. Typical operating temperatures are in the 1100–1300 K range, with the Globar being an exception at around 1500 K. Note that the Globar usually requires some form of assisted cooling and is seldom used for process applications.

6.4.2.2 Monochromatic Sources

The most common forms of 'monochromatic' source are the LED (light emitting diode) and the laser. Today, these types of source are becoming popular for a wide range of spectral measurements for both instruments applied to specific applications and dedicated chemical sensors. LEDs are typically limited to NIR spectral measurements although some devices are now available that operate in the short wavelength end of the IR region. Lasers are now available operating in the mid-IR, from 3–30 μm. Tunable lasers have always been the dream of spectroscopists. The concept of having a small, high-powered, narrow beam monochromatic IR light source that can be tuned rapidly to other wavelengths is considered the ultimate for many process-related applications. Today, solid-state IR lasers are tunable over both narrow (TDL devices) and relatively broad wavelength (QCL) ranges. While the tuning range of TDL devices is typically very small, often no more than a few nanometers, it is sufficient to be able to scan through the narrow line spectra of certain molecular gas species. The devices are useful for the measurement of low concentrations of specific gases and vapors, with detection limits down to parts-per trillion (ppt) being reported. The concept of using an array containing several laser elements has been made available for handling gas mixtures. Tunable laser sources have been used for monitoring species such as hydrogen fluoride, light hydrocarbons, and combustion gases. The broadly tunable QCL laser is expected to have a major impact on the design of IR instruments in the next 2 years. While current systems are still limited in range and costs are still high, the projection is that tuning ranges in the future could cover from 6 μm (1665 cm^{-1}) to 12 μm (830 cm^{-1}), and possibly broader, and at costs in the \$10 000–20 000 range. While this does not encompass the entire defined IR range, it does cover a majority of chemical species, and it is focused on the fingerprint region, the most important region of the spectrum for most compounds.

Other non-continuum sources, which are effectively multiwavelength 'monochromatic' (because of line width) or multiple line sources, such as gas discharge lamps, have been used, and are often considered to be useful for calibration purposes. Another novel source, which is polychromatic in character, but is *monochromatic* or *species-specific* in behavior, is the molecular equivalent of the electrodeless discharge lamp. The source utilizes a low-pressure gas, the same as the analyte species, and the molecules of the gas are excited within a radiofrequency field. The excitation includes transitions in the vibrational energy states of the molecule, and the source emits radiation in the IR region at the characteristic, resonance wavelengths of the molecular species. In this manner, a spectrometric measurement system can be developed where the analytical wavelengths of the analyte are identical and resonantly coupled to that of the source – comparable in nature to an atomic absorption measurement. This is similar in concept to a correlation filter photometer, except that the source provides the filter function. Commercial instrumentation has been developed based on this technology, for medical applications, but it has potential for use in other applications, including the measurement of combustion gases in an industrial environment.

6.4.2.3 Single Element Detectors

Most standard infrared spectrometers currently operating in the mid-IR spectral region feature single element detectors. Exceptions are analyzers that are equipped with multiple detectors dedicated to 'single' wave-

length measurements, either within a spectrograph arrangement where the detectors are placed at the physical analytical wavelength locations, or where multiple element detectors are used with dedicated filters. The selection of detector type is dependent on the spectral region being covered [22–24]. There are two main types of detector in use for mid-IR measurements – thermal detectors and photon detectors. In the mid-IR region thermal detectors are used for most routine applications. These are preferred for most process applications because they operate at room temperature. The most common thermal detectors are the pyroelectrics, such as DTGS (deuterated triglycine sulfate) and lithium tantalate. Lithium tantalate detectors are used widely for low cost sensing devices (including fire alarms), where the detector element can cost as little as $10 (US). They are also used in filter-based instruments and in low-cost scanning instruments. MEMS-based thermopiles are also available and again these tend to be used for low cost filter-based instruments. Most process FTIR instruments feature the more expensive and more sensitive DTGS detector, which provides between one and two orders of magnitude increase in sensitivity over lithium tantalate, dependent on implementation. However, for general-purpose process analytical applications it is helpful to thermally stabilize the DTGS detector, typically at temperatures around 25–28 °C. This is an optimum temperature providing maximum stability, which is important for this particular detector, which becomes inoperable at temperatures above 41 °C. Detector packages with built-in single-stage thermo-electric coolers (TEC, also known as Peltier devices) are available for this purpose. A parameter known as D^* is often used to define detector performance, and DTGS detectors with a D^* as high as 10^8 are available. Recently, a MEMS-based pyroelectric-style detector was introduced and this is reported to have a D^* around 10^9, a factor of up to 10× improvement over traditional DTGS detectors

Thermal detectors by nature have a relatively slow response, and are not efficient for fast data acquisition rates. Also, the same applies in certain measurements that feature low energy throughput, as in some reflectance measurements, high-resolution gas analyses, and mid-IR fiber optic-based measurements or applications where a light-pipe conduit is used for interfacing an instrument to a sampling point. Traditionally, for such applications featuring low light levels or high-speed acquisition, the MCT (mercury–cadmium–telluride) detector is used. If measurements are to be made within the 4000–2000 cm^{-1} spectral region a lead selenide detector can be used. For many applications the detector can be used at room temperature. A sensitivity increase can be obtained by reducing the detector temperature with a single stage thermoelectric cooler.

The MCT is capable of providing between one and two orders of magnitude (or more) sensitivity over the pyroelectric detectors, offering D^* values in the 10^9–10^{11} range. The MCT detector, which can operate at longer wavelengths than the lead selenide detector, operates at subambient temperatures, and is usually cryogenically cooled to liquid nitrogen temperatures. When operating with liquid nitrogen cooling it is necessary to manually fill a Dewar cold-trap. These normally have an eight-hour capacity, but extended fill versions are available, providing up to 24 hours usage. Obviously, the requirement to manually fill the Dewar and the limited capacity of the Dewar pose limitations on the usage of this style of detector in a process environment. Although a commercial system exists for autofilling the Dewar, the detector system tends to be limited to laboratory-based applications. While this type of detector system has been used in some specialized industrial applications, it is mainly limited to high-resolution gas monitoring, chemical reaction studies and specialized semiconductor analyses.

Alternatives to cryogenic cooling include thermoelectric cooling (multistage Peltier devices) and thermomechanical cooling devices, such as Sterling coolers, recirculating microrefrigerators and Joule–Thompson (J–T) coolers. None of these solutions are ideal: Peltier coolers (TECs) are operated as up to three-stage devices, Sterling engine coolers have limited mechanical lifetimes and are very expensive, and J–T coolers often exhibit vibrational coupling problems. Of these, the thermoelectric coolers are often the most straightforward to implement, and can be purchased as standard detector packages. However, they do not achieve liquid nitrogen temperatures (77 K), which are necessary in order to gain optimum sensitivity. This limits the spectral range to a spectral limit of 1000–700 cm^{-1} in the mid-IR and reduces the detector's performance

closer to that of a DTGS detector. However, the high-speed data acquisition benefits of the detector are retained. Both the three-stage Peltier cooled detectors and Sterling cycle engine cooled detectors have been used for process and environmental monitoring applications with dedicated FTIR instruments, and both are commercially available from the main IR detector manufacturers

6.4.2.4 Array Detectors

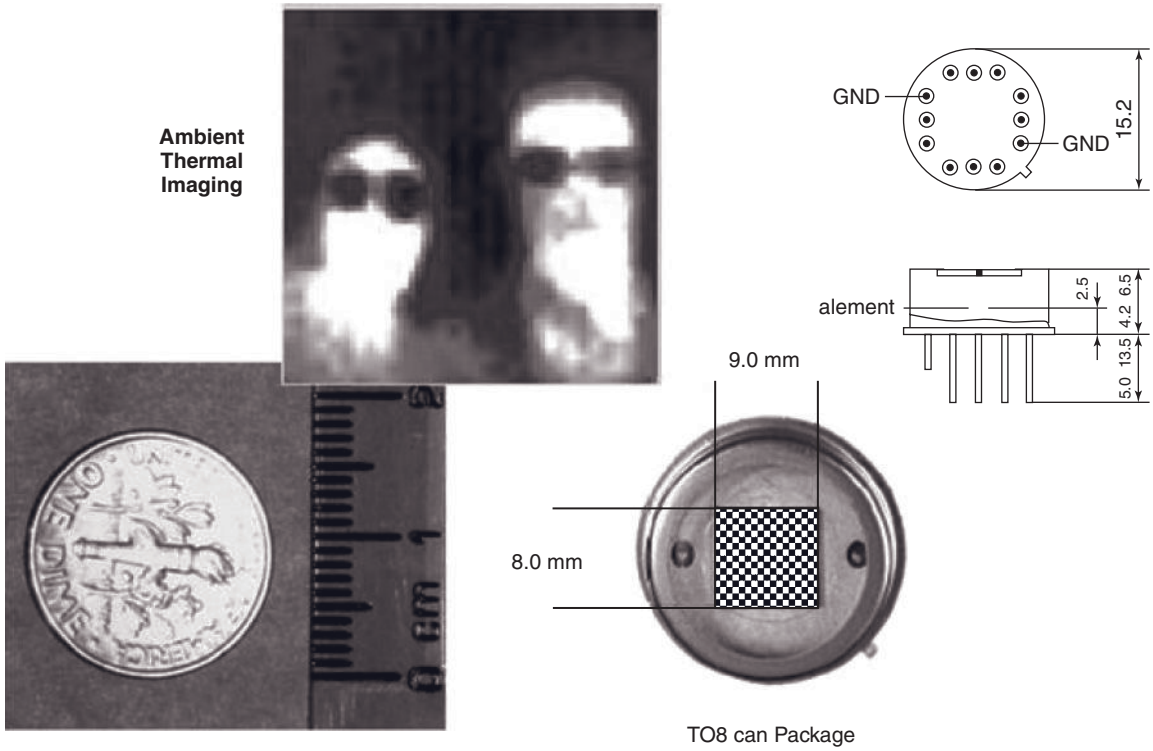
Until recently, the only detector array elements in common use were the silicon photodiode arrays, and these were initially used for UV-vis measurements, and later with extended performance into the short-wave NIR. InGaAs (indium-galium-arsenide) detector arrays operating in the mid-range NIR, which were initially developed for military imaging applications, have been used extensively in telecoms applications. Originally, these devices lacked consistency (pixel-to-pixel), they were very expensive, and were limited to 128 and 256 element arrays. Recently, arrays based on lead salts (lead sulfide and lead selenide) have also been introduced, and these have provided extended performance into the mid-IR region – down to around 5 μm (2000 cm^{-1}) – which is adequate for measuring the 'hydride' fundamental (OH, NH and CH) vibrations, for cyano compounds, and for certain combustion gases.

MCT arrays have been used by the military for night vision imaging for many years, and versions of the detector have been available for imaging-based applications (hyper-spectral imaging). These are limited to such highly specialized applications and the price is cost-prohibitive for any process-related measurements. In the past few years, other array detectors devices operating in the mid-IR spectral region have become available. These are MEMS-based structures that function as pyroelectric devices, and at least four European-based suppliers have introduced arrays featuring 128 and 256 pixels respectively. One new company based in the UK (Pyreos, Edinburgh) is offering arrays with a reported detector sensitivity with a D^* of 10^9. These arrays are being offered in evaluation packages integrated with an LVF.

6.4.3 New technologies for IR Components and instruments

There is a constant move towards the improvement of instrumentation technology. This ranges from new detector and source technologies (as indicated in the previous section), to new ways to make IR measurements and new ways to fabricate instruments. In the area of components, there are new MEMS-based components, new tunable laser devices, new tunable filter devices, both broad and narrow range, and new detector arrays. Many of these devices are currently relatively high cost. However, most are made from scalable technologies, such as semiconductor fabrication technologies, and as such if the demand increases the cost can come down. Many can follow Moore's law of scaling. In the future, for a dedicated IR sensing system, it is feasible to consider costs in the low $100s. For example, an integration of 2-D MEMs-based thermal imaging array with advanced thin film devices can yield the equivalent of a spectrometer, or a spectral detector [25] within a single detector package (Figure 6.8). The device as illustrated can be configured to provide a 120 channel output covering as most of the mid-IR spectral range (2.5–15 μm, 4000–665 cm^{-1}). An example output for an integrated vapor analyzer system is shown for the combination BTX (benzene–toluene–xylene); useful for ambient air monitoring in areas of fuels and hydrocarbon production. Note that similar miniaturized integrated spectral detectors of this type are available for both UV-visible and NIR applications.

In the area of instrument design there are examples of small interferometers being designed for portable applications. In some cases, there is a compromise in resolution, but this is more than compensated by the convenience of the smaller size. One good example which takes technology to the edge is a new MEMS-based FTIR that is in the development stages at Block MEMS (Marlboro, MA, USA). This miniature FTIR is intended for portable applications, and in its final format it will be utilized in a conceptual device known



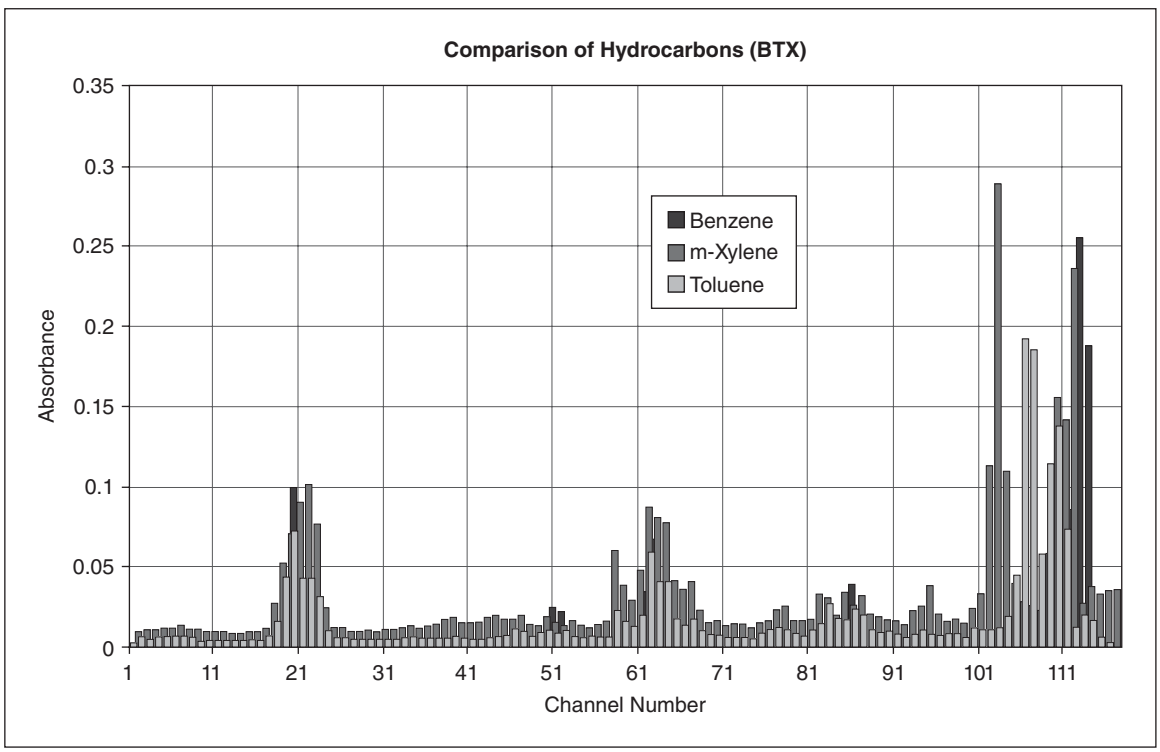


Figure 6.8 *Example of a MEMs based IR spectral detector; component dimensions and example output (120 channel output for BTX).*

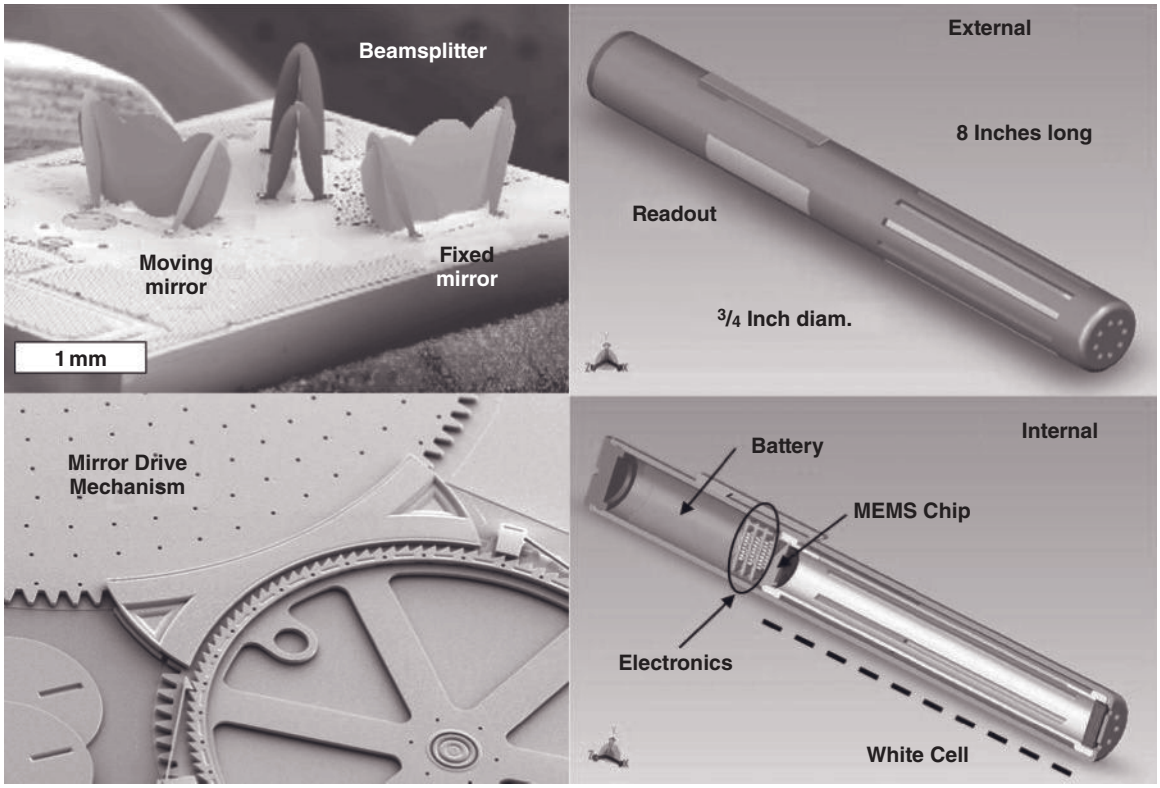


Figure 6.9 *A MEMS-based FTIR spectral engine for the ChemPen concept.*

as the ChemPen. The latter system is an integrated spectral measurement system that includes a folded path gas call to be used for chemical vapor detection (Figure 6.9). While the introduction of this device is a few years away, the basic MEMS components have been fabricated and a fully functional scanning miniature interferometer has been implemented.

6.4.4 Requirements for process infrared analyzers

The operating environment for instrumentation designed for industrial process-related measurements is understood, and guidelines (written and otherwise) exist for their implementation. While there may be compelling reasons to adapt traditional instrumentation technology used in the laboratory for process analytical applications, there are also many reasons not to do so. A common assumption is that it is a packaging issue and all that is needed for a process analyzer is to repackage the laboratory instrument in an industrial-grade enclosure. While this concept may succeed for early prototyping it seldom works for a final process analyzer. At issue there are more factors in making a process analyzer than simply the repackaging of an existing instrument. Important issues are the measurement philosophy, the overall system design philosophy and changes in the operational concepts. It is necessary to classify the mode of operation and the operating environment of the analyzer, and also it is beneficial to examine the term *process analysis*.

One important distinction is a logical separation between the terms *instrument* and *analyzer*. An instrument is totally open-ended and flexible in terms of sampling and operation, and typically generates data that

requires detailed interpretation. This is particularly the case with IR instrumentation. An analyzer is a total system, which normally includes automated sampling, or at least preconfigured sample handling, operation with predefined instrument settings, a defined user interface, and a prescribed method for the measurement and calculation of results. Process instrumentation tends to be the world of analyzers and plant engineers understand analyzer-based terminology. Process analysis can be defined as an analytical procedure that is performed to provide information on raw material quality (screening), the progress of a process – whether it is batch or continuous – and composition/quality of the final product. This can range from dedicated laboratory-based off-line testing (including portable analyzers) to continuous on-line measurements. Several clearly defined platforms of process analyzer exist and these can be delineated as follows:

6.4.4.1 Laboratory-based Analyzers

For many applications, laboratory based analyses may be developed on standard laboratory instruments. While these are not always the most ideal, they do provide a starting point, and once well defined, a custom-designed or dedicated analyzer can be produced. For infrared methods of analysis the main issues are the ease of sample handling, preferably without modification of the sample, a simple operator interface and the automated calculation and reporting of results. One of the benefits of an IR analysis in that a single instrument is multifunctional, and can be configured for a wide range of samples. As a consequence, analytical methods tend to be '*recipe-based*' and are usually developed by a support group. Most instrument companies provide macro generation software that allows the procedures to be automated with little operator input, and provide a documented audit trail in the event that a measurement is questioned. Once standardized, a dedicated analyzer can be produced just for that one measurement. This is an important requirement for regulated analytical procedures. While, there are minimum instrument performance requirements for most analyses, the most common sources of variability and nonreproducibility originate from sample handling errors. Where practical, automated methods of sample handling are preferred to eliminate or reduce this source of error.

6.4.4.2 Analyzers for Near-line or At-line Plant Measurements

Plant-based analyzers often share a common heritage with laboratory-based instruments, but their design requirements are more rigorous and demanding. Laboratory-based instruments are normally packaged in a molded plastic or sheet metal housing, with standard power and data cable connections. For operation in a plant environment, it is important to protect the instrument from the operating environment, and to ensure that the instrument conforms to factory and worker safety mandates. These can have a significant impact on the way that an instrument/analyzer is packaged and the type of electrical connections used for power and data transmission.

Most process analyzers are required to function 24 hours a day, and must work reliably for all work shifts. It is unacceptable for an analyzer to fail unexpectedly especially late at night. This requirement places a demand for a high degree of reliability, which in the end is a design issue – normally achieved by minimizing critical moving parts, and by using high quality components with long mean-time-between-failure (MTBF). This favors newer methods of measurement that have minimized or eliminated the need for critical moving parts … and as a result, the more recent FTIR systems designed for plant or mobile applications are preferred over traditional laboratory-based optical benches for such analyzers.

The requirement for simplicity of sampling and analyzer operation is even more essential within a plant environment. Sampling must be highly reproducible and require only a minimum of operator skill. One must not second-guess an operator when designing either the sample–analyzer interface or the user interface of the analyzer. Keystrokes for any operation must be kept to a minimum, and where practical, devices such

as bar code readers should be implemented. All visual displays must be unambiguous, and any cautions, warnings or error messages must be clearly displayed. Wherever possible, the local language must be represented in any displayed messages – Microsoft Windows based computer systems provide for international languages and symbolic languages.

Most analyzers incorporate a personal computer (PC) for data acquisition, control, data analysis, presentation/display of results and communications. Many are integrated via Ethernet-based networks. Two approaches are used: either the computer is integrated into the analyzer package, including the visual display and keyboard/control panel (an abbreviated keyboard or a preprogrammed equivalent), or a separate industrial-grade computer is used. In the first approach high-performance single-board computers are available based on a passive back plane mounted inside the analyzer. Alternatively, rack-mounted PCs and/or environmentally hardened PCs are used in systems requiring separate computers. The focus on computer hardware includes keyboards and input devices, which may feature integrated touch-sensitive screen displays.

If the analyzer is required to operate in a true production environment then it may be necessary to mount it within an environmental enclosure, which is sealed to provide protection from water and dust. Not all manufacturing plants offer a controlled operating environment and so the analyzer must be designed to operate over a wide range of humidity (from 0 to 100%, noncondensing) and over a wide temperature range, at minimum from around 0–40 °C, sometimes wider. If there is the risk of exposure to flammable vapors, either incidentally or continuously, then the analyzer must conform to fire safety codes. One standard approach is to place the analyzer within a purged and pressurized stainless steel enclosure. As technologies change, and as miniaturization takes place, these requirements can be reduced by paying attention to the way that the analyzer is designed, and by understanding the requirements for intrinsically safe electronics. A truly miniaturized system can be completely sealed and can operate with low voltage, often under 5 V, and with minimal current draw.

Designing a process analyzer does not stop at the enclosure, and the internal components. The electronics, the spectrometer and all associated optics, must be made to be insensitive to the operating environment. This is particularly important for infrared instruments that feature sensitive optics and also high temperature energy sources. This may be achieved by design, where all components must operate reliably independent of the operating environment of the instrument/analyzer, which can include wide variations in humidity and temperature, and the presence of vibration, over a broad range of frequencies. Even moisture-sensitive optical windows, such as KBr optics, can be adapted for use in areas of high humidity by the application of a thin film hydrophobic coating, such as Parylene.

6.4.4.2 On-line Process Analyzers

The basic requirements of an on-line analyzer are very similar to those defined above for an off-line process analyzer. The main difference is that the analyzer usually operates 24 hours a day and unattended. This places higher dependence on reliability and plant safety issues. Any maintenance must be planned in advance and is ideally carried out during a scheduled plant shutdown. Components, such as sources and lasers that are known to have a limited lifetime, must be replaced regularly during scheduled maintenance. Premature replacement reduces the likelihood of an unexpected failure. Replacement of serviceable items must be simple, and must not require any sophisticated alignment procedures. An example is the laser used in an FTIR instrument, which must not require any tedious alignment, typically associated with laser replacement in laboratory-based instruments. Also, an average electrical technician or maintenance engineer must be able to replace this component.

Most on-line analyzers are installed as permanent fixtures. Bearing this in mind, environmental issues associated with temperature and vibration become more critical, and the requirement for conformance to

safety standards may be enforced by the need for system certification. The latter requires an assessment of the working environment, and the potential for fire and/or explosion hazard, relative to the anticipated presence of flammable vapors or gases. Safety hazard and local electrical design code compliance, which includes CE Mark for Europe, Canadian Standards Association (CSA) for Canada, and Underwriter's Laboratory Association (UL) for the United States, may be required for both off-line and on-line analyzers for certain operating environments. Certification is usually conducted by recognized organizations such as Factory Mutual (FM, USA) or TÜV (Germany). Figure 6.10 is an example of a fully configured FTIR-based analyzer (the Analect HSS), equipped with onboard sample handling and onboard data processing that complies with the standard safety codes for the United States and Europe (originally CENELEC).

System installation in a permanent location may require a sample conditioning system featuring some degree of automation, such as automatic cleaning (the system illustrated above features such a system) and outlier sample collection and the need to interface to an existing control system process computer. The latter may require that the system operates with a standardized communications protocol, such as Modbus, for the chemical industry. Certain specialized industries use different protocols, such as the semiconductor industry, which uses SECS and SEC-II protocols. A 'standardized' approach designated the *Universal* Fieldbus is another method/protocol for process analyzers which is being supported by certain hardware manufacturers.

One approach where the analyzer is housed within a more benign environment is based on the use of an optical fiber-based sample interfacing. As an example the analyzer may be located within the plant control room, possibly within a standard 19-inch (48 cm) electronics rack. With this type of configuration the sample cell and sample conditioning system are placed within the hazardous location and the instrument is located in a general purpose, or a safe area, where the optical fibers are used for the analyzer–sample cell/sampling

Figure 6.10 *Fully configured and certified FTIR process analyzer system; example shown for the Analect HSS System.*

system interface [26]. In the past, optical fibers have been limited in their use in the mid-IR region because of only a small number of materials available providing IR transmission. Note that there is only limited use of optical fibers in the mid-IR region as covered by most FTIR analyzers. Currently, common choices are zirconyl fluoride glasses, which provides a spectral window from the NIR down to around 2000 cm^{-1} (5.0 μm), the chalcogenide glasses, which may be used from around 3300 to 800 cm^{-1} (3.03–12.5 μm), and silver halide, AgCl : AgBr solid solution PIR fibers that cover the range 3300–550 cm^{-1} (3.03–18 μm). Note for some applications, a practical alternative to optical fibers is light-pipe technology (optical conduit), where the sample interface is interconnected to the instrument via a series of light pipes with flexible connections. At least two manufacturers offer this approach as an option. For such installations, there are fewer requirements for the analyzer to conform to safety standards, and the rack-mounted format is quite suitable, assuming that a rack assembly is available to house the analyzer.

The functional specifications of most on-line process analyzers are similar to those used for off-line applications, and consequently, the design requirements are similar. A very practical design concept is to use common instrument measurement technology throughout, providing all three analyzer platforms as discussed above. This enables laboratory-based methods development with the ability to handle both off-line and on-line plant applications as required. A factor that is important from an implementation point of view and for on-going serviceability is the requirement to be able to support instrument calibration off-line, by providing a practical approach to calibration and method transferability between analyzers. For some applications this might make the final analyzer undesirably expensive – FTIR could be a case in point. One solution is to consider the use of simpler technology, including a filter-based instrument for the final installation, where all of the initial work is performed on a more sophisticated laboratory instrument. Once completed, the calibrations are moved over to the simpler instrument by the use of a suitable calibration transfer function.

In the design of a process spectrometer there are many attributes and design features to consider. Adapting FTIR technology to an industrial environment can provide some of the greatest challenges. Consequently, FTIR can be considered as a suitable model for the discussion of the issues involved. In this case, the main issues are safety, reliability and serviceability, vibration, temperature, optics, sources and detectors, sample interfacing and sampling, and calibration.

- *Analyzer safety:* There are several potential safety hazards, and the source is one – providing localized temperatures in the range of 1100–1500 K. The laser and its power supply, which is a potential high voltage spark hazard, are other examples. Methods for thermal isolation and instrument purging can address the source issues. The discharging of any charge within the capacitance circuitry of the laser power supply addresses the spark hazard issue.
- *Reliability and serviceability:* FTIR instruments traditionally are not known for long-term reliability, based on lab experience. However, this is changing, and vendors have focused on these issues even for laboratory instruments. Attention must be paid to optical stability, and the reliability of key components, such as optics (nonhygroscopic preferred), electronics, lasers and sources. If a component has to be replaced, then a nonspecialist technician must be able to perform the task. Again, vendors are addressing this issue with laboratory-based instruments where users can often exchange out these key components using pre-aligned optical modules.
- *Vibration:* Vibration can be a major disruption and destructive influence for any form of instrumentation. FTIR is a special case because the fundamental measurement is vibration sensitive. Care must be taken to ensure that the mirror drive, and associated optics and components are immune to vibration. Note that mirror mounts, and even detector mounts can be a source of vibrational interference. From a more basic standpoint, constant vibration can work components loose, and so the extensive use of a screw bonding adhesives, is strongly recommended.

- *Temperature:* Temperature changes can result in dimensional changes, which inevitably cause problems if not addressed, for optomechanical assemblies within an instrument. Temperature compensation is usually required, and careful attention to the expansion characteristics of the materials of construction used for critical components is essential. This includes screws and bonding materials. If correctly designed, the optical system should function at minimum over typical operating range of 0 to 40 °C. Rapid thermal transients can be more problematic, because they may result in thermal shock of critical optical components. Many electronic components can fail or become unreliable at elevated temperatures, including certain detectors, and so attention must be paid to the quality and specification of the components used.
- *Optics:* Issues relative to hygroscopic optics and the need to pay attention to mirror mounts relative to vibration and/or thermal effects have already been addressed. Zinc selenide is an important alternative material, especially when antireflection (AR) coated. If potassium bromide absolutely has to be used for its lower transmission range (down to 400 cm^{-1}) then a protective coating such as Parylene must be used. Most process analyzers use protective windows between the spectrometer and the sample interface. If used, back reflections from the window surfaces into the interferometer must be avoided because these will cause photometric errors. Such reflection can be eliminated by wedging or tilting the optical windows, relative to the beam path.
- *Sources and detectors:* Specific discussions of sources and detectors have been covered elsewhere in this article. The issues here are more service and performance related. Most sources have a finite lifetime, and are service replaceable items. They also generate heat, which must be successfully dissipated to prevent localized heating problems. Detectors are of similar concern. For most applications, where the interferometer is operated at low speeds, without any undesirable vibrational/mechanical problems, the traditional lithium tantalate or DTGS detectors are used. These pyroelectric devices operate nominally at room temperature and do not require supplemental cooling to function, and are linear over three or four decades.
    - Certain applications requiring higher spectrum acquisition rates or higher sensitivity may require a higher performance detector, such as a MCT detector or the lead selenide detector (5000–2000 cm^{-1}). The MCT detector is capable of high-speed acquisition and has a very high sensitivity; however, the main issues here are practical in terms of implementation. The detector ideally must be cooled down to liquid nitrogen temperatures (77 K), and the detector exhibits nonlinearity with absorbances greater than one (unless electronically linearized). MCT detector packages with single or multistage TE coolers and MCT detectors integrated with Sterling cycles coolers are available if the use of an MCT detector is required. The TE cooled devices tend to have a limited spectral range, dependent on the number of cooling stages used – more stages the greater the range. The Sterling cycle coolers are expensive and have a limited lifetime and so their use may be limited to high value-added applications.
    - The DTGS detector does not require cooling to operate; however, it will stop functioning as a detector, at temperatures above 41 °C, because of depolarization. Note that internal instrument temperatures can be up to 10 degrees higher than the external ambient temperature, and so there is a need for concern about the DTGS detector. A solution is to thermally stabilize the DTGS detector in a process analyzer, typically with a one-stage Peltier TE cooling device.
- *Transfer optics:* The interface of the IR beam with the sample, especially in a classified hazardous environment, can be a major challenge. Gas samples are not too difficult although it is important to pay attention to the corrosivity of the gases, relative to the windows and any internal optics, as with a folded path length cell. Liquids offer different challenges. For on-line applications users prefer to minimize the use of valves, bypass streams and auxiliary pumps, especially over long distances between the stream and the analyzer. At times there is a benefit to sample the stream either directly or as close as possible

to the stream, and in such cases there is the need for transfer optics between the sample cell and the spectrometer/analyzer.

  - There are a couple of quite elaborate schemes for 'piping' the light from the spectrometer to the sample and back. While these do work, they are typically cumbersome, and they can be awkward to implement, and maintaining alignment and optical throughput can be a problem. The use of fiber optics has already been mentioned, and it is worthwhile to consider their use in context; there are practical limitations as previously mentioned. Recent work has involved hollow core fiber light guides, which may provide a better transmission range than IR transmissive fiber-optic materials, and with much lower attenuation losses. In this implementation, the hollow fiber guides have much of the flexibility of fiber optics, but with most of the transmission characteristics of a light pipe interface.

- *Diagnostics:* There are many sources of error that can develop within a process analyzer. Component failure is usually relatively easy to diagnose (as long as it is not the computer) and so on-line diagnostics can be straightforward. More insidious problems are caused by the analysis itself, or by electronic component aging (does not fail but operates out of specification), or when a sampling device gives problems, such as corrosion or the coating of windows of a cell. Many of these problems are application dependent, but often the failure mode is anticipated, and preventative measures implemented to reduce the impact. Note that the FTIR instrument is a full-spectrum device, providing a significant amount of spectral over-sampling, and as such is very beneficial for application diagnostics. For example, regions of the spectrum may be used to detect the presence and identity of contaminants or unexpected material decomposition. This is a unique advantage offered by the IR spectroscopy in process applications.
- *Calibration:* Most process analyzers are designed to monitor concentration and/or composition. This requires a calibration of the analyzer with a set of prepared standards or from well-characterized reference materials. The simple approach must always be adopted first. For relatively 'simple' systems the standard approach is to use a simple linear relationship between the instrument response and the analyte/standard concentration [27]. In more complex chemical systems, it is necessary to adopt either a matrix approach to the calibration (still relying on the linearity of the Beer–Lambert law) using simple regression techniques, or to model the concentration and/or composition with one or more multivariate methods, an approach known as chemometrics [28–30].
- *Standardization:* The instrument response function can vary from analyzer to analyzer. If calibration transfer is to be achieved across all instrument platforms it is important that the instrument function is characterized, and preferably standardized [31]. Also, at times it is necessary to perform a local calibration while the analyzer is still on-line. In order to handle this, it is beneficial to consider an on-board calibration/standardization, integrated into the sample conditioning system. Most commercial NIR analyzers require some form of standardization and calibration transfer. Similarly, modern FTIR systems include some form of instrument standardization, usually based on an internal calibrant. This attribute is becoming an important feature for regulatory controlled analyses, where a proper audit trail has to be established, including instrument calibration.

Overall, most of the requirements for a process spectrometer/analyzer are straightforward to implement, but they do require attention at the design level. Another important area, which is FTIR specific, is the user interface and the need to provide for industry standard data communications. Standard software packages do exist for process instrumentation. For prototype development, and even for the front-end interface in a stand-alone mode of operation, software products, such as National Instruments' LabView and the Mathworks MatLab, are also important instrumentation development tools. Note that National Instruments also provides important computer-based electronics and hardware that meet most of the computer interfacing, and system control and communications needs for modern instrumentation. For practical installations, a product known

as Symbion, Developed by Axiom Analytical (Tustin, CA) operates as a universal software interface for all forms of spectroscopic instrumentation. The company also offers versions of the software optimized for PAT applications.

The design issues discussed so far are based on the requirements of process analyzers operating in a traditional industrial environment that are governed by the need to comply with established electrical, fire and safety standards. For analyzers operating in the high-technology fabrication industries (semiconductors, optics, vacuum deposition, composites, etc.), consumer-associated industries, such as the automotive after market, and environmental, safety and hygiene applications (United States Environmental Protection Agency, United States Occupational Safety and Health Administration, etc.), the needs and requirements are often very different. Instrument reliability and performance are always important, and operator safety requirements are much the same in all industries, although the actual standards may differ, dependent on the regulatory agencies involved. For operation in some specialized areas of application, the need to follow well-defined operating procedures and operating protocols is essential, and full documentation and audit trails are a requirement. For some applications, ISO 9000 certification of the analyzer company may be a prerequisite.

6.4.5 Sample handling for IR process analyzers

Many references have been made regarding the importance of the sample, how it is handled and the manner in which it is presented to the IR analyzer. Sample handling for IR spectroscopy has become a major topic of discussion for more than 30 years, and is well documented in standard texts [32–38]. IR spectroscopy, when compared to other analytical techniques, tends to be unique relative to the large number of sample handling technologies that are available. This is in part due to the fact that samples can be examined in any physical state. Selection of the most appropriate sampling technique is extremely important relative to the overall success of a process-related application. Although the modes of operation may be very different, this applies equally to both off-line and on-line styles of analyzers. The only difference for on-line applications is that the sample handling method must be sufficiently robust to ensure optimum performance during 24-hour operation. Attention to sample corrosion, and pressure and temperature effects is essential. This often requires the special attention to the materials of construction used for cell fabrication. The use of specialized and even exotic materials may be required for both the optical elements and the cell construction. Traditional IR optics (Table 6.2) are generally unsuitable, and optical materials such as sapphire, cubic zirconia, and even diamond are required for demanding applications; in these cases the material selection is based on the optical transmission range that is required for the measurement (Table 6.3). Likewise,

Table 6.2 *Common IR transmitting window materials*

Material	Useful range (cm^{-1}) (transmission)	Refractive index at 1000 cm^{-1}	Water solubility (g/100 ml, H_2O)
Sodium chloride, NaCl	40 000–590	1.49	35.7
Potassium bromide, KBr	40 000–340	1.52	65.2
Cesium iodide, CsI	40 000–200	1.74	88.4
Calcium fluoride, CaF_2	50 000–1140	1.39	Insoluble
Barium fluoride, BaF_2	50 000–840	1.42	Insoluble
Silver bromide, AgBr	20 000–300	2.2	Insoluble
Zinc sulfide, ZnS	17 000–833	2.2	Insoluble
Zinc selenide, ZnSe	20 000–500	2.4	Insoluble

Table 6.3 *Common IR transmitting materials used for IR optics and ATR Internal reflectance elements (IREs)*

Material	Useful range (cm^{-1}) (transmission[a])	Refractive index at 1000 cm^{-1}	Water solubility (g/100 ml, H_2O)
Zinc sulfide, ZnS	17 000–833	2.2	Insoluble
Zinc selenide, ZnSe	20 000–550	2.4	Insoluble
Cadmium telluride, CdTe	20 000–320	2.67	Insoluble
AMTIR[b]	11 000–625	2.5	Insoluble
KRS-5[c]	20 000–250	2.37	0.05
Germanium, Ge	5500–600	4.0	Insoluble
Silicon, Si	8300–660	3.4	Insoluble
Cubic zirconia, ZrO_2	25 000–~1600	2.15	Insoluble
Diamond, C	45 000–2500, 1650–<200	2.4	Insoluble
Sapphire	55 000–~1800	1.74	Insoluble

Materials such as cubic zirconia, diamond and sapphire may be used for transmission windows for special applications.
[a] The usable range is normally less than the transmission range because of the higher optical attenuation that results from the extended optical path length of an IRE.
[b] AMTIR: Infrared glass made from germanium, arsenic and selenium – also known as chalcogenide glass.
[c] Eutectic mixture of thallium iodide/bromide.

materials such as nickel, Monel and Hastelloy are used for cell construction when corrosive materials are analyzed.

One of the most important factors in the selection of the sample handling technique is to attempt to analyze the sample, as it exists, without any form of chemical or physical modification. For gases and certain liquids, simple transmission cells, often with a flow-through configuration meet these requirements.

However, many liquid samples have a high IR absorption cross-section, and strong IR absorption occurs even for short path lengths of the material. This situation causes problems for mid-IR measurements in a process application. In the laboratory, it is sometimes possible to reduce the sample path length for such samples down to a capillary film. This is impractical for monitoring applications that would require a flow-through configuration. This problem does not apply in the NIR, where sample path lengths are typically one or two orders of magnitude larger (typically between 1 and 100 mm) than in the mid-IR spectral region.

A solution to this path length issue is to use the technique known as ATR or simply, just internal reflectance. During the 1980s and 1990s there was a focus on the ATR for both off-line and on-line sampling [35–38]. There are several different configurations available that provide for manual sample loading, continuous flow methods, and dip and insertion probes. The technique is based upon the principle of internal reflectance within a transparent medium of high refractive index. The medium is cut or shaped to provide the equivalent of facets, or reflective surfaces, and the final component is called an internal reflectance element or IRE. As the infrared radiation reflects from the internal surfaces, there is an interaction at the surface with any material (the sample) in intimate contact with that surface. The radiation effectively penetrates in the region of approximately 1 and 3 μm into the surface of the contact material (the sample). Within this penetration depth the equivalent of absorption occurs at the characteristic frequencies of the sample, and as a consequence a spectrum of the sample is generated. The actual final 'effective path length' is the accumulated sum of the absorptions as a function on the number of internal reflections. There is a range of materials available that are suitable for the construction of the IRE, and the final selection is based on the index of refraction of the IRE material, the effective range of optical transparency, and the surface strength and inertness of the IRE, relative to the sample. A selection of suitable materials is included in Table 6.3,

and this includes diamond, which along with sapphire and cubic zirconia, is particularly applicable to process monitoring applications.

ATR offers many benefits for the measurement of samples in the mid-IR because it can handle a wide range of materials and sample characteristics. It can be applied to liquids, powders, slurries, foams and solids. It is particularly well suited to liquids because a high degree of intimate contact with the surface can be maintained with a liquid. An issue that one needs to be aware of is the potential for surface contamination and fouling, especially for on-line, flow-through applications. Fouling of the surface not only causes interferences and/or potential energy losses, but it can also change the photometry of the experiment by modifying the refractive index of the surface. The propensity for fouling is application dependent, and is also a function of the surface characteristics of the IRE and the chemical nature of the sample. A material such as zinc selenide has polar sites at the surface (it is effectively a salt, and zinc and selenide ions can exist on the surface) and ionic interactions with certain sample types can occur. In such cases, it is important to try to implement algorithms for detecting material build-up on the surface. Cleaning schemes can be incorporated, and by definition, this requires some form of automated sampling system. For some applications, a simple flushing or back flushing with a cleaning solvent is all that is required. Certain commercial analyzer systems incorporate an ultrasonic cleaning option into the sampling area for cases where there is an incidence of strong material adhesion to the surface. Another issue relative to surface contamination effects on zinc selenide is that care must be exercised in the event that the stream contains an oxidizing agent. In the presence of an oxidant the surface layers of zinc selenide can be converted to selenium dioxide, which may show up as a darkening or a brown-to-purplish stain. This material thickness can be very small, of the order of molecular layer thickness, but dependent on the exposure and the amount of degradation it might cause a lowering of the apparent optical throughput. If in doubt about the choice of material when oxidizing agents are present in the stream, the use of a diamond IRE may be a safe choice.

It is a general perception that sample handling is easier in the NIR region compared to the mid-IR. To some extent this is true, because longer path lengths may be used for liquid sample handling, and direct reflectance methods, such as diffuse reflectance, may be used for convenient sampling of solids. Diffuse reflectance is used for mid-IR applications in the laboratory, but in general it is difficult to implement for on-line applications. Certain applications have been documented where mid-IR diffuse reflectance has been applied to mineral characterization, and this has been used successfully as a field-based application. In practice, spectral artifacts generated by anomalous surface reflections tend to limit the use of reflection techniques for solids in process applications.

6.4.6 Issues for consideration in the implementation of process IR

6.4.6.1 Low Maintenance and Turnkey Systems

The average process engineer is extremely conservative and is hard to convince when it comes to the implementation of process analyzers – and process IR is no exception. Unless there is an agreement to use a process analyzer in an experimental role, it is normally a requirement to supply the analyzer system as a turnkey product. Also, the product is expected to operate 'out-of-the-box', with minimum technical support required at the time of installation-very much like installing a modern refrigerator in the home. Once installed, the analyzer system must operate with minimum operator attention. Maintenance, as noted earlier, must fit into a scheduled maintenance program. Failure modes must be understood, and the recommended time frame for replacement parts must be determined, with an adequate supply of spare parts. The process engineer is responsible for down time and lost production. An analyzer company must not be perceived as the cause of such problems. With an FTIR system, components such as sources and lasers are often the most likely to fail and understanding of the MTBF for these is important.

An on-board embedded computer is welcomed in a process analyzer; however, fragile components such as integrated hard disk drives are not recommended. The best approach for on-board storage is solid-state storage, such as flash random access memory (RAM) or battery-backed RAM. Note that while the use of flash memory is now widespread and solid state drives are readily available, it is important to note that the number of reads and writes to a flash device can be limited. In a continuous process this might be important because of the high frequency of data collection. Moving parts in analyzers are always subject to critical assessment, and technologies that minimize the need for moving parts are preferred. If FTIR is the only option for a process analyzer, then it is important to re-enforce design features and attributes that address system stability and reliability. There is a trend towards miniaturization with no moving parts. It is early days yet for some of these technologies, but with time they are expected to provide a practical alternative to FTIR.

Maintenance issues are important to address, and the requirement for high cost replacement parts must be kept to a minimum, unless the frequency of replacement is once every three to four years. Today, process FTIR analyzers are not necessarily a problem, as long as components are simple to replace. The normal replacement items are the source and laser, and the replacement of these must be kept simple. If an interferometer alignment is required, then this must be automated via an integrated mechanized auto-align procedure.

6.4.6.2 Cost per Analyzer Point

One aspect that has limited the use and introduction of process analyzers, and in particular FTIR analyzers, has been the high cost of the analyzer, and the resultant high cost per analysis point. Many process analyzer installations are financially evaluated in terms of cost per analysis or analysis point, as well as cost of ownership and the rate of payback. Typically, these costs need to be kept as low as possible

The use of fiber optics and fiber-optic multiplexing can increase the number of analysis points, and hence can reduce the overall costs related to a single analyzer. This approach has been used successfully with NIR instrumentation, where typically up to eight points can be handled. As noted earlier, the use of fiber optics with IR Fourier transform instruments has in the past been limited. New fiber materials with improved optical throughput are available, and also with the considered use of IR lasers, the role of fibers for IR applications is expected to increase. Although in the past commercial multiplexers have been available for mid-IR fiber systems, their use has not been widespread.

The opportunity to reduce the cost per analysis point can be very important for certain applications. One option with FTIR is to use sample stream multiplexing with a manifold system. For gas-based systems this is very straightforward, and does not impose a high cost overhead. Liquid systems are more complex, and are dependent on the nature of the stream involved, and material reactivity, miscibility and viscosity are important factors to consider. As noted previously, with the introduction of new, miniaturized technologies, the hope is that newer, less expensive devices can be produced, and these can provide the needed multiplicity to reduce the cost per analysis point to an acceptable level.

6.4.6.3 Overall Cost Considerations

The cost per analysis point and the projected payback are only two metrics used to evaluate the feasibility of installing a process analyzer. If a spectrometer replaces an existing service intensive analyzer then that is a positive situation. Also, because of the flexibility of infrared analyses, it is feasible that a mid-IR analyzer might replace several existing analyzers; it is a case where direct chemical information from IR can be more important than certain inferential methods that only provide indirect information. Also, if the only other

option is lengthy off-line laboratory analyses, then it is easier to justify installing a process IR analyzer. However, if it is necessary to maintain complex chemometric methods, as with some of the NIR applications, then that can be a large negative factor. Cost of ownership is an important metric and is often estimated over 5- or 10-year time frames. Analyzers operating in the mid-IR, or even in the first overtone (FT-NIR) are typically much easier to calibrate than traditional NIR analyzers because individual absorption bands can be monitored, often free from interferences. In such cases a chemometrics-based calibration is unnecessary in the mid-IR. Consequently, cost benefits associated with mid-IR analyzers may be a positive selling point compared to an NIR solution.

6.5 Applications of Process IR Analyzers

Up to this point IR spectroscopy has been discussed relative to the requirements of a process analyzer, and what is needed to implement IR technologies. It is also important to evaluate the IR instrumentation relative to specific applications. IR provides important information that relates directly to the molecular structure and composition, of both organic and inorganic materials. Virtually all classes of chemical compounds have functional groups that provide characteristic absorptions throughout the classic mid-IR spectral region. Similarly, many parallel applications exist for inorganic compounds, which typically also provide very characteristic and often very intense IR signatures. A detailed discussion of potential and actual applications, both industrial and otherwise, is too long to cover in any detail here. Table 6.4 indicates strong candidate applications of mid-IR that can be extended to process, process-related or QC measurements.

Table 6.4 *Example applications for process IR spectroscopy*

Application area	Materials and products	Techniques and applications
Refinery production	Distillates, crude oils, reforming and cat cracking, and blending	Liquid and gas – transmission Composition monitoring
Fuels	Natural gas, LPG, propane/butane, distillates, gasoline, diesel, fuel oil, syn-fuels, additives	Liquid and gas – transmission Composition monitoring, Octane and cetane number measurements
Solvents	High purity products, mixtures	Composition monitoring Vapor analysis
Oils and lubricants	Mineral and synthetic oils, greases, coolants, hydraulic fluids	Blending analysis Additive chemistry Oil degradation monitoring
Specialty gas products	High purity gases, gas mixtures	Composition monitoring Impurity monitoring
General chemicals	Organic/inorganic, liquids and solids	Composition monitoring Impurity monitoring
Chemical reaction monitoring	Polymerizations, esterifications and other condensation reactions, diazo reactions, oxidation and reduction	On-line and dip-probe applications

(continued overleaf)

Table 6.4 *(continued)*

Application area	Materials and products	Techniques and applications
Plastics and polymers	Monomers, homopolymers and copolymers, and formulated products	Composition monitoring and physical properties, such as crystallinity, conformation, monomer units, and end groups
Polymer products	Adhesives, adhesive tapes, sealants, latex emulsions, rubber materials, plastic fabrication, etc.	Composition monitoring Rate of cure monitoring Product QC
Consumer products	Liquids detergents and soaps Personal care products Cosmetics Polishes – waxes, etc.	Blending control Composition monitoring Raw materials screening Quality control
Specialized products	Water treatment products Dyes and pigments Textiles Pulp and paper Agrochemicals	Production control Composition monitoring Raw materials screening Quality control
Environmental	Solid and liquid waste Drinking and waste water Priority pollutants	Trace analysis Vapor analysis
Combustion gases	Regulatory applications, Fire hazard assessment	Composition monitoring – gas Trace gas and vapor analysis
Ambient air monitoring	Workplace, ventilation systems	Composition monitoring – gas Trace gas and vapor analysis
Food products	Dairy products, General food additives Natural oils and fats Beverages Flavors and fragrances Fermentation reactions	Blending control Composition monitoring Raw materials screening Quality control
Pharmaceutical products and development	Prescription drugs Dosage forms Active ingredients Product synthesis	Reaction monitoring Dryer monitoring – vapors Blending control Composition monitoring Raw materials screening Quality control
Aerosol products	Packaging, propellants, final product assay	Raw materials screening Quality control Gas mixture analysis
Packaging	Paper, cardboard, plastics, films, adhesives	Raw materials screening Physical properties, such as crystallinity – polymers Quality control
Refrigeration	Automotive and industrial, monitoring of refrigerants	Gas composition analysis
Semiconductors	Process gases, plasma gases, substrate analysis for contaminants	Gas composition analysis Raw materials screening Trace analysis Quality control

6.6 Process IR Analyzers: a Review

Process IR spectrometers can be used throughout a large number of industry segments, ranging from their use in traditional chemicals and petrochemicals manufacturing to fine chemicals and specialized manufacturing. The latter includes semiconductor, pharmaceutical and consumer-oriented products, such as packaged goods and food products. A potential end-user has to balance the pros and cons of mid-IR versus NIR. In recent years NIR has become accepted in many industry segments as a near-line/at-line tool, and as an on-line continuous monitoring technique. However, there are areas not well covered by NIR, such as gas-phase monitoring applications, especially for the detection of low concentrations of materials. For these, mid-IR spectroscopy, used in the form of either traditional analyzers, or as FTIR instrumentation, have become the techniques of choice.

Gas and vapor applications go beyond product manufacture, and include industrial health and safety, hazard monitoring, and, as appropriate, the needs of various government agencies, including those that focus on the environment (United States Environmental Protection Agency), energy (Department of Energy) and homeland security (Department of Defense). In some cases the monitoring devices are not required by the agency directly, but are used as tools of compliance for industry as a whole.

The petrochemical and chemical industries are the traditional industries for IR analyzers. The needs are extremely diverse and the versatility associated with process IR spectral measurements make the analyzers a good fit. Products manufactured by the petrochemical and chemical industries range from commodity materials (basic chemicals and raw materials) to fully formulated or compounded end-use products. Value-added products (formulated products) often require an advanced spectral analyzer, such as a process FTIR, whereas commodity chemicals are often served by a traditional IR photometer.

It has always been assumed that manufacturing companies want to make plants more efficient and wish to maximize profitability during production. While this is true, it is not an automatic assumption that a company will install expensive equipment to make manufacturing more efficient. In general, IR analyzers are viewed as cost-savers, providing information rapidly, often capable of providing it in 'real-time', and with a minimal amount of operator intervention. In many cases, mid-IR analyzers can be easier to justify than other forms of process analyzers. Adding new technology might be just a small part of an overall modernization scheme – it depends on the industry, and it depends on the company. Note that there is a requirement for a return on investment (ROI) and this can vary significantly from industry-to-industry: it ranges from a payback within 6 months to 4 or 5 years.

Beyond the ROI issues, one of the premises for the use of advanced process analyzers is improved product quality. For many industries, and in particular industries close to consumers, there is the need to define procedures and installation qualification/ operational qualification/performance qualification (IQ/OQ/PQ) issues, good laboratory practice and good manufacturing practice (GLP/GMP), and audit trails, etc. Such requirements open the door for the use of improved analytical methodology, which can be readily provided by the implementation of good process analyzer technology. It is worth noting that many instrument vendors now address these issues directly within the instruments and the operating software of the instrument.

At times there has been a focus on a specific application, and instruments have been developed based on either a new or an existing technology. While such devices may appear to be limited to the application at hand, with some modification, either to the optics or the sampling they can be applied more generally. Many instrument developments are either sponsored by the Federal Government (for military, aerospace, energy, and environmental applications) or are spawned from projects undertaken by engineering consulting firms for major manufacturing companies. One example is a refrigerant analyzer that was developed for the automotive industry for the determination of Freon/halocarbon type in vehicle air conditioning systems – a requirement initiated from an EPA mandate. In this particular case, specific filters were used to define the

analytes, and the packaging was small and inexpensive. This type of system can lend itself to many process applications with only minor adaptations. The filter wheel shown in Figure 6.6C is an example of a component that can be customized for such applications. Examples such as this are applications driven, and the products are often developed quite independently of the traditional process analyzer business. Today, we are seeing a similar phenomenon with the telecommunications industry, where IR and NIR spectral devices were designed to monitor laser throughput in fiber optics networks. With some adaptation, many of these devices can be expanded to meet industrial and environmental applications. The benefits to such technologies are; small size, low cost potential, environmental ruggedness and high reliability (both required for the telecommunications industry).

6.7 Trends and Directions

Over time, a large number of traditional laboratory instruments have been morphed to meet industrial needs for QC applications. Example applications include raw material, product quality control and also some environmental testing. In such scenarios laboratory instruments appear to work adequately. Having said that, there are issues to be addressed: the need for immediate feedback and the need for smaller, cheaper and more portable measurements. There is a growing interest in the ability to make measurements in almost any area of a process, with the idea that better production control can lead to a better control of the process and of the quality of the final product. Also, there is growing concern about the operating environment, worker safety and environmental controls. In the past few years there has been a notable increase in the number of handheld and portable FTIR instruments (as well as handheld NIR and Raman instruments). Many of these instruments were spawned from the requirement to have analytical-grade technologies available on-site for public safety testing in cases of chemical spills (Hazmat) and also chemical agent detection for homeland security and defense (DHS and DoD) applications. The availability of such instrumentation is expected to increase to the point where the need for laboratory instrumentation may be significantly reduced. Sending the sample back to the lab for testing will become a phrase from the past.

The ideal scenario would be to have the power of a traditional IR analyzer but with the cost and simplicity of a simple filter device, or even better to reduce the size down to that of a sensor (such as the spectral detector mentioned earlier) or a simple handheld device. This is not far-fetched, and with technologies emerging from the telecommunications industry, the life science industry and even nanotechnology, there can be a transition into analyzer opportunities for the future. There is definitely room for a paradigm shift, with the understanding that if an analyzer becomes simpler and less expensive to implement then the role of analyzers/sensor can expand dramatically. With part of this comes the phrase ‘good enough is OK’ – there is no need for the ultimate in versatility or sophistication. Bottom line is that even in process instrumentation, simple is beautiful.

Expanding on this theme, spectrometer devices with lower resolution, shorter spectral ranges, and with multiple wavelength selection can be customized to a wide range of analyses. Devices already exist, sometimes implemented in different arenas, in particular for aerospace, military and defense applications, and in certain medical applications. Technologies outlined in this chapter can be modified and changed with this understanding. Specialized enabling technologies exist, such as thin-film optical coatings, chip-scale devices, microengineering and micromachining, and alternative light transmission and detection technologies. Most come from outside of the traditional instrument industry (as already suggested), and in the future, new generations of instruments and spectral sensing devices will be introduced from companies other than the established instrument vendors.

References

1. D.A. Burns, E.W. Ciurczak (eds), *Handbook of Near-infrared Analysis*, 3rd edition, Practical Spectroscopy Series, Marcel Dekker, New York, 2007.
2. G. Dent, J.M. Chalmers, *Industrial Analysis with Vibrational Spectroscopy*, Royal Society of Chemistry, Cambridge, 1997.
3. J.M. Chalmers, P.R. Griffiths, *Handbook of Vibrational Spectroscopy*, 5 volumes, John Wiley & Sons, Chichester UK, 2002.
4. J.P. Coates, 'Vibrational Spectroscopy: Instrumentation for Infrared and Raman Spectroscopy', *Appl. Spectrosc. Rev.*, **33**(4), 267–425 (1998).
5. P. Fredericks, L. Rintoul, J.P. Coates, 'Vibrational Spectroscopy; Instrumentation for Infrared and Raman Spectroscopy,' in *Ewing's Analytical Instrumentation Handbook*, 3rd edition, Ed. Jack Cazes, Marcel Decker, New York, NY, pp. 163–238, 2005.
6. J.P. Coates, The many faces of vibrational spectroscopy in the 21st century, *American Laboratory*, **1**(1), 2–10 (2008), available on-line from www.iscpubs.com
7. J.J. Workman, Jr, D.J. Veltkamp, S. Doherty, *et al*., Process analytical chemistry, *Anal. Chem.*, **71**, 121R–180R (1999).
8. J.J. Workman, Jr, 'Review of Process and Non-invasive Near-infrared and Infrared Spectroscopy', *Appl. Spectrosc. Rev.*, **34**(1&2), 1–89 (1999).
9. J.P. Coates, P.H. Shelley, 'Infrared Spectroscopy in Process Analysis,' in *Encyclopedia of Analytical Chemistry*, Ed. R.A. Meyers. John Wiley & Sons, Ltd., Chichester UK, pp. 8217–8240, 2000.
10. N.B. Colthrup, L.H. Daly, S.E. Wiberley, *Introduction to Infrared and Raman Spectroscopy*, Academic Press, San Diego, 1990.
11. D. Lin-Vien, N.B. Colthrup, W.G. Fateley, J.G. Grasselli, *Infrared and Raman Characteristic Group Frequencies of Organic Molecules*, Academic Press, San Diego, 1991.
12. J.P. Coates, A practical approach to the interpretation of infrared spectra, in *Encyclopedia of Analytical Chemistry*, R.A. Meyers (ed.), John Wiley & Sons, Ltd., Chichester UK, pp. 10815–10837, 2000.
13. J.J. Workman, Jr, Interpretive spectroscopy for near infrared, *Appl. Spectrosc. Rev.*, **31**(3), 251–320 (1996).
14. R.F. Curl and F.K. Tittel, Tunable infrared laser spectroscopy, *Annu. Rep. Prog. Chem.*, Sect. C, **98**, 1–56 (2002).
15. R. Maulini, M. Beck, J. Faist, E. Gini, Broadband tuning of external cavity bound-to-continuum quantum-cascade lasers. *Appl. Phys. Lett.*, **84** (10): 1659 (2004).
16. I. Howieson, E. Normand, M.T. McCulloch, Quantum-cascade lasers smell success, *Laser Focus World*, **41**(3): S3 (2005).
17. Editorial: Tunable QC laser opens up mid-IR sensing applications, *Laser Focus World*, PennWell Publications (2006).
18. E. Normand, I. Howieson, M.T. McCulloch, Quantum-cascade lasers enable gas-sensing technology, *Laser Focus World*, **43**(4): 90–92 (2007).
19. P.A. Wilks, High performance infrared filtometers for continuous liquid and gas analysis, *Process Control Qual.*, **3**, 283–293 (1992).
20. P.R. Griffiths, J.A. de Haseth, *Fourier Transform Infrared Spectrometry*, Wiley-Interscience, New York, 2008.
21. J.R. Ferraro, K. Krisnan (eds.), *Practical Fourier Transform Infrared Spectroscopy: Industrial and Chemical Analysis*, Academic Press, San Diego, 1989.
22. R.D. Hudson, Jr, Infrared system engineering, in *Pure and Applied Optics*, S.S. Ballard (ed.), Wiley-Interscience, New York, 1969.
23. W.L. Wolfe, G.J. Zissis, *The Infrared Handbook*, revised edition, The Infrared Information and Analysis (IRIA) Center, Environmental Research Institute of Michigan, Office of Naval Research, Dept. of the Navy, Washington DC, 1985.
24. E.J. Lerner, Infrared detectors offer high sensitivity, *Laser Focus World*, **32**(6), 155–164 (1996).
25. J.P. Coates, Spectrometric and photometric detectors: opening the doors to miniaturized spectroscopy, *Lab Int.*, **September**, 25–27 (2008).
26. A. Ganz, J.P. Coates, Optical fibers for on-line spectroscopy, *Spectroscopy*, **11**(1), 32–38 (1996).
27. J.P. Coates, Classical methods of quantitative analysis, in *Handbook of Vibrational Spectroscopy*, Volume 3, P.R. Griffiths and J. Chalmers (eds), John Wiley & Sons, Ltd., Chichester UK, pp. 2235–2257, 2002.

28. J.J. Workman, Jr, P.R. Mobley, B.R. Kowalski, R. Bro, Review of chemometrics applied to spectroscopy: 1985–1995, Part 1, *Appl. Spectrosc. Rev.*, **31**(1&2), 73–124 (1996).
29. The American Society for Testing and Materials (ASTM), Standard practice for infrared, multivariate quantitative analysis, Practice E1655–94 (1995). ASTM Annual Book of Standards, West Conshohocken, 755–779, Vol. 03.06, 1996.
30. K.R. Beebe, R.J. Pell, M.-B. Seasholtz, *Chemometrics: Practical Guide*, John Wiley & Sons, Ltd, New York, 1998.
31. J.J. Workman, Jr, J.P. Coates, Multivariate calibration transfer: the importance of standardizing instrumentation, *Spectroscopy*, **8**(9), 36–42 (1993).
32. T.J. Porro, S.C. Pattacini, Sample handling for mid-infrared spectroscopy, Part I: solid and liquid sampling, *Spectroscopy*, **8**(7), 40–47 (1993).
33. T.J. Porro, S.C. Pattacini, Sample handling for mid infrared spectroscopy, Part II: specialized techniques, *Spectroscopy*, **8**(8), 39–44 (1993).
34. P.B. Coleman (ed.), *Practical Sampling Techniques Techniques for Infrared Analysis*, CRC Press, Boca Raton, 1993.
35. J.P. Coates, A review of sampling methods for infrared spectroscopy, in *Applied Spectroscopy, A Compact Reference for Practitioners*, J. Workman and A. Springsteen (eds), Academic Press, pp. 49–91, 1998.
36. N.J. Harrick, *Internal Reflection Spectroscopy*, Harrick Scientific Corporation, Ossining, New York, 1987. (Original Publication by John Wiley & Sons, Ltd, New York, 1967.)
37. F.M. Mirabella, Jr (ed.), *Internal Reflection Spectroscopy: Theory and Applications*, Marcel Dekker, New York, Vol. **15**, 1993.
38. F.M. Mirabella, Attenuated total reflection spectroscopy, in *Modern Techniques in Applied Molecular Spectroscopy*, F.M. Mirabella, Ed., Wiley Scientific, New York, 1998, Chap. 4, pp. 127–184.

7

Raman Spectroscopy

Nancy L. Jestel

SABIC Innovative Plastics, Advanced Analytical Chemist, Global Spectroscopy Functional Leader, New R&E, Selkirk, NY, USA

7.1 Attractive Features of Raman Spectroscopy

Raman spectroscopy is particularly well suited for use in process monitoring and control. This chapter discusses Raman spectroscopy's attractive features as well as alerts the reader to aspects that may present challenges. The fundamental principles of the technique are reviewed. The reader will learn about instrumentation and options in order to make the most appropriate choices. Special aspects of performing quantitative Raman spectroscopy are discussed since these are required in many installations. Applications from many diverse fields are presented. The reader is encouraged to examine all of the areas since there are good lessons and stimulating ideas in all.

Raman spectra encode information about a sample's molecular structure and chemical environment. Fortunately, that information is readily accessible since Raman spectroscopy is an extremely flexible technique, both in what it can measure and how the measurements can be made.

7.1.1 Quantitative information

Under constant experimental conditions, the number of Raman scattered photons is proportional to analyte concentration. Quantitative methods can be developed with simple peak height measurements [1]. Just as with infrared calibrations, multiple components in complex mixtures can be quantified if a distinct wavelength for each component can be identified. When isolated bands are not readily apparent, advanced multivariate statistical tools (chemometrics) like partial least squares (PLS) can help. These work by identifying all of the wavelengths correlated to, or systematically changing with, the concentration of a component [2]. Raman spectra also can be correlated to other properties, such as stress in semiconductors, polymer crystallinity, and particle size, because these parameters are reflected in the local molecular environment.

Process Analytical Technology 2e Edited by Katherine Bakeev
© 2010 John Wiley & Sons, Ltd.

7.1.2 Flexible sample forms and sizes used as accessed without damage

Samples for Raman analysis may be solids, liquids, or gases, or any forms in between and in combination, such as slurries, gels, and gas inclusions in solids. Samples can be clear or opaque, highly viscous or liquids with lots of suspended solids, though certain solid particle sizes may be problematic. Little or no sample preparation generally is required for Raman experiments, which is particularly valuable in a process installation. Generally the sample is not destroyed or altered during the experiment, except for a few materials, such as black or fluorescent samples, that absorb enough laser light to be thermally degraded or are otherwise photoreactive at the laser wavelength. Even if a sample is heated by the laser, this problem often can be mitigated by using a larger sample volume and/or improving sample stirring or rotation.

Raman spectroscopy has no inherent sample size restriction, large or small, but in practice, it is fixed by the optical components used in each particular instrument design. The lower bound is set by the diffraction limit of light, roughly a few cubic micrometers depending on the numerical aperture of the optics used and the laser's wavelength [3]. The upper bound typically is controlled by practical considerations, such as ensuring sufficient laser power density at the sample. For example, Raman can be used to monitor the progression of batch, semibatch, or continuous reactions, from small-scale like the curing of an epoxy coating on a microscopic substrate to large-scale, like a polymer produced in a 25 000 gallon (94 635 L) vessel. Macroscopic to microscopic samples can be measured with the appropriate selections of laser wavelength, laser power, and optics.

7.1.3 Flexible sample interfaces

Raman spectroscopy offers a rich variety of sampling options. Fiber optics enable multiple remote measurement points, up to about 100 m away, collected simultaneously or sequentially with one instrument. Measurements can be made noninvasively or in direct contact with the targeted material. Different probe options are discussed and pictured in Section 7.4.5. With noninvasive or noncontact probes, there are several inches of air between the front lens of the probe and the sample. This configuration is particularly valuable for very dangerous or radioactive compounds, corrosives, or sterile or packaged items. Samples can be measured directly through clear glass or plastic bottles, bags, blister packs, or ampoules. Powders are most easily measured with a noncontact probe. Samples in opaque containers may be measured with wide-area illumination (WAI), spatially offset (SORS), or transmission Raman systems. However, SORS and transmission configurations currently are not readily available commercially, though that is anticipated to change soon; those used in literature reports generally were custom-modified from existing commercial systems.

Contact or immersion probes can be placed directly in the process. The fiber optics are protected behind a suitable window, such as sapphire, quartz, or glass, with the process material touching the window. Slipstreams are not required, but can be used. These probes are engineered to me*et almost* any specified chemistry or condition, including high or low temperature, high or low pressure, caustic or corrosive, viscous, slurries, or submerged. However, a noncontact probe might be a better solution if the process components, such as catalyst dust, molten polymer, or powder mixtures, are likely to build up on the window and attenuate signal. Immersion probes with solvent wash capabilities are available but may add to the installation's complexity. If a probe design without such washing capabilities is selected, it is important to ensure that the probe can be removed and cleaned when necessary. The relative time, effort, and expense involved in manual versus automatic options should be compared before choosing.

For smaller samples, Raman spectra can be collected through a microscope. Process microspectroscopy systems, such as might be used for semiconductor chip manufacturing or pharmaceutical high throughput

screening (HTS), generally are fully automated. The resolution of the system is determined by the optics and laser wavelength, but generally is around a few micrometers. The same sample interface does not have to be used at each point, though there are limits based on the system's optics. The flexible sampling configurations and long fiber lengths possible give Raman spectroscopy considerable advantages over mid-infrared (MIR), and even near-infrared (NIR) spectroscopy for process applications.

7.1.4 Attractive spectral properties and advantageous selection rules

Raman spectra with sharp, well-resolved bands are easily and quickly collected over the entire useful spectral range, roughly 50–4000 cm^{-1}, though the exact range depends on instrument configuration. Bands can be assigned to functional groups and usually are easily interpreted. For quantitative work, simple univariate calibration models often are sufficient. Calibration models can range from 0.1–100% for most compounds, and can reach detection limits of 100 parts per million (ppm) for strong or resonance enhanced Raman scatterers. For low level concentrations, it can be beneficial to use signal enhancement techniques, such as surface enhanced Raman spectroscopy (SERS) substrates, changes in laser wavelengths to activate resonance Raman (RRS), or the combination of the two, surface enhanced resonance Raman spectroscopy (SERRS) [4–6].

The differences in selection rules between Raman and infrared spectroscopy define the ideal situations for each. Raman spectroscopy performs well on compounds with double or triple bonds, different isomers, sulfur-containing and symmetric species. The Raman spectrum of water is extremely weak so direct measurements of aqueous systems are easy to do. Polar solvents also typically have weak Raman spectra, enabling direct measurement of samples in these solvents. Some rough rules to predict the relative strength of Raman intensity from certain vibrations are [7]:

- totally symmetric ring breathing >> symmetric > asymmetric
- stretching > bending
- covalent > ionic
- triple > double > single
- heavier atoms (particularly heavy metals and sulfur) > lighter atoms.

A practical rule is that Raman active vibrations involve symmetric motions, whereas infrared active vibrations involve asymmetric motion. For example, a symmetric stretching or bending mode would be Raman active, but an asymmetric stretching or bending mode is IR active. Thus, it is clear why highly symmetric molecules, particularly homonuclear diatomic species, such as –C–C–, –C=C–, –N=N–, or –S–S–, generate such strong Raman scatter and are correspondingly weak in the infrared. Some classic strong Raman scatterers include cyclohexane, carbon tetrachloride, isopropanol, benzene and its derivatives, silicon, and diamond. By comparison, some strong IR absorbers are water, carbon dioxide, alcohols, and acetone.

7.1.5 High sampling rate

Acquisition times commonly vary from seconds to minutes, often with negligible time between acquisitions, even when measuring multiple locations simultaneously (multiplexing). The dedication of different areas on the charge coupled device (CCD) detector to each measurement point makes this possible. The detectors used for MIR and NIR instruments cannot be multiplexed in the same fashion and must measure multiple samples sequentially.

7.1.6 Stable and robust equipment

Most process Raman instruments have few, if any, moving parts and thus are quite stable and robust. This reduces the number of parts likely to need replacement. The laser and detector shutter are the most common service items. Special utilities and consumables are not required. Proper enclosure selection allows instruments to survive harsh environments, such as installation outdoors or around vibrations. An instrument built to accurate specifications and well installed is unlikely to require much care or daily attention. However, it is important to remember that there is no such thing as a maintenance-free instrument. Limited additional maintenance will ensure robust performance for years.

7.2 Potential Issues with Raman Spectroscopy

Like all techniques, Raman spectroscopy has its limitations and disadvantages.

7.2.1 High background signals

Fluorescence is one of the biggest, most frequent challenges to collecting quality Raman spectra, but does not impact MIR or NIR techniques. Common approaches to mitigate the problem include using longer laser wavelengths and summing numerous short acquisitions. Fluorescence problems are not always predictable and often occur in samples not typically described as fluorescent. If the source of the fluorescence can be identified, sample treatment options may help. However, there are impressive examples of fluorescence problems being reduced when the transmission mode is used, so instrument changes may need to be considered also; see Section 7.5 [8]. Regardless, fluorescence remains one of the first issues a feasibility study must address [9].

It is quite difficult to collect Raman spectra of black materials. The sample is usually degraded, burned, or otherwise damaged from absorption of the intense laser energy. Any Raman signal is masked by strong blackbody radiation. Many of these samples are equally challenging for MIR or NIR. Raman spectra of highly colored materials can be similarly difficult to collect, though the technique has been used to great advantage in examinations of paintings and pottery [10–12].

7.2.2 Stability

Subtle changes to the position and shape of sharp Raman bands are indicative of small changes in the local chemical environment. This makes the technique very sensitive and suitable for challenging chemical problems, but it also puts substantial demands on instrument stability. Small changes in laser wavelength or instrument environment could appear as wavenumber shifts and be mistaken for chemical changes. Instrument fluctuations can impact quantitative results considerably [13]. Besides using the highest quality, most stable instrument available for such applications, there are other approaches that can be used to reduce the potential for problems.

One of the most common approaches to enhance stability is to mathematically align reference or pseudoreference peaks, such as from a solvent or unchanging functional groups, prior to use in a calibration model. In other cases, it can help to mathematically reduce peak resolution slightly, analogous to using peak areas instead of peak heights. If an instrument has not yet been purchased, changes to the grating to increase resolution by sacrificing some spectral coverage may help minimize sensitivity to the fluctuations. Signals from solvent also can be used like an internal standard [14]. While these approaches may seem like hurdles, they are analogous to learning to handle NIR spectra with multiplicative scatter correction or second deriva-

tives. Similar problems plague NIR instruments, but are less noticeable on the broad, ill-defined peaks. While Raman instrument stability is excellent for many applications, users must remain alert to the possibility of a problem in long-term use and develop strategies to mitigate this risk.

7.2.3 Too much and still too little sensitivity

Raman spectroscopy's sensitivity to the local molecular environment means that it can be correlated to other material properties besides concentration, such as polymorph form, particle size, or polymer crystallinity. This is a powerful advantage, but it can complicate the development and interpretation of calibration models. For example, if a model is built to predict composition, it can appear to fail if the sample particle size distribution does not match what was used in the calibration set. Some models that appear to fail in the field may actually reflect a change in some aspect of the sample that was not sufficiently varied or represented in the calibration set. It is important to identify any differences between laboratory and plant conditions and perform a series of experiments to test the impact of those factors on the spectra and thus the field robustness of any models. This applies not only to physical parameters like flow rate, turbulence, particulates, temperature, crystal size and shape, and pressure, but also to the presence and concentration of minor constituents and expected contaminants. The significance of some of these parameters may be related to the volume of material probed, so factors that are significant in a microspectroscopy mode may not be when using a WAI probe or transmission mode. Regardless, the large calibration data sets required to address these variables can be burdensome.

Raman band intensity is a known function of temperature, but affects high and low frequency bands unequally. Quantitative results will be biased if calibration and process samples are measured at different temperatures. A separate temperature measurement may enable the results to be corrected. Even qualitative interpretations of band changes, such as inferring that a species has been consumed in a reaction, may be skewed by temperature-related changes in population distributions. Similarly, Raman spectra are very sensitive indicators of hydrogen bonding changes, which are affected by temperature. It is quite possible to accidentally turn the spectrum and instrument into an extremely expensive thermometer. It is possible to obtain temperature and composition from a single spectrum if carefully done, which is an advantage or disadvantage depending on the application. Regardless, temperature must be considered during the development and implementation of a calibration model.

These issues are not unique to Raman spectroscopy, but users may not be as aware of potential problems as for other techniques. Good understanding of the chemistry and spectroscopy of the system being modeled along with well-designed calibration experiments can help prevent problems. The power of the technique can be a hindrance if not recognized and managed.

Conversely, Raman spectroscopy has been described as an insensitive technique. Except with resonance enhanced materials, very strong scatterers, or special enhancement techniques, Raman spectroscopy is generally considered to have a detection limit around 0.1% [15]. The technique often is described as being poorly suited to detecting small changes in concentration and this has hindered the technique's use. It is hard to generalize since the limit depends on the situation and equipment configuration. It is important to test sensitivity and detection limits for target applications during feasibility studies and to recognize the demands those requirements put on equipment.

7.2.4 Personnel experience

In the past, process Raman applications largely were developed and implemented by formally trained Raman spectroscopists, aided by process engineers and other personnel. However, there are far fewer Raman spectroscopists than chemical, process, or production engineers or other types of personnel. Increasingly, these

professionals are identifying Raman spectroscopy as a potential sensor for their systems, arranging for or conducting feasibility studies, specifying equipment, and overseeing installation, model development and validation, and routine operation of the system. Instrument manufacturers are recognizing this shift and beginning to offer systems and services tailored to support this type of customer. However, it still may be necessary to work with a Raman spectroscopist, either an employee or a consultant, on sophisticated, cutting-edge applications; more straightforward applications also may go faster and smoother with technically experienced assistance.

7.2.5 Cost

Instruments are available at more price points and from more vendors now than a decade ago, enticing ever more users to try Raman. A 2007 overview of Raman instrumentation lists prices from US$12 000–500 000 depending on configuration and category (handheld, laboratory, or process) – an astonishing range [16]. Generally, handheld or portable instruments are least expensive. For process analyzers, the configuration differences include the number of measurement points required and distance of each from the analyzer, the kinds of probes used, the demands on an enclosure from weather and electrical classification, the system used to communicate with a facility's host computer, and required redundancies. In a chapter on process Raman spectroscopy, Pelletier wrote that the 'analyzer may cost from $50 000 to $150 000, and the integration of the analyzer into the production facility can cost as much as the analyzer itself' [15]. As the competition between vendors plays out, users no longer need to wonder whether they can afford a Raman instrument; instead, they need to ask if the one they can afford can deliver the performance required. As with anything else, flexibility and maximal performance generally cost more.

In addition, the cost of ownership should not be neglected. Lasers will need to be replaced and are expensive. As with the systems themselves, there is a wide range in laser costs depending on the quality required. Popular lasers, such as diode lasers at 785 nm and Nd : YAG lasers at 532 nm, have estimated lifetimes of just 1–1.5 years. Though lasers can sometimes be repaired, the usual replacement cost in a process system is $10 000–$20 000. While a NIR process unit might require the light source to be changed every 6–12 months, even the most expensive light bulb is practically free compared to a laser. The costs of reduced manufacturing yield, poor product quality, or unsafe sampling or operating conditions can easily justify the laser replacement expenses for many installations, but that decision needs to be made at the start of a project. Additionally, it is important to calculate the cost of calibrating the system. In some instances, it may be necessary to establish a small-scale version of the process to measure unusual samples and alter other variables outside the normal ranges in order to ensure a robust calibration model. If this is not already available, it can be a considerable expense. No matter how a firm calculates its required return on investment, such initial and ongoing expenses demand that an installation delivers substantial benefits.

7.3 Fundamentals of Raman Spectroscopy

Raman spectroscopy is based on an inelastic scattering process involving an energy transfer between incident light and illuminated target molecules. A small fraction of the incident light is shifted from its starting wavelength to one or more different wavelengths by interaction with the vibrational frequencies of each illuminated molecule. The molecule's vibrational energy levels determine the size of the wavelength shift and the number of different shifts that will occur. This amounts to a direct probe into the state and identity of molecular bonds, and is why Raman spectroscopy is so useful for chemical identification, structure elucidation, and other qualitative work.

Stokes Raman shift occurs if energy is transferred from the incident light to the molecule, resulting in red-shifted scattered light, with a longer wavelength and lower energy. In the molecule, a ground state electron is excited from a lower vibrational energy level to a higher one. Anti-Stokes Raman shift occurs when energy is transferred to the light from the molecule. The observed light is blue-shifted, with a shorter wavelength and higher energy. A ground state electron in the molecule is excited from a higher vibrational energy level through a virtual state and ends at a lower level. The molecular bonds vibrate at a lower frequency. In either Stokes or anti-Stokes scattering, only one quantum of energy is exchanged. The absolute energy difference between incident and scattered radiation is the same for Stokes and anti-Stokes Raman scattering.

The Raman effect is weak. Only approximately one in 10^8 incident photons is shifted. Of these few Raman-shifted photons, even fewer (relative fraction at room temperature) are anti-Stokes shifted because fewer molecules start in the required excited state. The Boltzmann distribution describes the relationship between temperature and the fraction of molecules in an excited state. As temperature increases, the relative proportion of ground and excited states changes and the Stokes to anti-Stokes intensity proportion changes accordingly. The rest of those 10^8 photons are known as Rayleigh scatter. Since the electrons start and finish at the same vibrational energy level, the photons also start and finish at the same wavelength. Everything is unchanged by the encounter with a molecule. A representative vibrational energy level diagram is shown in Figure 7.1 [4]. The challenge of Raman spectroscopy is to detect the Raman 'needle' in the Rayleigh 'haystack.'

A monochromatic light source, such as a laser, is used to generate enough Raman scattered photons to be detected. A complete Raman spectrum is symmetric about the wavelength of the incident light, with the Stokes portion typically presented on the right side and anti-Stokes on the left side when plotted on an increasing wavelength axis. However, the Stokes and anti-Stokes portions will not be mirror images of each other because of their relative intensity differences. A complete Raman spectrum of carbon tetrachloride is

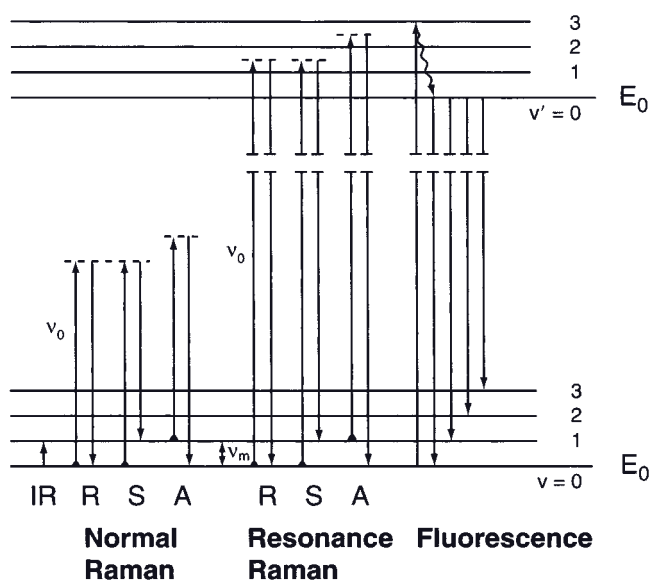


Figure 7.1 *Energy level diagram illustrating changes that occur in IR, normal Raman, resonance Raman, and fluorescence. Notation on the figure stands for Rayleigh scattering (R), Stokes Raman scattering (S), and anti-Stokes Raman scattering (A). Reprinted from Ferraro* et al. *(2003) [4] with permission from Elsevier.*

shown in Figure 7.2 [7]. Since classic Raman spectroscopy typically has been done at or near room temperature and since the Raman effect is inherently weak, most practitioners collect only the Stokes portion of the spectrum. Just as in infrared spectroscopy, the normal *x*-axis units for Raman spectra are wavenumbers, or inverse centimeters, which enables Raman spectra to be compared regardless of the wavelength of the incident light.

Energy transfer actually occurs as a two-photon event:

1. one photon interacts with the electron cloud of the molecule, inducing a dipole moment and modifying the vibrational energy levels of electrons and then
2. provokes emission of a second photon in response to those changes.

However, because this occurs through one quantum mechanical process, these should not be considered separate absorption and emission steps [17]. Raman instruments count the number of photons emitted as a function of their wavelength to generate a spectrum. This differs substantially from IR spectroscopy where incident light is actually absorbed by the vibrational energy transition of the molecule. IR instruments measure the loss of intensity relative to the source across a wavelength range. Raman and IR spectroscopy both probe the vibrational energy levels of a molecule, but through two different processes.

Accordingly, the selection rules for Raman and IR spectroscopy are different. In Raman spectroscopy, there must be a change in the molecule's polarizability upon excitation, whereas a change in dipole moment is required for IR. A dipole moment is the magnitude of the electronic force vector between the negative and positive charges or partial charges on a molecule. A permanent dipole moment exists in all polar mol-

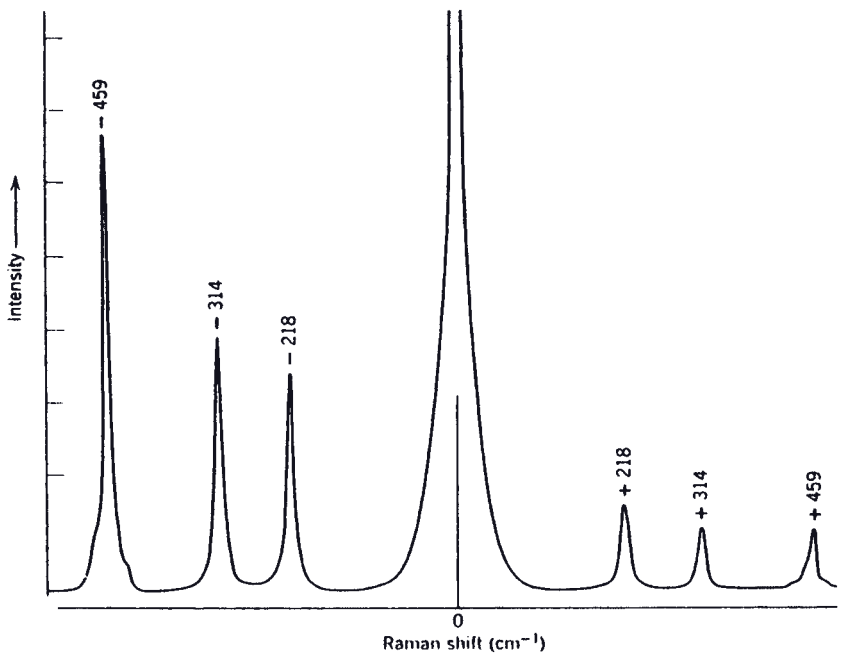


Figure 7.2 *Complete Raman spectrum of carbon tetrachloride, illustrating the Stokes Raman portion (on left, negative shifts), Rayleigh scattering (center, 0 shift), and the anti-Stokes Raman portion (on right, positive shifts). Reprinted from Nakamoto (1997) [7] and used by permission of John Wiley & Sons, Ltd., Chichester, UK.*

ecules. Changes in dipole moment mean that the magnitude of the charge or the distance between the charges has changed. When the oscillating electric field of a light source interacts with a molecule, the positively charged nuclei and negatively charged electrons move in opposite directions, creating a charge separation and an induced dipole moment [7]. Polarizability reflects the ease with which this occurs [18]. Changes in polarizability, or the symmetry of the electron cloud density, can occur for polar or nonpolar molecules. Since polyatomic molecules can move in any combination of x, y, or z axes, there are 3^2, or 9, possible directions of motion to evaluate. A change in any of them means Raman scattering can occur [7]. The mutual exclusion principle states that there are no shared transitions between IR and Raman for molecules with a center of symmetry. Both Raman and IR bands can be observed for molecules without a center of symmetry, though the relative intensities will be quite different. Raman overtones and combination bands are forbidden by the selection rules (symmetry) and generally are not observed.

7.4 Raman Instrumentation

Any Raman instrument consists of a laser as a monochromatic light source, an interface between sample and instrument, a filter to remove Rayleigh scatter, a spectrograph to separate the Raman scattered light by wavelength, a detector, and a communications system to report analysis results. Purchasers must make decisions about laser wavelength and stability, appropriate sampling mechanism, spectrometer design, and a communications system. Spectrometer design is less of a factor than in the past, since the sampling mechanism may determine the spectrometer available to go with it. The filter and detector choices also generally follow from these decisions. Electricity generally is the only utility required; purge gases, liquid nitrogen, and cooling water are not required. Normally, attaching the fiber optics and plugging in an electrical cord are the only connections required to get the equipment running. Few adjustments or alignments are required.

Common characteristics of all viable process Raman instruments include the ability to measure at multiple locations, simultaneously or sequentially, and in a variety of electrically classified or challenging environments with no special utilities. The instrument must be reasonably rugged and compact, and not require much attention, servicing, or calibration [19]. Modern process Raman instruments essentially are turnkey systems. It is no longer necessary to understand the optical properties and other instrument details to successfully use the technique. The general design options will be described briefly, but more information is available elsewhere [20,21]. This section will focus primarily on special instrumentation considerations for process installations.

7.4.1 Safety

Safety must be the first consideration of any process analytical installation. Electrical and weather enclosures and safe instrument-process interfaces are expected for any process spectroscopy installation. The presence of a powerful laser, however, is unique to process Raman instruments and must be addressed due to its potential to injure someone. Eye and skin injuries are the most common results of improper laser exposure. Fortunately, being safe also is easy. Because so many people have seen pictures of large industrial cutting lasers in operation, this is often what operations personnel erroneously first envision when a laser installation is discussed. However, modern instruments use small footprint, comparatively low power lasers, safely isolated in a variety of enclosures and armed with various interlocks to prevent accidental exposure to the beam.

It is important and worthwhile to spend time educating project participants about the kinds of lasers used, their power, safety classifications, proper handling and operation, and risks [22–24]. Ample free educational resources are available on the internet. Operations and maintenance personnel must receive regular laser

safety training commensurate with their expected level of interaction with the instrument. This might range from general awareness of a potential hazard in the area to standard operating procedures to lock-out/tag-out procedures when working on the instrument, laser, or fiber optics. Government laser safety regulations have specific requirements depending on the laser wavelength and power [22,25–27]. Common requirements include restricting access to the laser and its emission points, both to prevent accidental laser exposure and ensure uninterrupted operation, and wearing wavelength specific safety glasses whenever the laser beam is exposed. Appropriate safety warning signs and labels must be used.

Vendors who sell process Raman equipment have implemented many safety features as part of their standard instrument configurations. However, it remains the responsibility of the purchaser to ensure a safe environment and compliance with their company, industry, and governmental safety regulations. Anything involved with an industrial process has associated hazards that must be controlled to eliminate or reduce risks. Those originating from the laser in a Raman instrument are readily managed when given due consideration and must not be ignored.

7.4.2 Laser wavelength selection

Raman signal intensity is proportional to the fourth power of the inverse of the incident wavelength. As Table 7.1 illustrates, the use of a 785 nm NIR laser instead of an ultraviolet (UV) one at 480 nm means that only 14% of the possible photons could be generated. It is desirable to use the shortest laser wavelength possible to maximize the number of Raman photons available for detection.

Fluorescence is a common interference and its severity tends to increase with shorter wavelengths. By using longer wavelength lasers, many fluorescence problems can be avoided, but this comes at the expense of greatly reduced Raman intensity and increased risk of overheating the sample. CCD detectors have limited response functions at longer wavelengths. Lasers and detectors should be selected together in order to minimize limitations on spectral coverage. Project feasibility studies are an excellent time to evaluate different laser wavelengths. The selection of a laser wavelength to balance the trade-offs between maximizing Raman signal and minimizing fluorescence is a critical decision.

A dramatic enhancement in Raman signal can be obtained when the laser wavelength falls in the sample's electronic absorption band. This is referred to as resonance Raman spectroscopy (RRS) and it greatly increases sensitivity and improves detection limits. All vibrational bands in the spectrum are not enhanced equally since the resonance is only with the chromophoric functional group. It makes the system highly selective, though that could be a disadvantage if more than one component needs to be measured [7]. Resonant systems are prone to local sample heating so care must be taken to ensure that samples are not exposed to the laser too long or that there is an efficient sample cooling mechanism. One should take advantage of RRS whenever possible.

7.4.3 Laser power and stability

Higher laser power is preferred since it proportionally reduces acquisition time (signal strength). The maximum sampling rate in a process application is limited by acquisition time, which depends on the molecule's Raman scattering efficiency, though adjustments to the standard data transfer hardware and software

Table 7.1 *Decrease in Raman intensity as a function of laser wavelength*

Wavelength λ (nm)	480	532	633	785	840	1064
Intensity $(1/\lambda)^4$ (relative %)	100%	66.3%	33.1%	14.0%	10.7%	4.1%

might be needed for very high rates. High laser power can cause local sample heating, which may damage static samples unless heat can be dissipated. The ready availability of quality lasers has made process Raman instruments possible and further improvements will translate directly into increased sensitivity and power of Raman spectroscopy.

Laser stability is a key differentiator between commercial instruments. Some vendors choose to use the most stable lasers (constant power, wavelength, and mode) available whereas others use less robust lasers but continuously correct for fluctuations with math and programming. The first option is more expensive but conceptually easier to understand; the second is less expensive but quite challenging for the instrument designer and programmer, and sometimes the user. Both are valid approaches, though many Raman spectroscopists prefer the first approach. However, even with very stable lasers, model robustness often benefits from some implementation of the second approach to eliminate the effects of long-term fluctuations and changes. The choice between these approaches will depend on the spectral sensitivity of the application, the analysis requirements, and project budget.

7.4.4 Spectrometer

Raman vendors often seek to establish competitive positions through their spectrometer's design, resulting in many variants on the basic dispersive (axial transmissive, concave holographic, Echelle, and Czerny–Turner) and interferometric (Fourier transform and Fabry–Perot) designs [28]. Most dispersive Raman systems use a silicon-based CCD as a detector. The system's resolution is a function of the grating, spectral range, detector width in pixels, and pixel size. The detector width is usually the limiting factor. In some instances, the sampling probe's design dictates which spectrometer can be used for a given application. Key spectrometer design evaluation criteria include optical throughput or efficiency, spectral coverage, spectral resolution, instrument stability and robustness, and instrument size. As with all process spectroscopic instruments, it is essential to use the right environmental enclosure. In general, process Raman instruments must be protected from weather and kept thermally stable. A variety of manufacturers offer process-ready systems that have been specifically engineered to handle common environmental challenges, such as temperature fluctuations, vibrations, weather, and electrical classification. Similarly, a convenient, automated mechanism must be included for calibrating the spectrometer wavelength, laser wavelength, and optical throughput or intensity axes [29].

Axial transmissive systems frequently are used in process installations, and are built around volume holographic gratings or grisms (combination grating and prism). Light travels along a single axis, passes through the grating, and is dispersed on the detector. Additionally, since the grating and all optics are mechanically fixed, these designs have excellent stability and a small footprint. Because there are few optical surfaces, the optical throughput is high. Simultaneous measurement of multiple locations is accomplished by dedicating a location on the detector for each. The holographic grating enables a complete spectrum to be split in half, usually around a low information portion of the spectrum like $1750\,cm^{-1}$, with each half directed to a different portion of the detector, offering comparable resolution, around $5\,cm^{-1}$, to other designs. Simultaneous acquisition of a complete spectrum makes it effective for measuring rapidly changing systems. The Rayleigh scatter is removed with a holographic notch filter, which is standard in the industry [30,31]. The grism enables numerous fibers to be imaged onto the detector simultaneously, thus offering the possibility of monitoring tens of channels simultaneously. However, WAI systems are limited to a single probe covering a narrower spectral range. In a related dispersive approach, concave holograms can be used as both the grating and imaging optics to build a one-element spectrograph. Since there are so few components, these instruments are compact, inexpensive, and stable. Over long wavelength ranges, some image quality problems may be observed [20]. In an alternative design, lasers at two wavelengths are used instead of two gratings to obtain a full Raman spectrum with good resolution and minimal fluorescence [32].

Fourier transform Raman (FT-Raman) systems are mainly used to avoid substantial fluorescent backgrounds by exciting the sample with long laser wavelengths. The systems generally use a 1064 nm laser, though a system is available at 785 nm. As with dispersive systems, the longer excitation wavelength greatly reduces signal intensity. The poor sensitivity of these systems has limited their process utility, but may be the best choice when fluorescence is severe. A high laser power often is used to improve sensitivity, but that frequently heats the sample excessively and can trigger unexpected photochemical reactions. The instruments typically use indium gallium arsenide (InGaAs) or germanium detectors since the photoelectric response of silicon CCDs at these wavelengths is negligible.

7.4.5 Sample interface (probes)

There have been substantial changes in Raman sample interfaces recently. The approaches now can be divided broadly into two categories based on the sampling volume. The probes intended to sample small volumes include the more traditional noncontact or standoff probes, immersion probes, and optical microscopes. Large volume sampling approaches are newer and include WAI and SORS probes and transmission configurations.

All of the approaches generally use fiber-optic cables, up to approximately 100 m, between measurement point and instrument, with a dedicated path to carry laser light to the sample and a separate path to return the Raman signal to the spectrometer. Silica Raman scatter always is generated when the laser travels through fiber optics. It will overwhelm any analyte signal if it is not removed before the laser reaches the sample. For similar reasons, the laser must be blocked from the return fiber. The optical path of one commercial probe design is shown in Figure 7.3. Though this task can be accomplished in different ways, all probes use optics, which must be protected from dirt and weather with a suitable enclosure or housing. Most probes use a 180° backscatter configuration, which refers to the angle between the sample and the Raman scatter collection optics; transmission measurements use a 0° configuration. Because the illumination and collection optics occupy essentially the same space in backscatter designs, Raman spectroscopy can be performed using a single probe inserted directly in a process stream.

Raman spectroscopy's flexible and minimally intrusive sampling configurations reduce sample pretreatment demands. Sampling interfaces can be engineered for almost any environment, but these requirements must be clearly defined before probes are built. No probe can be made to survive all environments so there may be some design trade-offs. In a noncontact system, the laser usually comes to a focus several inches

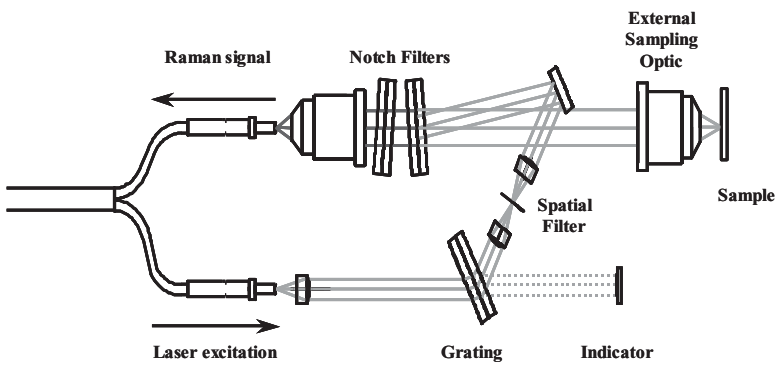


Figure 7.3 *Optical path of a commercial Raman probe. Adapted, with permission, Copyright © 2004 Kaiser Optical Systems, Inc.*

away from the probe's optical face. A clear, process-compatible sight glass may be placed in between the probe and the focal point if necessary, which also must be kept sufficiently clear. Depending on the focal distance, it may be necessary to enclose the laser beam on noncontact probes to address safety concerns. Different optics can be attached to the face of the noncontact probe, much like conventional cameras use different lenses. Alternatively, the fibers can be used to launch the laser into the optical path of a microscope, providing high spatial resolution. In a contact or immersion system, the probe is inserted through a fitting and its optical window touches the targeted material. The laser beam travels down the empty shaft, coming to a focus just outside the optical window at the probe's tip. Many immersion probes have a second window as a safety backup to ensure the process remains contained in the event of a window seal failure. Since the sample does not have to pass through a narrow gap, viscous streams or ones with high solids' content often can be measured. Probes have been designed with motorized focus adjustments, automatic solvent or steam washes or air jets to keep the window clear, and sealed in water and pressure safe containers. The main sampling challenge is ensuring the interface is designed to keep the window clear enough for light to pass through.

Four styles of commercial probes are shown in Figure 7.4. In Figure 7.4a, a laboratory scale immersion probe is shown and it is installed in a laboratory scale fermentation reactor in Figure 7.4b. The immersion optics on this probe easily can be removed and replaced with a noncontact optic for flexible methods development studies. In Figure 7.4c, a commercial scale immersion optic is shown; it is installed in a glass reactor in Figure 7.4d. In Figure 7.4e, a gas phase probe is shown with a flow through cell and it is installed in process piping in Figure 7.4f. A WAI probe is shown installed in a pharmaceutical tablet coater in Figure 7.4g. The probe has been coated along with all of the pharmaceutical tablets. After the arm holding the probe is withdrawn from the vessel, it will be hosed down and returned to service. The choice between the approaches usually depends on which implementation would be the safest, easiest, and most robust.

The size, number and arrangement of the fibers differ between the approaches. For small sample volume arrangements, the cables typically contain a single fiber for each path; a spare pair may be included in the same cladding. The specific fiber selected will determine the volume sampled, and this normally ranges between a few cubic micrometers to 500-μm diameter spot size [33,34]. The small sample volume probes work extremely well when the sample is homogeneous on a size scale equal to or smaller than the volume probed. However, industrial production materials often are not homogeneous. A common solution to this problem is to collect Raman spectra continuously while the sample is rotated under the probe. Another way around this problem is to illuminate and collect signal from a larger sample volume. The need for this was recognized many years ago [35]. See Chapter 3 for advice on proper sampling. The WAI approach uses a bundle of commingled illumination and collection fibers. The illumination fibers and associated optics effectively probe a 3–6-mm diameter spot, instead of 0.5 mm; the signal from all of the collection fibers is added together to give one spectrum [33,34,36]. An example is shown in Figure 7.4g.

SORS probes interrogate a large sample volume despite illuminating only a single 200–300-μm spot [37]. The laser photons propagate through the sample and Raman signal is detected using a bundle of fiber optics arrayed in a ring at a set diameter from the illumination spot. The exact volume probed is difficult to determine since it really depends on how deep the laser photons travel and on how the light bounces around. Inverse SORS probes illuminate a ring on the sample and collect Raman signal from the center of the ring [38].

Though the use of transmission geometry is common for many other spectroscopic techniques, it has not been widely used for Raman spectroscopy [39] In this case, illumination and collection optics are on opposite sides of the sample. The actual generation and travel of Raman photons through the sample is convoluted, but it is safe to conclude that the bulk of the sample is probed [40,41]. The large sample volume probed results in reduced subsampling errors. In one example, the use of the transmission mode enabled at least 25% reduction in prediction error compared to a small sampling area probe [42]. The approach is insensitive

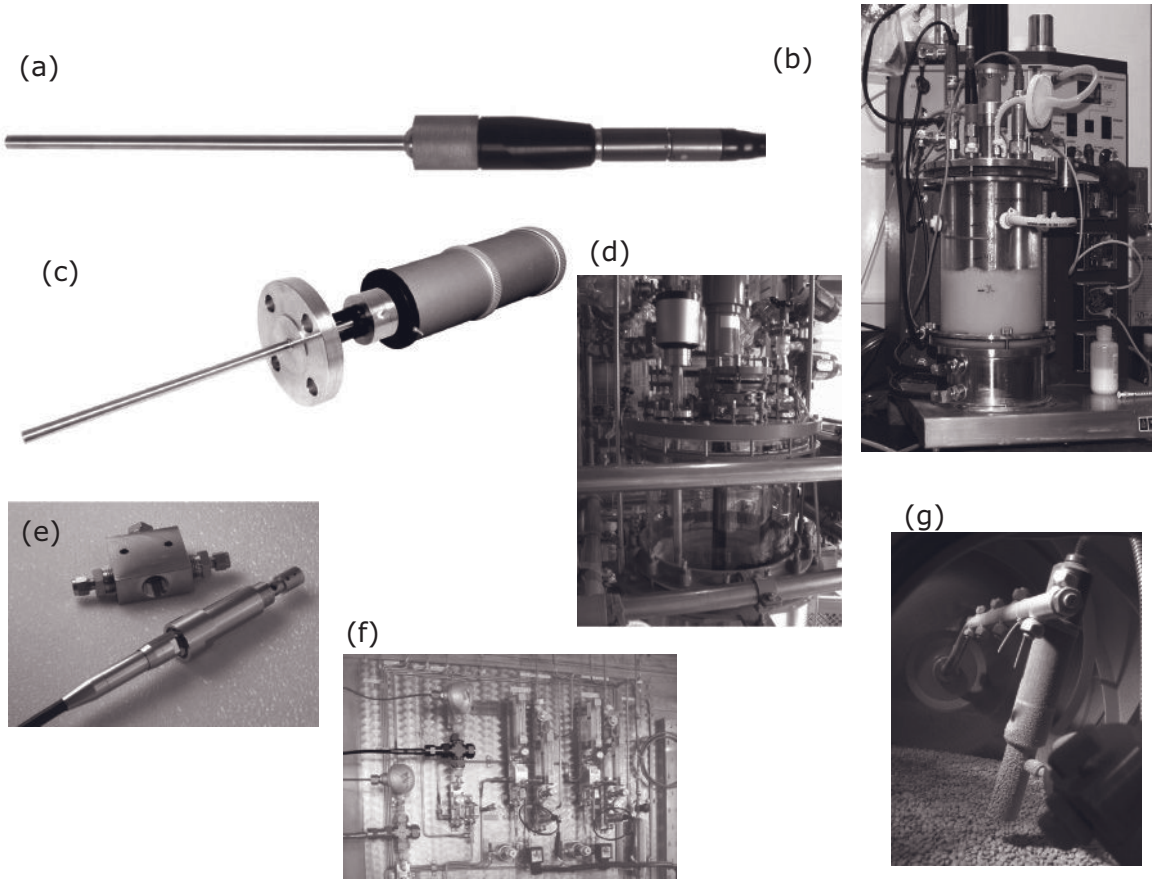


Figure 7.4 *Collection of commercial Raman probes designed for different installations: (a) laboratory scale probe with interchangeable immersion or noncontact optics, shown with immersion option; (b) probe shown in (a) installed in laboratory fermentation reactor; (c) production scale immersion probe; (d) probe shown in (c) installed in a glass reactor; (e) gas phase probe with flow through cell; (f) probe shown in (e) installed in process piping; (g) wide area illumination (WAI) noncontact probe after completion of a pharmaceutical tablet coating operation. Adapted, with permission, Copyright © 2004 Kaiser Optical Systems, Inc.*

to minor constituents, which is an advantage when quantifying the bulk concentration of a pharmaceutical tablet and a disadvantage when quantifying the coating on the tablet.

7.4.6 Communications

In order for the data generated by the analyzer to be useful, it must be transferred to the operation's centralized control or host computer and made available to process control algorithms. Vendor packages manage instrument control and can do spectral interpretation and prediction or pass the data to another software package that will make predictions. Most vendors support a variety of the most common communications

architectures. In addition to reporting predictions, the software also typically reports a variety of possible instrument errors.

7.4.7 Maintenance

The routine maintenance burden of a Raman system is quite low. Optical windows may need to be cleaned, and the difficulty of this should be addressed when selecting a sample interface. In some cases, automatic window cleaning systems can be implemented if it is a known issue. Periodically, the wavelength axis, laser wavelength, and intensity axis may need to be calibrated. Vendors provide different means for accomplishing this. The most likely repair activity is laser replacement. Some systems have a backup laser that turns on automatically if the primary laser fails. This lessens the impact of a failure if the unit is being used for closed loop process control. In addition, it is important for the instrument and calibration models to have strong self-check and fault detection routines, particularly since a Raman expert often is not available on site to diagnose problems. However, it is inevitable that the system will require model recalibration or updates at some point. The quality of plans to minimize the work associated with that may determine the project's long-term success. With a quality instrument and well-developed models, process Raman installations need less maintenance than many competing techniques.

7.5 Quantitative Raman

Quantitative Raman spectroscopy is an established technique used in a variety of industries and on many different sample forms from raw materials to in-process solutions to waste streams, including most of the applications presented here [1]. Most of the applications presented in the next section rely on quantitative analysis. Similar to other spectroscopic techniques, many factors influence the accuracy and precision of quantitative Raman measurements, but high quality spectra from representative samples are most important.

Raman often is evaluated as an alternative to an existing high performance liquid chromatography (HPLC) method because of its potential to be noninvasive, fast, simple to perform, and solvent-free. Raman was compared to HPLC for the determination of ticlopidine-hydrochloride (TCL) [43], risperidone [44] in film-coated tablets, and medroxyprogesterone acetate (MPA) in 150-mg/mL suspensions (DepoProvera, Pfizer) [45]; it was found to have numerous advantages and performance suitable to replace HPLC. In an off-line laboratory study, the relative standard deviation of the measurement of the composition of powder mixtures of two sulfonamides, sulfathiazole and sulfanilamide, was reduced from 10–20% to less than 4% by employing a reusable, easily prepared rotating sample cell [46].

The effect of different sample volumes on the quality of composition predictions was the subject of several studies. Scientists at AstraZeneca collected Raman spectra from 8-mm diameter tablets of different compositions illuminated in four different ways: as five separate points in the shape of a cross, as a small (2.5-mm diameter) ring, as a large (5 mm) ring, and as a circular area (5 mm) [47]. The samples were rotated during data acquisition, except for the point illumination. Illumination pattern had a substantial effect on prediction error. The large ring and area had equivalent errors, less than 2%, and both were smaller than either the small ring or point cross. The size-scale of the compositional heterogeneity, or subsampling, dictates which illumination pattern and size are required for accurate quantification. Similar findings were reached in a study using a commercially available probe that illuminated a circular area (3-mm diameter.) [34]. Wikström *et al.* measured blend quality simultaneously utilizing three different Raman probe configurations: an immersion probe and small- and large-illumination area noncontact probes [33]. Despite constant laser power at

the sample, a higher sampling rate could be achieved with the large area probe since its required acquisition time was 80% less than the other probes. The effective sampling volumes of the large and small area probes, respectively, were more than 15 000 and 12 times larger than that of the immersion probe. With the small area probe, useful results were obtained by defocussing, helpful information for those who already own one. Interestingly, the authors of these studies comment that the larger-area systems are more robust to variations in how the system is focused on the sample and thus more precise.

Fluorescence from pharmaceutical capsule shells and tablet coatings has hindered measurement of their composition by Raman spectroscopy. By switching from the conventional backscattering mode to a transmission mode, Matousek *et al*. demonstrated that fluorescence could be eliminated in many instances [8]. Backscattering- and transmission-mode Raman spectra of several samples are shown in Figure 7.5. Each spectrum was acquired in 10 s with 80 mW 830-nm laser power. Matousek *et al.* also speculate that 'signal acquisition times could be relatively easily shortened to well below 0.1 s' when the transmission mode is combined with optimized optics [8].

Newcomers to Raman spectroscopy often are amazed that useful spectra can be collected directly from materials packaged in clear glass bottles, plastic bags or blister packs. However, practitioners quickly become used to this, and want to be able to analyze materials through opaque packaging materials, ideally also those in dark colors. Opaque materials and turbid systems have been problematic to analyze, but new probe designs and larger sampling areas have improved the measurements. Kim *et al*. measured the concentration of povidone in commercial eyewash solution through a low-density polyethylene (LDPE) bottle with conventional and WAI (6-mm diameter) probes [48]. The relative standard deviations of the ratio of the LDPE and povidone peaks were 32.5% (conventional) and 3.2% (WAI), a dramatic difference. In the WAI system, the use of an external standard for laser intensity corrections enabled the standard error of prediction (SEP) to be reduced from 0.46% to 0.06%. WAI resulted in substantially better performance compared to illumination with a more traditional smaller spot size to determine the concentration of acetaminophen in syrup solution, which is a turbid solution at some concentrations [36]. De Beer presents a second-order polynomial calibration model to measure the concentration of salicylic acid in white petroleum jelly ointment as a suitable alternative to HPLC [49]. Successful predictions are achieved only when the drug particle sizes in the unknown samples match those in the calibration set. This suggests that scattering effects may be present. Though not discussed as a turbid system, it would be interesting to examine the system with other illumination techniques.

TiO_2 is used frequently to coat pharmaceutical capsules, but its very strong Raman signal can obscure the signal of other components in the mixture. Using transmission geometry, Matousek *et al*. measured the active pharmaceutical ingredient (API) concentration in TiO_2-coated capsules in 5 s with 1.2% relative root mean standard error of prediction (RMSEP) or 1.8% relative RMSEP in 1 s [41]. The authors present an interesting comparison of the different Raman spectra that can be obtained on the same sample, depending on the optics. As shown in Figure 7.6, the Raman spectrum of a capsule measured with a traditional Raman system is dominated by signal from the capsule's exterior. With defocusing, the signal from the capsule contents is evident. The contents' signal is intensified with a WAI probe and strongest in transmission mode. This serves as a reminder that there is not a single best optical configuration; the nature of the problem to be solved and the constraints on the solution will dictate the best choice.

Since a larger sample volume is presumed to be probed, the use of transmission mode has led to simpler, more accurate models requiring fewer calibration samples [50]. Scientists at AstraZeneca found that with a transmission Raman approach as few as three calibration samples were required to obtain prediction errors nearly equivalent to their full model [42]. For a fixed 10-s acquisition time, the transmission system had prediction errors as much as 30% less than the WAI system, though both approaches had low errors. It is hoped that this approach in combination with advanced data analysis techniques, such as band target entropy minimization (BTEM) [51], might help improve Raman's quantitative sensitivity further.

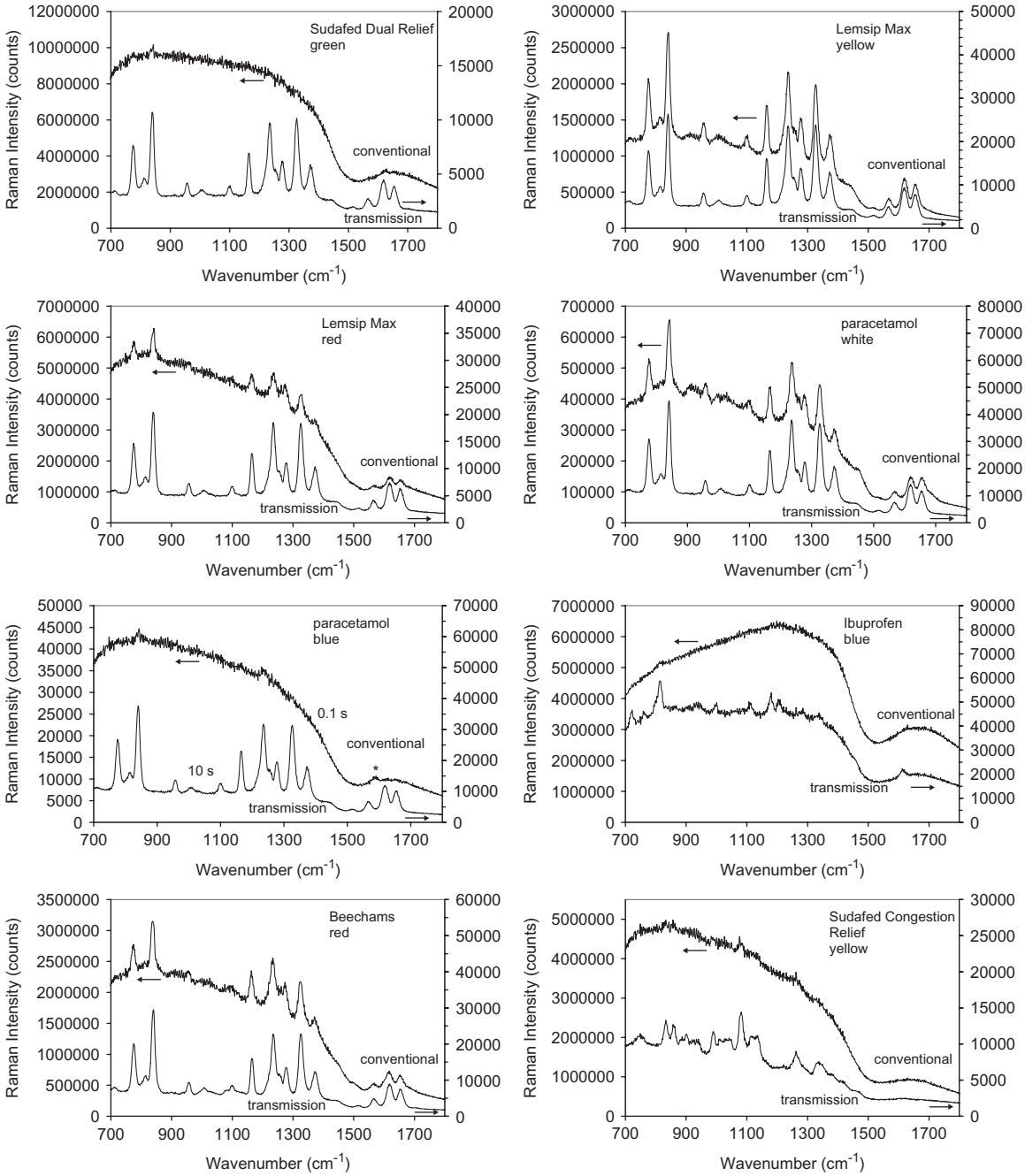


Figure 7.5 *Raman spectra of a series of tablets with different coatings to illustrate improvements in spectral quality using transmission-mode instead of conventional backscatter geometry. Bands from the capsule shell are marked with a * symbol. Reprinted from Matousek* et al. *(2007) [8] with permission from John Wiley & Sons, Ltd.*

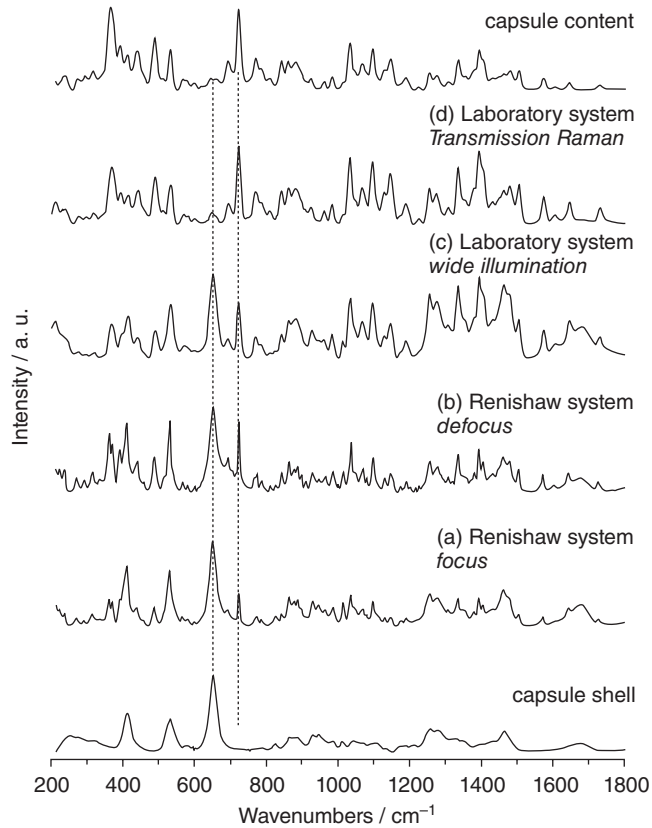


Figure 7.6 *Raman spectra of a pharmaceutical capsule using different measurement configurations. (Bottom) Capsule alone. (a) Commercial backscattering system with small sample area (b) Commercial backscattering system defocused to make sample area slightly larger (c) WAI backscattering probe (d) Transmission mode. (Top) Capsule contents. Reprinted from Eliasson* et al. *(2008) [41] with permission from Elsevier.*

7.6 Applications

Major categories of process Raman applications include reaction monitoring, in-process quality checks, and mobile or field point measurements. Quality control laboratory applications often are converted to a continuous process monitoring approach, or could simply be viewed as part of a larger production process.

Many spectroscopic techniques are increasingly used for chemical reaction monitoring because of their speed and detailed chemical data [52]. Common motivations for reaction monitoring are to ensure the correct raw materials were used, monitor reaction progress to a targeted product yield, detect intermediates and side-product generation, facilitate investigations of mechanisms and component relationships, optimize production conditions, and ensure high product quality. For in-process quality checks, the goal usually is to verify processing and transfer steps are proceeding as expected. With process analyzers, data is provided more frequently than off-line testing. The information often is intended to trigger a quick, automatic corrective action by an efficient process control system, particularly during rapidly changing reactions or operating conditions.

In other cases, safety or cleanliness issues associated with sampling for off-line analysis provide the motivation for the installation of a process Raman system. In many cases, the material being sampled is hazardous, requiring personnel to dress in a chemical wetsuit every time a sample is needed. Once the off-line analysis is completed, the excess sample must be safely disposed. Sample heterogeneity also can make it difficult to obtain a representative sample. See Chapter 3 for more on sampling. Additionally, in some environments it may be difficult to obtain a sample due to the risk of contaminating the vessel or process, such as with a bioreactor or fermentation tank. By collecting spectra through a viewing point, the reactor sterilization is never compromised. Alternatively, the probe can be inserted into the reactor and sterilized in situ. The time and risks associated with physical sampling make it desirable to find an alternative and help justify a switch to on-line monitoring.

One of the issues that may arise at the start of proposed process analytical projects is whether a sensor, let alone a sophisticated sensor, is needed at all. Often, a real-time sensor is needed only during the development of the process to understand the dynamics of the system and to gain knowledge about the effect of different variables on the final process output. Once sufficiently robust operating conditions are established, continuous monitoring may be unnecessary. The sensor can be removed and is not needed again until substantial modifications are made to the process or new operating conditions are employed. In other cases, when there is enough instability or variability in the system or the consequences of a process upset are too great, the sensor should not be removed and replaced with a transfer function based on other variables. One should carefully consider the approach that brings the most value to the organization.

7.6.1 Acylation, alkylation, catalytic cracking, and transesterification

Li *et al*. monitored and modeled the most well-known acylation reaction, synthesis of aspirin; wavelets were applied during analysis to remove fluorescent backgrounds [53]. Gosling of UOP patented the use of Raman spectroscopy to control a solid catalyst alkylation process [54]. Based on the measured composition of the stream, specific process parameters are adjusted to bring the composition back to a targeted value. Multiple probes are placed in the process, for example near a feedstock inlet, in the reaction zone, or after the reactor. In the process discussed, the catalyst can be deactivated faster than an on-line gas chromatograph (GC) can produce a reading. This reason, coupled with obtaining greater process efficiency, is strong motivation for a fast on-line system like Raman. United States patent 5,684,580 by Cooper *et al*. of Ashland Inc. discusses monitoring the concentration of benzene and substituted aromatic hydrocarbons in multiple refinery process streams and using the results for process control [55]. Xylene isomers can be differentiated by Raman spectroscopy, making it technically preferable to NIR. The production of biodiesel fuel from a transesterification reaction of soybean oil and ethanol was monitored and correlated to results from several proton nuclear magnetic resonance (NMR) methods, though trace contaminants were detected only by GC [56].

7.6.2 Bioreactors

The industrial importance of bioprocesses continues to grow [57]. From mead to modern beer and wine, it is an art to control microorganisms to produce the desired product. As the biochemical and bioprocess fields progress, the science behind the art is becoming understood and playing an increasingly important role [58]. Raman spectroscopy was used to quantify the production of the carotenoid astaxanthin from the yeast *Phaffia rhodozyma* and from the algae *Haematococcus pluvialis* in a bioreactor [59]. The bioreactor was operated normally for more than 100 hours per reaction. The calibration curve covered total intracellular carotenoid concentrations from 1 to 45 mg/L. Though such concentrations are normally considered too low for Raman spectroscopy, carotenoids have extremely strong Raman signal. The precision was better than 5% and the

calibration model transferred across batches. The data suggest that models also could be built for other reaction components such as glucose. No attempt was made to spectrally differentiate between the various carotenoid species generated. The continuous measurements via in-line probe were easy, sufficiently fast, and accurate.

Picard *et al*. monitored *in situ* the production of ethanol from glucose and the yeast *Saccharomyces cerevisiae* as a function of high hydrostatic pressure [60]. A special reaction medium was used to overcome fluorescence problems rather than changing from a 514-nm laser. Ethanol production was both three times faster and 5% higher yielding when run at 10 MPa versus ambient conditions; rate and yield were lower at higher pressures. Though done on a nanoliter scale, the approach offers a promising means of monitoring metabolism and industrially important fermentations.

Lee *et al*. determined the concentration of glucose, acetate, formate, lactate, and phenylalanine simultaneously in two *Escherichia coli* bioreactions using Raman spectroscopy [57]. Spectra were collected through a viewing port added to the bioreactor, which enabled reactor sterilization and eliminated possible contamination from instrument contact. Bristol Myers Squibb presented promising results on the application of on-line Raman spectroscopy with PLS and principal component regression (PCR) models to predict glucose, lactate, glutamine, ammonia, total and viable cell counts, and protein titer in a Chinese hamster ovary (CHO) cell culture [61]. The work is intended to help make the process more efficient and to help monitor production.

Wood *et al*. have combined acoustic levitation and Raman spectroscopy with the intention of developing a field tool for environmental monitoring of algal blooms and nutrient availability [62]. Heraud discussed the most appropriate spectral preprocessing techniques for analysis of Raman spectra of single microalgal cells and developed a method to predict the nutrient status of those cells from *in vivo* spectra [63,64].

7.6.3 Blending

De Beer presented an excellent use of screening and optimization experimental design tools (i.e. DoE) with Raman spectroscopy to attempt to predict both the time to homogenously disperse ibuprofen in a suspension fluid and the ibuprofen concentration [65]. Though time to homogenization could not be modeled from the variables examined, substantial insights into the process still were gained. An on-line immersion probe was used to measure the powder blend uniformity of 1% azimilide dihydrochloride in a mixture of lactose, crospovidone, and magnesium stearate every 20 s [66]. The results were correlated with HPLC reference values on samples obtained using a sample thief. The impact of different probe designs on blend measurements was discussed above [33].

Wet granulation refers to changing the particle size distribution of a powder mixture by mixing in just enough liquid to bind the ingredients into the desired granule size. The API can change form, or phase, during this process, which might affect its efficacy. Wikström *et al*. used a large-area probe to determine the phase transformation's onset time and transformation profile as a function of mixing speed for anhydrous theophylline [33]. The transfer functions between tip speed and inverse onset time and between tip speed and transformation rate were established. The ability to investigate the relationship between processing parameter settings and the resulting blend quality and kinetics provides a powerful tool to both research and manufacturing operations.

7.6.4 Calcination

One of the now-classic examples of process Raman spectroscopy is monitoring the efficiency of converting the anatase form of titanium dioxide to the rutile form by calcining [67–69]. The forms have dramatically different spectra, enabling simple univariate band ratio models to be used. The uncertainty in the mean

estimate at the 95% confidence level was 0.1%, roughly four times more precise than the X-ray diffraction (XRD) reference method. Monsanto reported several analyzers installed globally for this application, each measuring multiple reactors via fiber optics of lengths up to 100 m. The units were described as reliable, valuable process control tools.

7.6.5 Catalysis

Raman spectroscopy is playing an increasingly important role in catalysis studies [70,71] It enables measurement of catalyst structure under different reaction conditions, helping to develop a molecular level structure-activity/selectivity model and improve understanding of catalyst poisoning mechanisms. Stencel notes that the scattering cross-sections of molecules adsorbed on noble metals are greatly enhanced, making it possible to measure submonolayer concentrations [72]. Many studies characterize the catalyst prior to and after use in an off-line mode [73]. Others directly monitor changes to the catalyst as a function of controlled process conditions, or employ several analytical techniques simultaneously to observe the catalytic reaction steps *in situ* [74–76]. In subsequent operando studies (real-time *in situ* measurements under real-life operating conditions), the catalyst's activity, selectivity, and structure and other reactions are monitored under actual catalytic operating conditions [77,78]. Finally, catalyst state, determined from Raman measurements, has been used as a process control variable, enabling dynamic response to feedstock composition changes and corrective actions to prevent catalyst deactivation [79].

The catalysts used for the oxidative dehydrogenation of ethane and propane are frequently studied, providing insights into appropriate operating conditions, the role of the catalyst supports, species present or absent under given conditions, kinetic parameters, and structure–activity/selectivity relationships [78,80–82]. Taylor *et al*. monitored the concentration of methane during its partial oxidation over platinum at 900–1100 °C in a stagnation-flow reactor [83]. Several improvements were made to the models after differences from the theoretical predictions were discovered. Liu *et al*. studied the differences in the partial methane oxidation over Rh/Al_2O_3 and Ru/Al_2O_3 catalysts pretreated at different temperatures [84]. The oxidation of isopropyl alcohol to acetone over a catalyst in a microfluidic reactor was monitored continuously under a variety of reaction conditions and reagent and catalyst concentrations [85]. The approach enabled approximately 48-hours worth of macroscale reaction to be completed and analyzed in 10 minutes, offering an enticing way to screen or optimize reactions.

Examples demonstrating a means to improve efficiency or lifetime have great industrial implications. Regeneration cycles were started and stopped automatically based on the measured concentration of coke on a Cr/Al_2O_3 catalyst during propane dehydrogenation [79]. Since regeneration burn off could operate at a lower temperature, longer catalyst lifetime was expected. The gas feed concentrations and regeneration temperatures were varied in a series of cycles to verify robustness. The inset in Figure 7.7 shows spectra with increasing coke concentrations during a propane dehydrogenation reaction cycle, while the upper and lower portions of the main figure illustrate the changes in coke level and propane conversion percentage over time, respectively, at different times and cycles. The oxidation of n-butane to maleic anhydride using a vanadium phosphorus oxide (VPO) catalyst was monitored using a similar approach in the same paper. The catalyst's activation level was monitored as a function of a series of alternating low and high reactant compositions. Tests in pilot-scale reactors are underway but are not yet published.

Bergwerff *et al*. used Raman microscopy to track the form and distribution of molybdenum complexes during the impregnation of $CoMo/Al_2O_3$ catalyst bodies [86]. The differences Han *et al*. found between off-line and *in situ* spectra of Mn_2O_3 nanocrystal catalysts used for methane combustion were suggestive of a reversible temperature-dependent phase change [75]. Raman spectroscopy was used to study the CrII species formed in one step of polyethylene production using the Cr/SiO_2 Phillips catalyst [87]. Careful laser wavelength selection enabled activation of the resonant and preresonant vibrational modes of surface CrO_2

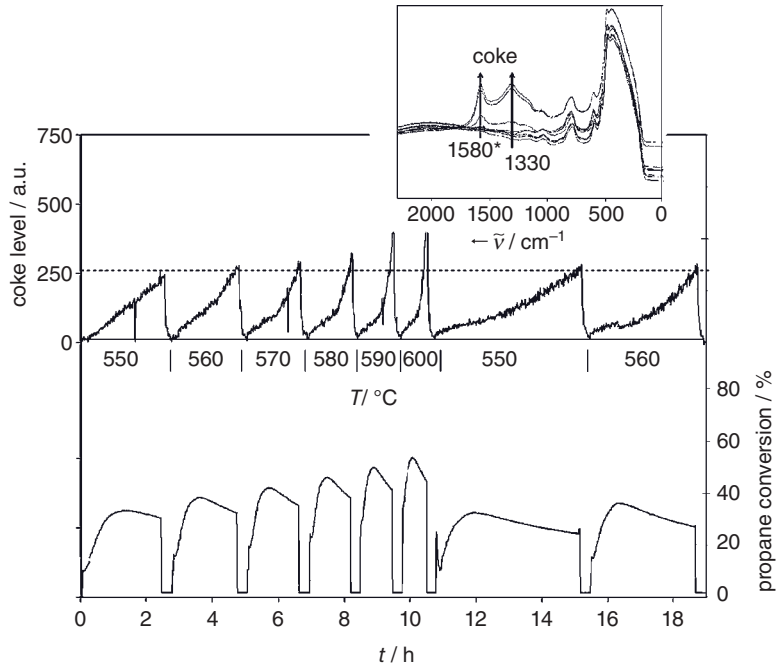


Figure 7.7 *Propane conversion percentage versus time (bottom), coke level versus temperature (top), and Raman spectra as coke level increases (inset). Coke level builds to a preset limit, after which a regeneration process is completed and propane production begins again. Reprinted from Bennici* et al. *(2007) [79]. Copyright Wiley-VCH Verlag GmbH & Co. KGaA. Reproduced with permission.*

species, enabling monitoring of the changes in the active centers in the presence of nitrogen and carbon monoxide. Damin *et al*. comment that the approach 'affords precious and direct information on the structure of the sites playing a role in catalysis' [87].

7.6.6 Chlorination

In the production of monochloroacetic acid from acetic acid and acetic anhydride, the concentration of the reaction intermediate, acetyl chloride, which also serves as a catalyst, must be controlled to minimize dichloroacetic acid by-product formation [88]. Recycled process streams are fed back into the process. Process safety was improved since manual sampling of the toxic, corrosive solution could be avoided. Several different laser wavelengths were tested during the project's feasibility stage to find one that reduced fluorescence problems to manageable levels. Samples were prepared and measured off-line to ensure sufficient variation in the calibration data set, particularly for those conditions outside safe operating limits. Subsequently, the spectra were treated using standard normal variate (SNV) to correct for intensity differences; it was not possible to use a band as an internal reference since everything was changing. These data treatments are common in NIR spectroscopy and are becoming more common for Raman spectra.

The use of Raman spectroscopy to monitor the reaction of chlorine gas and elemental phosphorous to produce phosphorous trichloride was first reported in 1991 and has been extensively described in patents and journals [67,89–93]. A FT-Raman spectrometer was selected to minimize fluorescence, but signal was

reduced unacceptably by the formation of a coating on the optical probe's window due to the instrument's high laser power. The problem was reduced by moving the probe to a flowing slipstream. Laser power could not be reduced without sacrificing too much measurement performance. With a switch to a high-throughput dispersive spectrometer with a shorter wavelength laser, the probe was returned to its original location. The scattering efficiency was improved, even with much lower laser power. Though the sample fluorescence that originally mandated the use of the FT instrument was observed at the end of the reaction, it did not interfere with process control [91].

Small details can impact how successfully an installation is viewed. The unforeseen window coatings and associated maintenance were not sufficient for the plant to stop using the analyzer, but they likely had reduced satisfaction with the technology. Several analysis routines were added to prevent reporting erroneous results, including recognizing substantial signal intensity drops from obscured process windows, probe misalignment, or excessive changes to the instrument or probe operating environment. Since it is simply not possible to predict and investigate every possible issue during feasibility studies, owners of any kind of on-line analyzer must be prepared to consider the need to make changes to the system over time.

7.6.7 Counterfeit pharmaceuticals

Counterfeit pharmaceuticals are an enormous, growing problem. They are stated to threaten public health and health systems, fund organized criminal networks, jeopardize the financial position of manufacturers, insurance providers, governments, and consumers, and create substantial potential liability and litigation [94,95]. Raman spectroscopy has been demonstrated to be an excellent option for counterfeit detection because of the lack of required sample preparation, ability to be configured in the field for fast data collection and interpretation with little operator training, and the ability to develop models for simple classification (genuine/counterfeit) or for identification and quantification of pharmaceutical and excipient ingredients [96]. Additionally, it may be possible for the analysis to be conducted through the packaging when required [97].

A number of chemometric tools have been employed for these classifications, including partial least squares – hierarchical cluster analysis (PLS-HCA) for Viagra tablets [98] and antimalarial artesunate tablets [99]. de Peinder *et al*. used partial least squares discriminant analysis (PLS-DA) models to distinguish genuine from counterfeit Lipitor tablets even when the real API was present [100]. The counterfeit samples also were found to have poorer API distribution than the genuine ones based on spectra collected in a cross pattern on the tablet.

Though extremely promising, the successful analysis of genuine and counterfeit pharmaceuticals by Raman spectroscopy is not always guaranteed. Ricci *et al*. found that Raman spectroscopy lacked sufficient sensitivity to detect some lower concentration components that were detected by mass spectrometry [101] Fluorescence of genuine artesunate samples, but not counterfeits, plagued a study using a portable Raman instrument [101], while the coatings on tablets of genuine and counterfeit Cialis tablets had to be removed to eliminate fluorescence in a study using a 633-nm laser [102]. As discussed previously, the use of transmission mode instead of the conventional backscattering mode eliminated fluorescence from many tablet coatings [8]. If the blister pack is metallized on one side, hindering the use of the transmission mode, the SORS mode still is possible [103]. Transmission mode, WAI probes, and SORS and its variants also offer the opportunity to probe through opaque containers. Additionally, these approaches sample a larger volume, tending to lead to simpler, more accurate models requiring fewer calibration samples [50]. It is important to remember that there are many ways for a drug to be considered counterfeit and Raman will not be suitable to address all of them. However, its many advantages make it likely to be routinely employed in the future as well-engineered systems help address its limitations.

7.6.8 Extrusion

Extrusion is associated with the polymer and food industries, though it also is used to manufacture pharmaceuticals. Marrow *et al.* of ExxonMobil have filed a patent application on correlating Raman spectra collected at the extruder with a polymer property, such as melt index, composition, or molecular weight, and using that in a feedback loop to control the operation [104]. Tumuluri *et al.* discuss the quantification of drug content in extruded films by Raman as an alternative to wet chemistry procedures like extractions [105]. Calibration curves were developed for two model drugs, clotrimazole and ketoprofen, both off-line and on-line. As anticipated, the off-line prediction models were simpler with lower SEPs than on-line models (approximately 0.7 wt% versus 1 wt%), but both approaches were successful. Normalization to a band from the polymer matrix was required. The 1655 cm^{-1} ketoprofen band shifted, perhaps caused by a crystalline transformation during extrusion, from a drug-polymer interaction, or another cause. The ease of use, responsiveness, robustness, and predictive ability of Raman was compared with NIR, MIR, and ultrasound sensors for composition measurements at or in an extruder using a variety of polymer systems [106–109]. Raman measurements were suitable in each case, with standard errors ranging from 0.15 wt% to 1 wt%, depending on the system. Fluctuating fluorescent backgrounds sometimes arise from degraded material, additives, stabilizers, dyes, and pigments in the mixtures. When present, these complicate analysis and remain hurdles to routine industrial use.

7.6.9 Forensics

More examples of forensic applications of Raman spectroscopy have been published recently. It has been used to identify individual crystals of drugs and excipients on paper currency [110], multilayer paint chips, inks, plastics [111], and fibers [112]. A study demonstrated the feasibility of quantifying acetaminophen in the presence of many excipient types [113]. Other studies seek to identify particulates, such as illicit or abused drugs, in fingerprints lifted at a crime scene [114,115].

7.6.10 Hydrogenation

Hydrogenation reactions are important to many industries, including petroleum, chemicals, and pharmaceuticals [116,117]. These reactions are highly exothermic and may involve unstable intermediates or side reactions, presenting considerable safety risks to manage. One risk mitigation approach is to monitor the reaction's progress, tracking the identity and concentration profile of all reaction components and making corrective adjustments to reaction conditions as necessary to maintain the reaction within safe limits.

Literature examples include one-, two-, and three-step hydrogenation reactions. In each case, the reactant and product concentrations were monitored, as were intermediates or side reaction products as necessary. Simple peak height or area measurements were sufficient to generate reaction profiles for the reduction of cyclohexene [116] and of 1-chloro-2-nitrobenzene [117]. However, for more spectrally complex systems, such as the reduction of carvone and of 2-(4-hydroxyphenyl) propionate, multivariate curve resolution (MCR) was required [117].

Wiss [117] and Tumuluri [116] both determined the reaction end point and detected undesired side products or excessive intermediate concentration increases from spectra. Tumuluri went a step further, calculating the reaction kinetics for different catalysts to determine the most efficient and effective one. Interestingly, the reaction with one of the catalysts followed a different mechanism than the others examined. In one reaction, a long induction period was noticed and traced to an unanticipated side reaction. The structure of the side product could not be definitively determined because the MCR data reduction was too noisy, but its presence was clearly detected. Additional experiments specifically designed to study that stage of the

reaction could be done if desired. Similarly in Wiss' studies, the product, 2-chloroaniline, is less Raman-active than the starting material, 2-(4-hydroxyphenyl) propionate, so the product concentration profile was noisier than the reactant profile. Hamminga had similar experiences with the hydrogenation of γ-butyrolactone, of 1-butanal, and of diethyl maleate [118]. It is important to remember that reaction profiles can be generated only for species detected spectroscopically. Selection rules and reference spectra of pure compounds can help determine quickly if there are suitable bands. If only one side of a reaction, reactants or products, is detectable, it may be necessary to measure with multiple techniques, such as ATR-FTIR and Raman, in order to generate profiles of all components.

7.6.11 Hydrolysis

Super or near-critical water is being studied to develop alternatives to environmentally hazardous organic solvents. Venardou *et al.* utilized Raman spectroscopy to monitor the hydrolysis of acetonitrile in near-critical water without a catalyst, and determined the rate constant, activation energy, impact of experimental parameters, and mechanism [119,120]. Widjaja *et al.* tracked the hydrolysis of acetic anhydride to form acetic acid in water and used BTEM to identify the pure components and their relative concentrations [121]. The advantage of this approach is that it does not use separate calibration experiments, but still enables identification of the reaction components, even minor, unknown species or interference signals, and generates relative concentration profiles. It may be possible to convert relative measurements into absolute concentrations with additional information.

7.6.12 Medical diagnostics

Raman's insensitivity to water makes it well suited for examination of biological samples and this research area is developing rapidly [122]. Mello is developing spectroscopic microorganism identification procedures for disease diagnosis to replace traditional, slow, labor- and chemical-intensive methods [123]. Raman spectroscopy is being actively pursued as a tool for disease diagnosis using tissue [124,125], blood serum [126], spinal fluid [127], and noninvasive glucose measurements right through the skin [128]. There is evidence that Raman-detectable biochemical changes occur in tissue before the physical changes used for histological diagnosis [129,130]. This suggests that Raman spectroscopy might someday be used to guide complete excision of diseased tissue during surgery. The high demands of these applications on the instrumentation and data handling are already driving innovations in the field, such as SORS, transmission Raman, and Raman tomography, that have implications far beyond medical applications [37,131–133].

7.6.13 Microwave-assisted organic synthesis

Microwave-assisted synthesis is attractive to researchers for many reasons, including speed, yields, and the potential for reduced solvent use. Raman monitoring offers a convenient way to elucidate the chemical mechanism while instantly, continuously monitoring reaction kinetics. This enables rapid, data-driven process optimizations without concerns about safely and accurately sampling out of a microwave vessel stopped mid-reaction. Pivonka and Empfield of AstraZeneca Pharmaceuticals describe the continuous acquisition of Raman spectra of an amine or Knoevenagel coupling reaction in a sealed microwave reaction vessel at elevated temperatures and pressures [134].

The Leadbeater group has monitored numerous systems, including a variety of Suzuki coupling, organometallic ligand-substitution, and esterification reactions [135–138]. Researchers took less than a day to run the esterification reaction for the first time, optimize it on the millimole scale, and scale up with an automated stop-flow apparatus, generating almost six moles product in under 2.25 hours [138]. The reaction rate, and

thus yield, of Suzuki coupling reactions of phenylboronic acid and different aryl halides depended on the aryl halide used [135]. Spectroscopic results disproved common belief that longer reaction times would increase yields, saving considerable research effort. As advocates of the approach, they have published a detailed procedure to easily interface commercially available instruments and test their performance [139].

Each of the many ways to measure the temperature inside a microwave heating chamber has limitations. It would be useful to know the temperature in each phase of a heterogeneous mixture, instead of the average temperature, since different materials respond differently to microwave heating. Vaucher *et al.* solved this problem by making interesting use of the temperature dependence of Raman spectra to deliberately turn their fiber-optic Raman probe into a specialized thermometer by monitoring the changes in the phonons of a silicon and diamond powder mixture as it was heated up to approximately 750 °C [140,141]. This approach enabled determination of the heating mechanisms occurring in the vessel: Si powder is heated from microwave absorption, but transfers heat to the diamond, which does not absorb microwaves. Other advantages include the ability to obtain desired physical properties of the material by controlling the microstructure [141]. The simultaneous measurement of temperature and chemistry offers a powerful tool for microwave-based studies.

7.6.14 Mobile or field uses

Mobile or field applications of Raman have grown enormously. A simple internet search will turn up numerous companies selling portable or handheld Raman systems, a category of instrument that practically did not exist several years ago. Often these application environments are more demanding than in manufacturing facilities. Early applications focused on tools to enable the first personnel responding to an emergency to identify a potentially hazardous material [142]. Security applications, including detecting explosives and biological threats, have continued to develop [143–145], as have material authentication applications, whether in a manufacturer's receiving area, at a customs or police facility, or in a military arena [41,146,147]. In one model security-screening example, a SORS probe was used to identify the contents of a white sunscreen bottle as table sugar right through the closed bottle [148]. Instrument prices vary considerably; in general, lower cost instruments do not offer quite as high a performance level as a full laboratory instrument, but that is not a significant issue for many applications [149]. The lower price enables many more researchers access to the technology, which enables new problems and applications to be investigated.

7.6.15 Natural products

Natural products, from plants and foods to rocks and minerals, are complicated systems, but their analysis by Raman spectroscopy is a growing area. Most examples come from quality control laboratories, motivated to replace current time-consuming sample preparation and analysis steps with a less labor-intensive, faster technique; but most authors anticipated the eventual application to process control. Often a method will be practiced in a trading house or customs facility to distinguish between items perceived to be of different qualities, and thus prices.

Raman spectroscopy was found suitable for use in manufacturing operations related to cacao seeds, from the initial evaluation to final extracts [150]. Rubayiza *et al.* used differences in the diterpene kahweol lipid content to distinguish between samples of Arabica and Robusta coffee and between the geographic origin within each type [151]. Edwards *et al.* also probed Raman's ability to classify the origin of ginseng roots since certain varieties are endangered and their trade is forbidden [152]. Other researchers have examined oils [153–155], honey [156], conjugated linoleic acid concentration in milk fat [157], peppercorns [158], alginates [159], Devil's claw roots and associated medicinal products [160], irradiated starch gels

[161], essential oils [162–164], distilled spirits [165] and carrots [166]. Strehle *et al.* used a mobile Raman microspectroscopy unit to assess the possibility of a field-deployed unit to determine the optimal harvest time by measuring fennel oil in the field [164]. Raman has shown promise for use as an in-process test for the degree of fatty acid unsaturation in salmon in several forms: oil, ground, and fillet [167]. This level affects the commercial value of the fish, but traditional GC tests have proven too slow for manufacturing operations.

7.6.16 Orientation, stress, or strain

Paradkar *et al.* used on-line Raman spectroscopy to measure the molecular orientation of polypropylene (PP) fibers and crystallinity of high-density polyethylene (HDPE) fibers at different distances from the spinneret on a pilot-scale melt spinning apparatus. Birefringence [168] and differential scanning calorimetry (DSC) [169] were the reference methods. The 841/809 cm^{-1} band ratio was used to monitor PP. The polyethylene studies used the integrated intensity of the 1418 cm^{-1} band normalized to the area of the 1295–1305 cm^{-1} bands. These bands have been shown to be insensitive to the crystallization state of the material and thus are a suitable built-in internal standard. The identification of appropriate analysis bands is critical and often is supported by many laboratory-based studies. Additionally, the equipment was slightly modified to eliminate crystallinity calculation errors from laser polarization. The biggest experimental problem faced in both studies was keeping the fiber from moving in and out of the focal plane, resulting in unacceptably long acquisition times. Paradkar *et al.* solved that problem in a subsequent study at Dow Chemical on propylene homopolymers by passing the fiber through a ceramic guide to constrain the fiber's movement [170]. The relationships between crystallinity onset and rate, fiber spinning speed, and throughput were explored for two resins with different melt indices and used to validate a model of stress-induced crystallization [171]. Fibers spun from cellulose with *N*-methylmorphiline *N*-oxide (NMMO) were deformed while Raman spectra were collected [172]. The results revealed significant amorphous fractions that limit the mechanical performance of the material. The authors suggest that processing changes would be necessary to realize performance improvements in the fibers.

Orientation and crystallization also are important issues for polymer films and sheets. Ogale *et al.* discuss the use of Raman spectra to probe the development of crystallinity in polyethylene blown films [173] and to measure the nonisothermal crystallization half-time during blown film extrusions of linear low-density polyethylene (LLDPE) and of isotactic PP [174]. Separate studies discuss the influence of processing parameters, such as the take-up ratio and bubble air pressure, on the rate of crystallization of LLDPE [175] and coextrusions of low-density polyethylene (LDPE) with PP [176]. A laboratory study of drawn poly(L-lactic acid) films found that the amorphous regions could be oriented only with draw ratios greater than 4.5 [177]. A laboratory study of HDPE prepared using different injection velocities found Raman could be correlated to orientation as measured by small angle X-ray scattering (SAXS) methods [178].

Raman spectroscopy is a powerful tool for probing orientation, stress, and strain. Galiotis *et al.* have written a review on the determination of stress and strain in composites and fibers using Raman spectroscopy [179]. Young *et al.* discuss the complexities of correct interpretation of molecular orientation information encoded in Raman spectra of polymers [180]. Caution and a suitable number of control studies are necessary to prevent faulty conclusions.

Wright of Advanced Micro Devices discusses the use of Raman microspectroscopy to measure the integrity of a film on semiconductor wafers during manufacture in US patent 6,509,201 and combined the results with other data for feed-forward process control [181]. Yield is improved by providing a tailored repair for each part. Hitachi has filed a Japanese patent application disclosing the use of Raman spectroscopy to determine the strain in silicon semiconductor substrates to aid manufacturing [182]. Raman spectroscopy has a well established place in the semiconductor industry for this and other applications [183].

7.6.17 Ozonolysis

Several pharmaceuticals are manufactured using an ozonolysis step, which can be readily tracked with Raman spectroscopy [184]. Pelletier *et al.* demonstrated continuous quantitative concentration measurement of *trans*-stilbene reacting to give benzaldehyde and α-methoxybenzyl hydroperoxide, a model reaction [184]. The calibration range was approximately 0–300 mM. The components' concentration changes are detected, even when due to the system's pressure changes during sample collection for off-line analysis. Raman monitoring provided more information than off-line HPLC analysis and did not interfere with the reaction's progress.

Multivariate calibrations are powerful tools, but the number and type of calibration samples required often is prohibitive. To overcome this problem, Pelletier employed a powerful but relatively uncommon tool, spectral stripping. This technique takes advantage of existing system knowledge to use spectra of fewer, more easily generated samples. More applications of this approach can be expected.

7.6.18 Polymerization

Francisco *et al.* at DuPont discuss monitoring and control of semibatch polymerizations using monomer mixtures, enabling both better process control and higher product quality [185]. In a laboratory study, Edwards *et al.* determined the rate constant of the group transfer, or living, polymerization of methyl methacrylate by monitoring the decrease in monomer carbon-carbon double bond signal relative to the solvent signal [186]. Since moisture can quench this reaction, it is appealing to avoid breaking the vacuum to sample the system, as might be required for several other analysis techniques. During the thermal purification of methacrylic acid and/or methacrylic esters, some of the material can polymerize inside the production equipment, leading to costly, time-consuming production outages for cleanup. BASF employees discuss the application of Raman spectroscopy, along with several other methods, to monitor the impurities in the monomer solution that lead to the problem and adjusting the plant's operating conditions as required to manage them [187].

In patents, Raman spectroscopy frequently is used to monitor and control polymerization processes. Battiste at Chevron Phillips describes its use in the manufacture of polyolefins, including in slurry systems, in several patents [188–190]. He discusses the use of low-resolution Raman instruments as a cost-effective option because the quantitative bands are sufficiently separated in the systems being monitored. Battiste also discusses the connection between manufacturing process controls, Raman spectra, and final product properties. Marrow *et al.* at ExxonMobil also describe monitoring continuous solution polymerization and subsequent processing steps [191], a fluidized bed [192, 193], slurry [194], or high pressure [195] reactors, or mixing devices, such as extruders [104]. In many cases, the focus is on the correlation between manufacturing conditions and final product properties, enabling the process to be fine-tuned to specific targets. This effectively is a step beyond monitoring a specific chemical functionality.

Emulsion polymerization is used for 10–15% of global polymer production, including such industrially important polymers as poly(acrylonitrile-butadiene-styrene) (ABS), polystyrene, poly(methyl methacrylate), and poly(vinyl acetate) [196]. These are made from aqueous solutions with high concentrations of suspended solids. The important components have unsaturated carbon–carbon double bonds. Raman spectroscopy is well-suited to address these challenges, though the heterogeneity of the mixture sometimes presents challenges. New sample interfaces, such as WAI and transmission mode, that have shown promise in pharmaceutical suspensions are anticipated to help here also.

Wenz and colleagues at Bayer Polymers Inc. describe the use of Raman spectroscopy to determine the reaction end point for the emulsion production of ABS graft copolymer [197]. Early termination reduces product yield and results in an extra product purification step; termination too late reduces product quality.

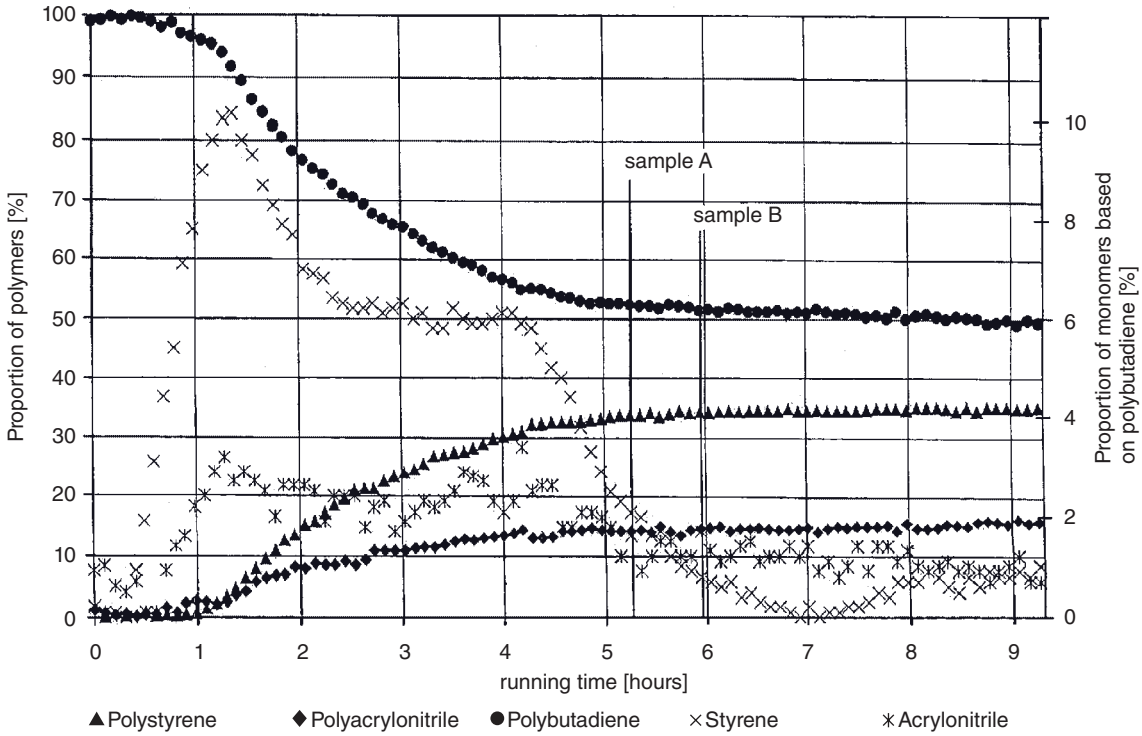


Figure 7.8 *Concentration profiles of polystyrene, polyacrylonitrile, polybutadiene, styrene, and acrylonitrile as a function of time since the reaction started. Reprinted from Wenz* et al. *(2001) [197].*

As Figure 7.8 illustrates, the reaction composition over time is not smooth and predictable, making it unlikely off-line analysis would ever be fast enough to ensure correct control decisions. Santos *et al.* monitored the conversion rate of the suspension polymerization of styrene. The conversion estimates did not use a reference method, but agreed well with gravimetric results until it was no longer possible to obtain representative samples for gravimetry [198,199]. Santos [198] and Reis [200] both present results showing a correlation between spectral features and particle size distribution (PSD), with the intent of developing an on-line PSD indicator to aid process control.

Bauer *et al.* describe the use of a noncontact probe coupled by fiber optics to an FT-Raman system to measure the percentage of dry extractibles and styrene monomer in a styrene/butadiene latex emulsion polymerization reaction using PLS models [201]. Elizalde *et al.* have examined the use of Raman spectroscopy to monitor the emulsion polymerization of *n*-butyl acrylate with methyl methacrylate under starved, or low monomer [202], and with high solids-content [203] conditions. In both cases, models could be built to predict multiple properties, including solids content, residual monomer, and cumulative copolymer composition. Another study compared reaction calorimetry and Raman spectroscopy for monitoring *n*-butyl acrylate/methyl methacrylate and for vinyl acetate/butyl acrylate, under conditions of normal and instantaneous conversion [204]. Both techniques performed well for normal conversion conditions and for overall conversion estimate, but Raman spectroscopy was better at estimating free monomer concentration and instantaneous conversion rate. However, the authors also point out that in certain situations, alternative techniques such as calorimetry can be cheaper, faster, and often easier to maintain accurate models for than Raman spectroscopy. In a subsequent article, Elizalde *et al.* found that updating calibration models after

instrument maintenance was better than building new models, analogous to many other process spectroscopies [205]. All instruments require service at some point, so it is an important reminder to consider the true requirements of an application and the availability of resources for recalibration before selecting a technique.

One of the challenges of building robust calibration models often is obtaining suitable, representative process samples that contain sufficient variation across and between the components of interest. Obtaining and establishing reference values for these samples often is one of the hardest parts of a process analytical technology project, regardless of the technique selected. It can be difficult to quench reaction samples fast enough to avoid introducing a bias between the spectroscopic and off-line reference values [206]. Reis *et al.* present two cases where synthetic samples were used successfully to build predictive models of monomer concentration during butyl acrylate and vinyl acetate homopolymerizations [207]. A separate paper compares prediction results using models built using process-generated versus synthetic samples [208]. Changes in solution heterogeneity as the reaction proceeds can hinder model development. Principal curves are demonstrated as being better at detecting relatively short periods of deviation from expected reaction profiles than T^2 and Q statistics [209]. A principal curve is a 'smooth curve through the middle of a dataset' and is a principal component if the curve is straight [210]. Medium heterogeneity also complicates spectral interpretation in the anionic dispersion polymerization of styrene and 1,3-butadiene. Jiang *et al.* apply several chemometrics approaches, including self modeling curve resolution (SMCR), parallel vector analysis (PVA), and window factor analysis (WFA), combined with mean-centered window principal component analysis (MCWPCA) [211]. Both PVA and WFA enabled accurate spectral and composition profiles to be extracted from the data.

It is interesting to consider the use of in situ Raman spectroscopy as a routine laboratory tool for research and development, as opposed to process monitoring. One of the scenarios developing is the extensive use of the tool in the laboratory to more fully characterize a system and its behavior under deviation conditions. With sufficient characterization, it may not be necessary to install a full process analyzer in the production facility. Simple sensors might be sufficient. The stereochemistry (*cis*-, *trans*-) of polymers produced from dicyclopentadiene (DCPD) using different catalysts was monitored in situ with FT-Raman [212]. This serves as an excellent tool during the product and process development processes to ensure that an optimal catalyst, for both yield and specificity, is selected. The production of sol–gel materials from silanes with epoxy and amine functionality was optimized based on Raman spectroscopic results, enabling control of solution viscosity [213]. Separate studies provided insights into the role of water and solvents in the reaction.

7.6.19 Polymer curing

On-line reaction monitoring of polymer curing chemistry helps scientists develop accurate transfer functions between a material's physical properties and its formulation and processing conditions. Scientists at Philips Research Laboratories have used Raman spectroscopy to help monitor the curing process of thin acrylic coatings used to convert spherical glass lenses into aspherical ones [214]. These are suitable for use as optical pick up lenses in Blu-ray DVD machines. The measurements were performed on free droplets of resin and on lenses in an actual lens production unit. The unanticipated conclusion of the experiments was that the mold holding the lens and coating deformed during polymerization to accommodate the volume reduction (shrinkage). It was expected that the mold would not move and that the reaction rate of a volume-constrained coating would be different than a free droplet.

Cole reviews the use of Raman spectroscopy to monitor the curing of different thermoset composite polymer systems, though he cautions that it is not suitable for certain materials [215]. Cruz *et al.* studied the effect of elevated pressure on the cure rate and the degree of cure of unsaturated polyester and other thermoset materials [216,217]. The approach worked well, though the experimental set up required addi-

tional temperature control modifications to better correlate with DSC results. In another case, the cure chemistry of polydicyclopentadiene was monitored as part of a feasibility study for real-time monitoring of reactive injection or rotational molding operations [218]. The synthesis of a phenol–formaldehyde resol resin was monitored through a reactor window as a function of a stepwise reaction temperature profile and in a series of reactions held at different temperatures [219]. The loss of reactants and gain of desired functional groups were readily monitored. Three samples then were selected and band ratios were used to follow the chemistry at four cure temperatures. Findings correlated with off-line NMR studies. Due to equipment availability, Musto *et al.* used quenched samples, rather than continuous monitoring, to generate reaction profiles of epoxy, anhydride, and ester groups during curing [220]. The results reveal both an induction period and suggest a side reaction that occurs at elevated cure temperatures or with skewed stoichiometric mixtures.

7.6.20 Polymorphs (crystal forms)

7.6.20.1 Identification and Quantification

There has been explosive growth in applications of Raman spectroscopy to the identification and quantification of polymorphs, or different crystal forms, in the pharmaceutical industry [221–227]. Form changes can affect effectiveness and safety so pharmaceutical manufacturers must understand how best to preserve the desired drug form through the manufacturing, distribution, and storage processes. The impact of problems can be reduced by examining material throughout the product development cycle. Sample sizes in such studies range from nanogram-sized crystals in HTS multiwell titer plates to full-scale production. The sampling technique selected will vary with the sample size.

Hilfiker *et al.* at Solvias used carbamazepine (CBZ) as a model compound to describe the use of Raman microscopy to characterize crystal forms, including during solvent evaporation experiments [228]. The spectra were processed into clusters by spectral similarity. The authors note that all published and several new crystal forms were identified during the study. Solvias' HTS uses a specific set of crystallization protocols that have tended to produce new polymorphs. Hilfiker notes that Raman microspectroscopy is 'an ideal analytical tool for high-throughput discrimination between crystal structures.' [229]. The ability to collect spectra directly and automatically in a microtiter plate with or without solvent and during evaporation is a major advantage over many other techniques.

Raman microspectroscopy is readily performed on multiple locations inside each well. As in other instances, the results might not be representative of the whole sample because of the small sample volume probed. Polarization effects can be pronounced, but may be mitigated by averaging the results from additional locations. An alternative is rotating the sample, but this usually is not practical for multiwell plates. Both options increase analysis time. Such problems appear to be minimized when handling bulk powders [222,223,230]. Several vendors sell systems preconfigured for automated analysis of microtiter plates and are typically integrated with optical microscopy.

The bottleneck in utilizing Raman shifted rapidly from data acquisition to data interpretation. Visual differentiation works well when polymorph spectra are dramatically different or when reference samples are available for comparison, but is poorly suited for automation, for spectrally similar polymorphs, or when the form was previously unknown [231]. Spectral match techniques, such as are used in spectral libraries, help with automation, but can have trouble when the reference library is too small. Easily automated clustering techniques, such as hierarchical cluster analysis (HCA) or PCA, group similar spectra and provide information on the degree of similarity within each group [223,230]. The techniques operate best on large data sets. As an alternative, researchers at Pfizer tested several different analysis of variance (ANOVA) techniques, along with descriptive statistics, to identify different polymorphs from measurements of Raman

peak positions and intensity alone [232]. In the classic trade-off between α- and β- statistical errors, all true polymorphs were correctly classified as such but some examples of the same polymorph were misidentified as different forms.

7.6.20.2 Polymorph Stability

After a potential drug has been identified, it is subjected to additional experiments, such as different solvents, temperatures, and pressures, to understand its specific behavior. Initial studies sought to identify the presence of different polymorphs, but the need to quantify the polymorphs naturally arose. Subsequent work tested the quantification models and probed how the polymorph concentration changes as a function of time or other variables. Kinetic or rate parameters often can be extracted. Studies were performed on the laboratory scale, such as under a microscope or in a bench crystallizer, up to large-scale manufacturing trials.

The stability of a suspension–emulsion of benzimidazole in oil was evaluated by quantifying the fractions of polymorphs A, B, and C as soon as prepared and after one year using a calibration model developed using pure polymorph samples [233]. The concentrations of mannitol in its amorphous form and each of three polymorphs were determined within 5% mass fraction, relative to salmon calcitonin [234]. The powders all were sized for delivery by respiration, less than 5 μm.

Examples of solvent-mediated transformation monitoring include the conversion of anhydrous citric acid to the monohydrate form in water [235,236], CBZ with water [237] and ethanol–water mixtures [238,239], and cocrystallization studies of CBZ, caffeine, and theophylline with water [240]. Raman spectroscopy was used to monitor the crystallization rate and solute and solvent concentrations as griseofulvin was removed from an acetone solution using supercritical CO_2 as an antisolvent [241]. Progesterone's crystallization profile was monitored as antisolvent was added [242].

In other cases, solvent-mediated transformations were functions of temperature. CBZ has been used in several studies [238,239,243,244]. Anquetil used different temperature profiles to selectively generate and quantify either of two forms in an HTS setup [243]. Raman spectra were collected *in situ* as CBZ samples were heated isothermally at a series of temperatures, converting them from form III to form I [244]. The data generated enabled a variety of kinetic models to be evaluated. Qu *et al.* used an immersion probe to understand the effect of temperature and solvent composition on the CBZ's transformation from anhydrous to dihydrate form [238,239]. The concentrations of each polymorphic form of flufenamic acid and of the ethanol–water solute were determined simultaneously as a function of crystallization time and temperature [245]. The transition temperature of flufenamic acid was estimated accurately from Raman data in another study [246]. *In situ* Raman spectroscopy was used to monitor continuously conversion of α-glycine to the γ-form in water and in NaCl solution at a variety of temperatures, stirring rates, and seed crystal sizes [247]. The transformation rate was faster in warm aqueous NaCl solution with a higher stirring rate and smaller seed crystals. The trend profiles for each form at two temperatures are shown in Figure 7.9 as an example. A logical extension of this study would be to develop a response surface model for these experimental variables and explore whether they interact significantly or respond nonlinearly. Without continuous in situ monitoring, such experiments would be too burdensome.

The solvent-mediated transformation of α-L-glutamic acid to the β-form was quantitatively monitored over time at a series of temperatures [248]. The calibration model was built using dry physical mixtures of the forms, but still successfully predicted composition in suspension samples. Cornel *et al.* monitored the solute concentration and the solvent-mediated solid-state transformation of L-glutamic acid simultaneously [249]. However, the authors note that multivariate analysis was required to achieve this. Additionally, they caution that it was necessary to experimentally evaluate the effect of solid composition, suspension density, solute concentration, particle size and distribution, particle shape, and temperature on the Raman spectra during calibration in order to have confidence in the quantitative results. This can be a substantial experi-

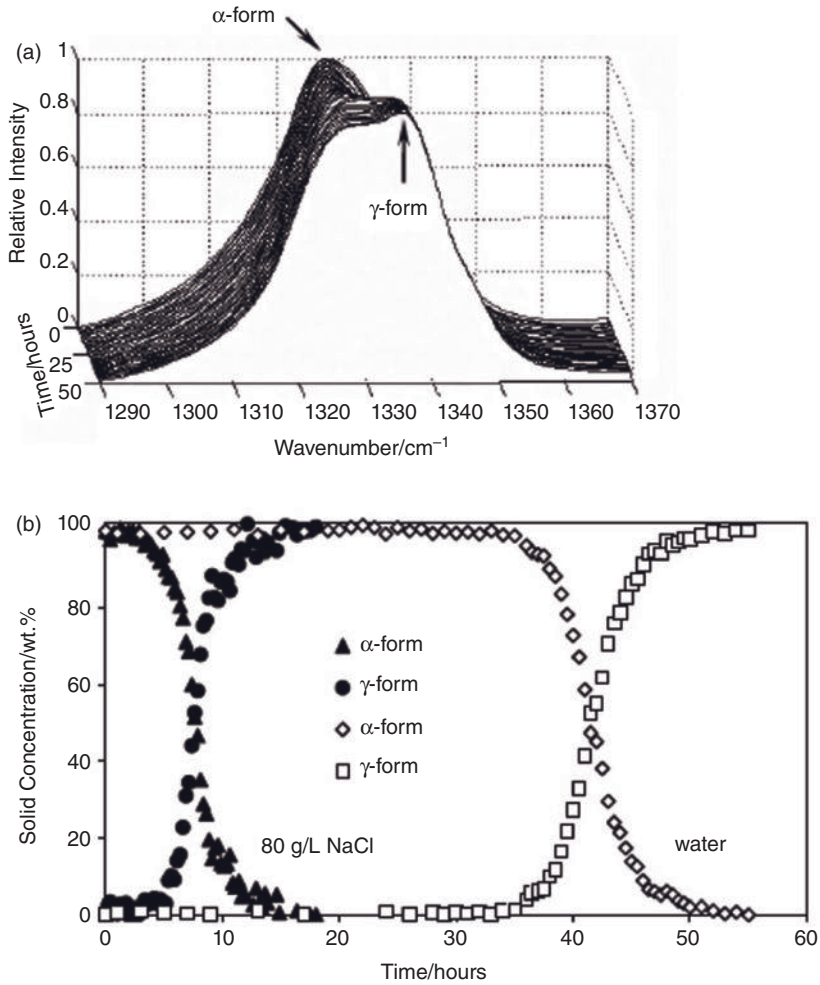


Figure 7.9 *Transformation of α-glycine to γ-glycine in aqueous NaCl solution at different temperatures. Reprinted from Yang* et al. *(2008) [247] with permission from John Wiley & Sons, Ltd.*

mental burden. However, if any of these factors are found to influence the results, it offers the possibility that those variables could end up being monitored inferentially from the Raman spectra.

Temperature-mediated transformations also occur. Ali *et al.* collected simultaneous Raman spectra and DSC curves of salmeterol xinafoate, detecting two forms [250]. Raman spectroscopy was used to follow the conversions of erythromycin A dihydrate as a function of temperature; conditions were identified where the Raman data provided more insights into the kinetics than powder XRD [251].

The transformation of famotidine from form B to form A was found to be controlled by pressure, temperature, and the interaction between those factors [252]. A subsequent study focused on identifying appropriate milling conditions to avoid conversion [253]. A study of milled Ti and C powders to form TiC found that Raman spectroscopy provided insights not available from XRD studies alone [254]. The successful complexation, or process-induced transformation, of nimodipine (NMD) and hydroxylpropyl-β-cyclodextrin

(HPbCD) during granulation was found to depend on the presence of sufficient ethanol; the results were related to those from a dissolution study [255].

7.6.20.3 Determining Appropriate Processing Conditions and Limits

Manufacturing operating conditions and control limits can be set based on the conditions required to convert, or transform, the compound from one polymorph to another. Scientists at the Merck Research Laboratories have used in situ Raman spectroscopy to follow the kinetics of polymorph conversion in a series of experiments designed to mimic manufacturing upset conditions [256]. At GlaxoSmithKline, the relationship between isolation temperature, cooling rate, and solvent ratio and the API's form and the induction time was used to gain insights into appropriate crystallization processing and alarm conditions and hold times during manufacturing [257]. The weak Raman signal from water enabled solvent–water slurries to be used successfully in the study. NIR and Raman spectroscopy each were used to quantify the conversion of theophylline monohydrate to the anhydrate form in a fluidized bed dryer as it occurred; NIR monitored the water loss and Raman tracked crystal structure changes [258]. Raman spectroscopy, NIR, and XRD were combined to provide a complete picture of changes in nitrofurantoin and theophylline during extrusion-spheronization pelletization [259].

Exposure to water, whether from aqueous granulation fluids or from exposure to humidity during manufacturing delays or storage, promoted unacceptable formation of an API's drug salt form [260]. Raman spectroscopy was used to monitor the pseudopolymorphic changes in theophylline under different storage conditions, including 0–100% relative humidity at ambient temperature for one week and a series of elevated temperature/time combinations [261]. The authors found that hydration occurs in a one-step random nucleation process, but dehydration occurs in two steps and involves the metastable form. The material is stable across a wide range of relative humidity, but converts within one day when those limits are exceeded. These types of studies are helpful in understanding how the material should be handled during processing, in designing the best packaging, and in determining shelf life.

It is important to remember that Raman's sensitivity makes many of these polymorph experiments possible, but it also can complicate the development of the calibration model and result in a burdensome number of calibration samples [227]. The results of many of the experiments were confirmed by comparing them to traditional methods. Future research hopefully will focus on means of reducing the calibration burden.

7.6.21 Product properties

In some cases, it is possible to predict the properties of finished products from their Raman spectra. Many of the patents related to polymers and Raman spectroscopy encompass the general idea of controlling the production process based on a calibration model correlating Raman spectra to specific polymer properties, such as product physical properties (melt index, density, viscosity, flexural and impact strength, yield stress, die swell, dye uptake, etc.) or composition (monomer or comonomer, catalyst, modifier, crosslinker, or additive concentration, etc.) [104,188–195,262,263]. Ito *et al.* developed a laboratory-based FT-Raman method to predict the viscosity of waterborne automotive paint emulsion samples as a first step towards on-line monitoring to predict final product performance, such as weather durability [264]. Other patent examples include the prediction of gasoline octane number and Reid vapor pressure [265,266]. In order to quickly and nondestructively measure the hole concentration in highly aluminum doped silicon carbide (SiC) bulk crystals, the width of the 965 cm^{-1} band, assigned as the longitudinal optical phonon–plasmon coupled mode, was calibrated against the charge carrier concentration [267]. The surface roughness of CdTe wafers was correlated to its Raman spectrum, but only when an 830-nm laser was used; data collected using 515- and 633-nm lasers were found to be uncorrelated [268].

7.6.22 Purification: distillation, filtration, drying

Distillations are perfect candidates for Raman process monitoring since many potential fluorescent or interfering species are reduced or removed. Dow Corning monitors the concentration of a chlorosilane distillation stream with Raman spectroscopy [269]. The use of on-line GC was discontinued because the 20–60 minute response time was too slow to detect important column dynamics. The GCs also required extensive maintenance due to the hydrochloric acid in the stream. The Raman system is used to predict the concentration of nine species. Many could not be detected or distinguished in the NIR. The Institut Français du Pétrole holds two relevant patents discussing the use of Raman spectroscopy to control the separation of aromatics by adsorption with simulated moving beds and of isomers, particularly aromatic hydrocarbons with 8–10 carbon atoms such as *para*-xylene and *ortho*-xylene [270,271].

The composition of a solid sucrose filter cake was monitored with a Raman immersion probe as it was washed with ethanol mother liquor to determine the best conditions to wash out the cyclohexane solvent impurity [272]. This is an excellent example of using Raman spectroscopy to rapidly understand a system and design appropriate operating conditions. Continuous monitoring of the commercial process might not be necessary if the process was sufficiently characterized during the development stage and relatively stable.

At Novartis, the chemical form of an API hydrate was monitored during a series of static and dynamic experiments investigating the wet drying process [273]. It was found unexpectedly that the final API form depended on both the relative humidity and solvent vapor concentrations for certain regimes. Additionally, the authors realized that the thickness of the wet cake may matter and that the proposed pan dryer type may not be the most efficient choice. Since these findings were made early, prior to manufacturing scale-up studies, it still would be possible to improve the manufacturing process. Walker *et al.* recorded Raman spectra using a remote Raman probe at a series of locations in a fluidized bed in order to generate three-dimensional maps of material density and particle concentration and composition [274]. Though this was a proof of concept study, the ability to monitor how the material composition changes as a function of location in the bed and over time could be a powerful tool for process improvement and optimization.

The hydration state of risedronate sodium was monitored continuously in a fluidized bed dryer and correlated to data on the physical stability of tablets made from the monitored material [275]. The final granulation moisture was found to affect the solid-state form, which in turn dictated the drug's physical stability over time. The process of freeze-drying mannitol was monitored continuously with in-line Raman and at-line NIR spectroscopies [276]. The thin polymer solvent coatings, such as poly(vinyl acetate) with toluene, methanol, benzene, and combinations of the solvents, were monitored as they dried to generate concentration/time profiles [277].

7.6.23 Thin films or coatings

Monitoring the results of chemical vapor depositions in situ is a common Raman application. Adem of Advanced Micro Devices, Inc. was granted a patent on the use of Raman spectroscopy to monitor the thickness, crystal grain size, and crystal orientation of polysilicon or other films as they are deposited on semiconductor wafers via low-pressure chemical vapor deposition (CVD) [278]. The spectra are acquired with a noncontact probe through a window in the loading door. When the thickness has achieved the targeted value, the deposition is stopped. If the crystal grain size or orientation is deemed unsuitable, the deposition temperature is adjusted accordingly. Scientists from Argonne National Laboratory and SuperPower Inc collaborated to use Raman spectroscopy to measure the composition of $YBa_2Cu_3O_{6+x}$ (YBCO) thin films on long metal tapes at the exit of a metal organic CVD (MOCVD) chamber [279]. Spectral differences could be seen in films before and after oxygen annealing and in film segments with misoriented YBCO grains.

The goal of the effort is to develop a control scheme that can identify defective tape and adjust process control parameters to return the tape to appropriate quality levels.

Mermoux *et al.* discuss the application of Raman spectroscopy to monitoring the growth of diamond films prepared under assorted CVD conditions [280]. The authors synchronized a pulsed laser and gateable detector in order to increase the Raman signal intensity enough to be detectable in the presence of the hydrogen emission spectrum from the plasma and blackbody radiation from the hot substrate. The signal enhancement was close to the duty cycle of the laser. Spectra were collected without interruption during 24 hours of deposition. The spectra enable simultaneous measurement of the deposition temperature, the development of nondiamond phases, and stress in the film. A previously measured spectrum of pure diamond at elevated temperatures and pressures was used as a reference to detect changes in key bands. Isotopically enriched ^{13}C methane was used to grow homoepitaxial diamond on a diamond anvil [281]. By measuring the 1281 cm^{-1} Raman band, present only in the ^{13}C diamond layer, the precise thickness of the deposited layer was determined. This enabled Qiu *et al.* to recognize that small quantities of nitrogen in the deposition gas more than double the growth rate [281].

The effect of different laser annealing parameters on an amorphous hydrogenated silicon thin film was probed *in situ* with Raman spectroscopy [282]. Scientists at Shell Solar established calibration curves to predict the S/Se-ratio, sulfur distribution, and the ratio of copper to the sum of indium and gallium content in thin films of $Cu(In,Ga)(S,Se)_2$ (CIGSSe) from Raman spectra [283]. The approach is intended to aid solar cell production and process development. Researchers at IMEC used Raman to monitor the oxidation of $Si_{0.75}Ge_{0.25}$ layers to assess the surface layer's homogeneity and confirmed the results with cross-sectional transmission electron microscopy (X-TEM) [284]. The elimination of water from the oxygen flow enabled more homogeneous, high quality surfaces to form. Huang *et al.* used the position and shape of the coupled A1 (LO) phonon as an indicator of the carrier concentration across a GaN wafer. The results correlated well with capacitance–voltage (CV) measurements and offer the possibility of a nondestructive manufacturing quality assurance tool [285].

The thickness of pharmaceutical tablet coatings was predicted using target factor analysis (TFA) applied to Raman spectra collected with a 532-nm laser, where the samples were photobleached in a controlled manner before spectra were acquired. The authors acknowledge numerous issues that limit the direct applicability of this approach to process control. These include potential damage or alteration of the samples from photobleaching, laser wavelength selection, and data acquisition time. However, most of the issues raised relate to the hardware selected for a particular implementation and do not diminish the demonstration [286].

7.7 Current State of Process Raman Spectroscopy

Raman spectroscopy was discovered over 75 years ago but has only been a viable process tool for 10–15 years. However, there has been an astounding increase in process Raman spectroscopy examples in the last five years. The United States Food and Drug Administration's (US FDA) endorsement of process analytical technology clearly set off an explosion of activity. Problems that sometimes sidelined Raman in the past, such as fluorescence or highly variable quantitative predictions from samples that were too small to be representative, are being re-examined and leading to new technology. In turn, that helps open, or perhaps reopen, new application areas. The availability of easy to use Raman instrumentation at many prices also helps with that.

However, process spectroscopy is, almost by definition, done to measure and control an industrial process. Most of the work is driven by business needs, such as improving profits or product quality. In competitive business environments, firms preserve every advantage possible by protecting valuable measurement systems

as trade secrets. Thus, firms often are reluctant to reveal process spectroscopy applications, whether successful or not. Often, companies will patent the work and will not publish in a scientific journal until after the patent is filed, if ever. Many applications, such as the classic titanium oxide monitoring paper [67–69], only are revealed years after implementation. Notable exceptions to this include the desire for positive publicity around improved safety or to direct the regulatory environment. Many pharmaceutical companies appear to be publishing Raman articles to help build industry and governmental understanding of the technique's value. Regardless of the studies' outcomes, their collective work to define Raman's role helps improve the technique for all users. As exciting as many of the recent examples are, it is important to remember that the current state of the art in the literature is quite likely far out of date or at least presents an incomplete picture.

References

1. M.J. Pelletier, Quantitative Raman spectroscopy, *Appl. Spectrosc.*, **57**, 20A–42A (2003).
2. J.M. Shaver, Chemometrics for Raman spectroscopy, in *Handbook of Raman Spectroscopy: From the Research Laboratory to the Process Line*, I.R. Lewis and H.G.M. Edwards (Eds), Practical Spectroscopy Series 28, Marcel Dekker, New York, 2001.
3. P.J. Treado and M.D. Morris, Infrared and Raman spectroscopic imaging, in *Microscopic and Spectroscopic Imaging of the Chemical State*, M.D. Morris (Ed), Practical Spectroscopy Series 16, Marcel Dekker, New York, 1993.
4. J.R. Ferraro, K. Nakamoto and C.W. Brown, *Introductory Raman Spectroscopy*, Academic Press, Boston, 2003.
5. W.E. Smith, Surface enhanced resonance Raman scattering, in *Pharmaceutical Applications of Raman Spectroscopy*, S. Šašic (Ed), Technology for the Pharmaceutical Industry Series, John Wiley & Sons, Ltd, Hoboken, 2008.
6. S. Šašic (Ed), *Pharmaceutical Applications of Raman Spectroscopy*, Technology for the Pharmaceutical Industry Series, John Wiley & Sons Ltd, Hoboken, 2008.
7. K. Nakamoto, *Infrared and Raman Spectra of Inorganic and Coordination Compounds. Part A: Theory and Applications in Inorganic Chemistry*, Wiley-Interscience, New York, 1997.
8. P. Matousek and A.W. Parker, Non-invasive probing of pharmaceutical capsules using transmission Raman spectroscopy, *J. Raman Spectrosc.*, **38**, 563–567 (2007).
9. M.J. Pelletier, Raman instrumentation, in *Analytical Applications of Raman Spectroscopy*, M.J. Pelletier (Ed), Blackwell Science, Malden, 1999.
10. H.G.M. Edwards, Raman spectroscopy in the characterization of archaeological materials, in *Handbook of Raman Spectroscopy: From the Research Laboratory to the Process Line*, I.R. Lewis and H.G.M. Edwards (Eds), Practical Spectroscopy Series 28, Marcel Dekker, New York, 2001.
11. F.R. Perez, Applications of IR and Raman spectroscopy to the study of medieval pigments, in *Handbook of Raman Spectroscopy: From the Research Laboratory to the Process Line,* I.R. Lewis and H.G.M. Edwards (Eds), Practical Spectroscopy 28, Marcel Dekker, New York, 2001.
12. P. Vandenabeele, H.G.M. Edwards, and L. Moens, A decade of Raman spectroscopy in art and archaeology, *Chem. Rev.*, **107**, 675–686 (2007).
13. R. Szostak and S. Mazurek, A quantitative analysis of liquid hydrocarbon mixtures on the basis of FT-Raman spectra registered under unstable conditions, *J. Mol. Struct.*, **704**, 235–245 (2004).
14. P.J. Aarnoutse and J.A. Westerhuis, Quantitative Raman reaction monitoring using the solvent as internal standard, *Anal. Chem.*, **77**, 1228–1236 (2005).
15. M.J. Pelletier, New developments in analytical instruments: Raman scattering emission spectrophotometers, in *Process/Industrial Instruments and Controls Handbook*, G.K. McMillan and D.M. Considine (Eds), McGraw-Hill, New York, 1999.
16. R. Mukhopadhyay, Raman flexes its muscles, *Anal. Chem.*, **79**, 3265–3270 (2007).
17. L.A. Nafie, Theory of Raman scattering, in *Handbook of Raman Spectroscopy: From the Research Laboratory to the Process Line*, I.R. Lewis and H.G.M. Edwards (Eds), Practical Spectroscopy Series 28, Marcel Dekker, New York, 2001.
18. D.J. Gardiner, Introduction to Raman scattering, in *Practical Raman Spectroscopy*, D.J. Gardiner and P.R. Graves (Eds), Springer-Verlag, New York, 1989.

19. J.P. Chauvel, W.W. Henslee and L.A. Melton, Teaching process analytical chemistry, *Anal. Chem.*, **74**, 381A–384A (2002).
20. J.B. Slater, J.M. Tedesco, R.C. Fairchild and I.R. Lewis, Raman spectrometry and its adaptation to the industrial revolution, in *Handbook of Raman Spectroscopy: From the Research Laboratory to the Process Line*, I.R. Lewis and H.G.M. Edwards (Eds), Practical Spectroscopy Series 28, Marcel Dekker, New York, 2001.
21. B. Chase, A new generation of Raman instrumentation, *Appl. Spectrosc.*, **48**, 14A–19A (1994).
22. OSHA Safety & Health Topics: Radiation – Laser Hazards, available http://www.osha.gov/SLTC/laserhazards/index.html accessed 16 November 2009.
23. Laser Institute of America, available www.laserinstitute.org accessed 16 November 2009.
24. S.D. Harvey, T.J. Peters and B.W. Wright, Safety considerations for sample analysis using a near-infrared (785 nm) Raman laser source, *Appl. Spectrosc.*, **57**, 580–587 (2003).
25. R.J. Rockwell, Jr. and J. Parkinson, State and local government laser safety requirements, *J. Laser Appl.*, **11**, 225–231 (1999).
26. American National Standards Institute (ANSI), American national standard for safe use of lasers, in Z136 Laser safety standards series. Z136.1 (2007).
27. US Occupational Safety and Health Agency (OSHA), *Guidelines for laser safety and hazard assessment in STD 01-05-001-PUB 8-1.7* (1991). Available http://www.osha.gov/pls/oshaweb/owadisp.show_document?p_table=DIRECTIVES&p_id=1705 accessed 16 November 2009.
28. P. Kotidis, R. Crocrombe and W. Atia, Optical, tunable filter-based micro-instrumentation for industrial process control. *Abstracts of Papers, Federation of Analytical Chemistry and Spectroscopy Societies (FACSS)*, Orlando, FL, USA, October 19–23, 2003, Federation of Analytical Chemistry and Spectroscopy Societies, Santa Fe, NM USA, 2002.
29. P.J. Gemperline, Rugged spectroscopic calibration for process control, *Chemom. Intell. Lab. Syst.*, **39**, 29–40 (1997).
30. D.E. Battey, J.B. Slater, R. Wludyka, H. Owen, D.M. Pallister and M.D. Morris, Axial transmissive f/1.8 imaging Raman spectrograph with volume-phase holographic filter and grating, *Appl. Spectrosc.*, **47**, 1913–1919 (1993).
31. N. Everall, H. Owen and J. Slater, Performance analysis of an integrated process Raman analyzer using a multiplexed transmission holographic grating, CCD detection, and confocal fiberoptic sampling, *Appl. Spectrosc.*, **49**, 610–615 (1995).
32. P. Niemelä, J. Sumen and J. Suhonen, Dual-laser CCD-Raman spectrometer applied to on-line measurement of paper coating paste, *Proc. SPIE-Int. Soc. Opt. Eng.*, **5826**, 406–411 (2005).
33. H. Wikström, I.R. Lewis and L.S. Taylor, Comparison of sampling techniques for in-line monitoring using Raman spectroscopy, *Appl. Spectrosc.*, **59**, 934–941 (2005).
34. M. Kim, H. Chung, Y. Woo and M. Kemper, New reliable Raman collection system using the wide area illumination (WAI) scheme combined with the synchronous intensity correction standard for the analysis of pharmaceutical tablets, *Anal. Chim. Acta*, **579**, 209–216 (2006).
35. D.L. Gerrard and H.J. Bowley, Instrumentation for Raman spectroscopy, in *Practical Raman Spectroscopy*, D.J. Gardiner and P.R. Graves (Eds), Springer-Verlag, New York, 1989.
36. S.C. Park, M. Kim, J. Noh, H. Chung, Y. Woo, J. Lee and M.S. Kemper, Reliable and fast quantitative analysis of active ingredient in pharmaceutical suspension using Raman spectroscopy, *Anal. Chim. Acta*, **593**, 46–53 (2007).
37. P. Matousek, I.P. Clark, E.R.C. Draper, *et al.*, Subsurface probing in diffusely scattering media using spatially offset Raman spectroscopy, *Appl. Spectrosc.*, **59**, 393–400 (2005).
38. P. Matousek, Inverse spatially offset Raman spectroscopy for deep noninvasive probing of turbid media, *Appl. Spectrosc.*, **60**, 1341–1347 (2006).
39. D.J. Gardiner and P.R. Graves (Eds), *Practical Raman Spectroscopy*, Springer-Verlag, New York, 1989.
40. P. Matousek and A.W. Parker, Bulk Raman analysis of pharmaceutical tablets, *Appl. Spectrosc.*, **60**, 1353–1357 (2006).
41. C. Eliasson, N.A. Macleod, L.C. Jayes, F.C. Clarke, S.V. Hammond, M.R. Smith, and P. Matousek, Non-invasive quantitative assessment of the content of pharmaceutical capsules using transmission Raman spectroscopy, *J. Pharm. Biomed. Anal.*, **47**, 221–229 (2008).
42. J. Johansson, A. Sparén, O. Svensson, S. Folestad, and M. Claybourn, Quantitative transmission Raman spectroscopy of pharmaceutical tablets and capsules, *Appl. Spectrosc.*, **61**, 1211–1218 (2007).
43. C.K. Markopoulou, J.E. Koundourellis, M.G. Orkoula, and C.G. Kontoyannis, Quantitative nondestructive methods for the determination of ticlopidine in tablets using reflectance near-infrared and Fourier transform Raman spectroscopy, *Appl. Spectrosc.*, **62**, 251–257 (2008).

44. M.G. Orkoula and C.G. Kontoyannis, Non-destructive quantitative analysis of risperidone in film-coated tablets, *J. Pharm. Biomed. Anal.*, **47**, 631–635 (2008).
45. T.R.M. De Beer, W.R.G. Baeyens, A. Vermeire, D. Broes, J.P. Remon, and C. Vervaet, Raman spectroscopic method for the determination of medroxyprogesterone acetate in a pharmaceutical suspension: Validation of quantifying abilities, uncertainty assessment and comparison with the high performance liquid chromatography reference method, *Anal. Chim. Acta*, **589**, 192–199 (2007).
46. M. López-Sánchez, M.J. Ruedas-Rama, A. Ruiz-Medina, A. Molina-Díaz and M.J. Ayora-Cañada, Pharmaceutical powders analysis using FT-Raman spectrometry: Simultaneous determination of sulfathiazole and sulfanilamide, *Talanta*, **74**, 1603–1607 (2008).
47. J. Johansson, S. Pettersson and S. Folestad, Characterization of different laser irradiation methods for quantitative Raman tablet assessment, *J. Pharm. Biomed. Anal.*, **39**, 510–516 (2005).
48. M. Kim, H. Chung, Y. Woo and M.S. Kemper, A new non-invasive, quantitative Raman technique for the determination of an active ingredient in pharmaceutical liquids by direct measurement through a plastic bottle, *Anal. Chim. Acta*, **587**, 200–207 (2007).
49. T.R.M. De Beer, W.R.G. Baeyens, Y. Vander Heyden, J.P. Remon, C. Vervaet and F. Verpoort, Influence of particle size on the quantitative determination of salicylic acid in a pharmaceutical ointment using FT-Raman spectroscopy, *Eur. J. Pharm. Sci.*, **30**, 229–235 (2007).
50. N.A. Macleod and P. Matousek, Emerging non-invasive Raman methods in process control and forensic applications, *Pharm. Res.*, **25**, 2205–2215 (2008).
51. E. Widjaja and M. Garland, Use of Raman microscopy and band-target entropy minimization analysis to identify dyes in a commercial stamp. Implications for authentication and counterfeit detection, *Anal. Chem.*, **80**, 729–733 (2008).
52. S.P. Gurden, J.A. Westerhuis and A.K. Smilde, Monitoring of batch processes using spectroscopy, *AIChE J.*, **48**, 2283–2297 (2002).
53. W.H. Li, G.W. Lu, L. Huang, H. Gao and W.M. Du, Aspirin synthesis on-line measurement by Raman spectroscopy, *Acta Phys.-Chim. Sin.*, **19**, 105–108 (2003).
54. UOP LLC USA, Control of solid catalyst alkylation process using Raman spectroscopy. Inventor: C.D. Gosling. 04 Mar 2003. 10 pp. (incl. 1 fig.). Appl. 02 Aug 2000. Int. Cl. G01N 35/08. US Patent Specification 6,528,316
55. Ashland Inc., Hydrocarbon analysis and control by Raman spectroscopy. Inventors: J.B. Cooper, J.P.E. Flecher, and W.T. Welch. 4 Nov 1997. 36 pp. (incl. 20 fig.). Appl. 1 May 1995. Int. Cl. G01N 21/65. US Patent Specification 5,684,580
56. G.F. Ghesti, J. Lemos de Macedo, I.S. Resck, J.A. Dias and S.C.L. Dias, FT-Raman spectroscopy quantification of biodiesel in a progressive soybean oil transesterification reaction and its correlation with ^{1}H NMR spectroscopy methods, *Energ. Fuel*, **21**, 2475–2480 (2007).
57. H.L.T. Lee, P. Boccazzi, N. Gorret, R.J. Ram and A.J. Sinskey, In situ bioprocess monitoring of *Escherichia coli* bioreactions using Raman spectroscopy, *Vib. Spectrosc.*, **35**, 131–137 (2004).
58. U. von Stockar, S. Valentinotti, I. Marison, C. Cannizzaro and C. Herwig, Know-how and know-why in biochemical engineering, *Biotechnol. Adv.*, **21**, 417–430 (2003).
59. C. Cannizzaro, M. Rhiel, I. Marison and U. von Stockar, On-line monitoring of *Phaffia rhodozyma* fed-batch process with in situ dispersive Raman spectroscopy, *Biotechnol. Bioeng.*, **83**, 668–680 (2003).
60. A. Picard, I. Daniel, G. Montagnac and P. Oger, In situ monitoring by quantitative Raman spectroscopy of alcoholic fermentation by *Saccharomyces cerevisiae* under high pressure, *Extremophiles*, **11**, 445–452 (2007).
61. S. Varma, S. Sakhamuri, M. Hausladen, Z. Li and S.S. Lee, Simultaneous determination of multiple components in CHO cell culture using Raman spectroscopy. *Abstracts of Papers, 234th ACS National Meeting, Boston, MA, USA, August 19–23, 2007*, American Chemical Society, Washington, DC, BIOT-256.
62. B.R. Wood, P. Heraud, S. Stojkovic, D. Morrison, J. Beardall and D. McNaughton, A portable Raman acoustic levitation spectroscopic system for the identification and environmental monitoring of algal cells, *Anal. Chem.*, **77**, 4955–4961 (2005).
63. P. Heraud, B.R. Wood, J. Beardall and D. McNaughton, Effects of pre-processing of Raman spectra on *in vivo* classification of nutrient status of microalgal cells, *J. Chemom.*, **20**, 193–197 (2006).
64. P. Heraud, J. Beardall, D. McNaughton and B.R. Wood, *In vivo* prediction of the nutrient status of individual microalgal cells using Raman microspectroscopy, *FEMS Microbiol. Lett.*, **275**, 24–30 (2007).
65. T.R.M. De Beer, W.R.G. Baeyens, J. Ouyang, C. Vervaet and J.P. Remon, Raman spectroscopy as a process analytical technology tool for the understanding and the quantitative in-line monitoring of the homogenization process of a pharmaceutical suspension, *Analyst*, **131**, 1137–1144 (2006).

66. D.S. Hausman, R.T. Cambron and A. Sakr, Application of Raman spectroscopy for on-line monitoring of low dose blend uniformity, *Int. J. Pharm.*, **298**, 80–90 (2005).
67. M.A. Leugers and E.D. Lipp, Raman spectroscopy in chemical process analysis, in *Spectroscopy in Process Analysis*, J.M. Chalmers (Ed), Sheffield Academic Press, Sheffield, 2000.
68. I.M. Clegg, N.J. Everall, B. King, H. Melvin and C. Norton, On-line analysis using Raman spectroscopy for process control during the manufacture of titanium dioxide, *Appl. Spectrosc.*, **55**, 1138–1150 (2001).
69. Tioxide Group Services Ltd, Calcination of titanium dioxide. Inventors: J.P. Besson, P.W.B. King, T.A. Wilkins, M.C. McIvor and N.J. Everall. 10 Jan 2001. 6 pp. (incl. 0 fig.). Appl. 9 Sep 1996. Int. Cl. C09C 1/36; C01G 23/08; G01J 3/44. EP Patent Specification 0 767 222 B1
70. M.A. Bañares, In situ Raman spectroscopy, in *In Situ Characterization of Catalysts*, B.M. Weckhuysen (Ed), American Scientific Publishers, Stevenson Ranch, 2004.
71. I.E. Wachs, Raman spectroscopy of catalysts, in *Handbook of Raman Spectroscopy: From the Research Laboratory to the Process Line*, I.R. Lewis and H.G.M. Edwards (Eds), Practical Spectroscopy Series 28, Marcel Dekker, New York, 2001.
72. J.M. Stencel, *Raman Spectroscopy for Catalysis*, Van Nostrand Reinhold, New York, 1990.
73. R. Nakamura and H. Frei, Visible light-driven water oxidation by Ir oxide clusters coupled to single Cr centers in mesoporous silica, *J. Am. Chem. Soc.*, **128**, 10668–10669 (2006).
74. A. Brüeckner, G. Scholz, D. Heidemann, M. Schneider, D. Herein, U. Bentrup and M. Kant, Structural evolution of $H_4PVMo_{11}O_{40}.xH_2O$ during calcination and isobutane oxidation: new insights into vanadium sites by a comprehensive in situ approach, *J. Catal.*, **245**, 369–380 (2007).
75. Y.-F. Han, L. Chen, K. Ramesh, E. Widjaja, S. Chilukoti, I.K. Surjami and J. Chen, Kinetic and spectroscopic study of methane combustion over α-Mn_2O_3 nanocrystal catalysts, *J. Catal.*, **253**, 261–268 (2008).
76. C. Zhao and I.E. Wachs, Selective oxidation of propylene over model supported V_2O_5 catalysts: Influence of surface vanadia coverage and oxide support, *J. Catal.*, **257**, 181–189 (2008).
77. M.O. Guerrero-Pérez and M.A. Bañares, From conventional in situ to operando studies in Raman spectroscopy, *Catal. Today*, **113**, 48–57 (2006).
78. A. Brüeckner and E. Kondratenko, Simultaneous operando EPR/UV-vis/laser-Raman spectroscopy – a powerful tool for monitoring transition metal oxide catalysts during reaction, *Catal. Today*, **113**, 16–24 (2006).
79. S.M. Bennici, B.M. Vogelaar, T.A. Nijhuis and B.M. Weckhuysen, Real-time control of a catalytic solid in a fixed-bed reactor based on in situ spectroscopy, *Angew. Chem., Int. Ed.*, **46**, 5412–5416 (2007).
80. C.L. Pieck, M.A. Bañares and J.L.G. Fierro, Propane oxidative dehydrogenation on VO_x/ZrO_2 catalysts, *J. Catal.*, **224**, 1–7 (2004).
81. A. Christodoulakis, E. Heracleous, A.A. Lemonidou and S. Boghosian, An operando Raman study of structure and reactivity of alumina-supported molybdenum oxide catalysts for the oxidative dehydrogenation of ethane, *J. Catal.*, **242**, 16–25 (2006).
82. T.V.M. Rao and G. Deo, Kinetic parameter analysis for propane ODH: V_2O_5/Al_2O_3 and MoO_3/Al_2O_3 catalysts, *AIChE J.*, **53**, 1538–1549 (2007).
83. J.D. Taylor, M.D. Allendorf, A.H. McDaniel and S.F. Rice, In situ diagnostics and modeling of methane catalytic partial oxidation on Pt in a stagnation-flow reactor, *Ind. Eng. Chem. Res.*, **42**, 6559–6566 (2003).
84. Y. Liu, F.-Y. Huang, J.-M. Li, *et al.*, *In situ* Raman study on the partial oxidation of methane to synthesis gas over Rh/Al_2O_3 and Ru/Al_2O_3 catalysts, *J. Catal.*, **256**, 192–203 (2008).
85. S.-A. Leung, R.F. Winkle, R.C.R. Wootton and A.J. deMello, A method for rapid reaction optimisation in continuous-flow microfluidic reactors using online Raman spectroscopic detection, *Analyst*, **130**, 46–51 (2005).
86. J.A. Bergwerff, L.G.A. van de Water, T. Visser, P. de Peinder, B.R.G. Leliveld, K.P. de Jong and B.M. Weckhuysen, Spatially resolved Raman and UV-visible-NIR spectroscopy on the preparation of supported catalyst bodies: controlling the formation of $H_2PMo_{11}CoO_{40}^{5-}$ inside Al_2O_3 pellets during impregnation, *Chem. Eur. J.*, **11**, 4591–4601 (2005).
87. A. Damin, F. Bonino, S. Bordiga, E. Groppo, C. Lamberti and A. Zecchina, Vibrational properties of Cr^{II} centers on reduced Phillips catalysts highlighted by resonant Raman spectroscopy, *ChemPhysChem*, **7**, 342–344 (2006).
88. L.G. Weyer, D. Yee, G.I. Pearson and J. Wade, Installation of a multiplexed Raman system in a chemical process, *Appl. Spectrosc.*, **58**, 1500–1505 (2004).
89. Monsanto Company, Apparatus for preparing phosphorus trichloride. Inventors: M.S. Feld, D.O. Fisher, J.F. Freeman, G.J. Gervasio, M.A. Hochwalt, L.F. Laskowski and E.E. Thomsen. 9 Nov 1993. 8 pp. (incl. 5 fig.). Appl. 16 Dec 1991. Int. Cl. G01N 21/00. US Patent Specification 5,260,026

90. Monsanto Company, Process for preparing phosphorus trichloride. Inventors: M.S. Feld, D.O. Fisher, J.J. Freeman, G.J. Gervasio, M.A. Hochwalt, L.F. Laskowski and E.E. Thomsen. 10 May 1994. 8 pp. (incl. 5 fig.). Appl. 28 Sep 1992. Int. Cl. C01B 25/10. US Patent Specification 5,310,529
91. I.R. Lewis, Process Raman spectroscopy, in *Handbook of Raman Spectroscopy: From the Research Laboratory to the Process Line*, I.R. Lewis and H.G.M. Edwards (Eds), Practical Spectroscopy Series 28, Marcel Dekker, New York, 2001.
92. J.J. Freeman, D.O. Fisher and G.J. Gervasio, FT-Raman on-line analysis of PCl_3 reactor material, *Appl. Spectrosc.*, **47**, 1115–1122 (1993).
93. G.J. Gervasio and M.J. Pelletier, On-line Raman analysis of PCl_3 reactor material, *Journal of Process Analytical Chemistry*, **III**, 7–11 (1997).
94. A.K. Deisingh, Pharmaceutical counterfeiting, *Analyst*, **130**, 271–279 (2005).
95. F.M. Fernandez, M.D. Green and P.N. Newton, Prevalence and detection of counterfeit pharmaceuticals: a mini review, *Ind. Eng. Chem. Res.*, **47**, 585–590 (2008).
96. M.R. Witkowski, The use of Raman spectroscopy in the detection of counterfeit and adulterated pharmaceutical products, *Am. Pharm. Rev.*, **8**, 56.(2005).
97. C. Ricci, C. Eliasson, N.A. MacLeod, P.N. Newton, P. Matousek and S.G. Kazarian, Characterization of genuine and fake artesunate anti-malarial tablets using Fourier transform infrared imaging and spatially offset Raman spectroscopy through blister packs, *Anal. Bioanal. Chem.*, **389**, 1525–1532 (2007).
98. M. de Veij, A. Deneckere, P. Vandenabeele, D. de Kaste and L. Moens, Detection of counterfeit Viagra® with Raman spectroscopy, *J. Pharm. Biomed. Anal.*, **46**, 303–309 (2008).
99. M. de Veij, P. Vandenabeele, K.A. Hall, *et al.*, Fast detection and identification of counterfeit antimalarial tablets by Raman spectroscopy, *J. Raman Spectrosc.*, **38**, 181–187 (2007).
100. P. de Peinder, M.J. Vredenbregt, T. Visser and D. de Kaste, Detection of Lipitor® counterfeits: A comparison of NIR and Raman spectroscopy in combination with chemometrics, *J. Pharm. Biomed. Anal.*, **47**, 688–694 (2008).
101. C. Ricci, L. Nyadong, F. Yang, F.M. Fernandez, C.D. Brown, P.N. Newton and S.G. Kazarian, Assessment of hand-held Raman instrumentation for *in situ* screening for potentially counterfeit artesunate antimalarial tablets by FT-Raman spectroscopy and direct ionization mass spectrometry, *Anal. Chim. Acta*, **623**, 178–186 (2008).
102. S. Trefi, C. Routaboul, S. Hamieh, V. Gilard, M. Malet-Martino and R. Martino, Analysis of illegally manufactured formulations of tadalafil (Cialis®) by ^{1}H NMR, 2D DOSY ^{1}H NMR and Raman spectroscopy, *J. Pharm. Biomed. Anal.*, **47**, 103–113 (2008).
103. C. Eliasson and P. Matousek, Noninvasive authentication of pharmaceutical products through packaging using spatially offset Raman spectroscopy, *Anal. Chem.*, **79**, 1696–1701 (2007).
104. ExxonMobil Chemical Company, On-line properties analysis of a molten polymer by Raman spectroscopy for control of a mixing device. Inventors: D.G. Marrow, D.A. Yahn, and T.R. Veariel. 20 pp. (incl. 10 fig.). Appl. 22 Jul 2005. Int. Cl. G01J 3/44; G01N 21/65. US Patent Application Publication 2007/0019190 A1
105. V.S. Tumuluri, M.S. Kemper, I.R. Lewis, S. Prodduturi, S. Majumdar, B.A. Avery and M.A. Repka, Off-line and on-line measurements of drug-loaded hot-melt extruded films using Raman spectroscopy, *Int. J. Pharm.*, **357**, 77–84 (2008).
106. P.D. Coates, S.E. Barnes, M.G. Sibley, E.C. Brown, H.G.M. Edwards and I.J. Scowen, In-process vibrational spectroscopy and ultrasound measurements in polymer melt extrusion, *Polymer*, **44**, 5937–5949 (2003).
107. S.E. Barnes, E.C. Brown, M.G. Sibley, H.G.M. Edwards and P.D. Coates, Vibrational spectroscopic and ultrasound analysis for the in-process monitoring of poly(ethylene vinyl acetate) copolymer composition during melt extrusion, *Analyst*, **130**, 286–292 (2005).
108. S.E. Barnes, E.C. Brown, M.G. Sibley, H.G.M. Edwards, I.J. Scowen and P.D. Coates, Vibrational spectroscopic and ultrasound analysis for in-process characterization of high-density polyethylene/polypropylene blends during melt extrusion, *Appl. Spectrosc.*, **59**, 611–619 (2005).
109. I. Alig, D. Fischer, D. Lellinger and B. Steinhoff, Combination of NIR, Raman, ultrasonic and dielectric spectroscopy for in-line monitoring of the extrusion process, *Macromol. Symp.*, **230**, 51–58 (2005).
110. K.Y. Noonan, M. Beshire, J. Darnell and K.A. Frederick, Qualitative and quantitative analysis of illicit drug mixtures on paper currency using Raman microspectroscopy, *Appl. Spectrosc.*, **59**, 1493–1497 (2005).
111. J. Zieba-Palus, R. Borusiewicz and M. Kunicki, PRAXIS – combined μ-Raman and μ-XRF spectrometers in the examination of forensic samples, *Forensic Sci. Int.*, **175**, 1–10 (2008).
112. J. Thomas, P. Buzzini, G. Massonnet, B. Reedy and C. Roux, Raman spectroscopy and the forensic analysis of black/grey and blue cotton fibres, *Forensic Sci. Int.*, **152**, 189–197 (2005).

113. M.-L. O'Connell, T. Howley, A.G. Ryder, M.N. Leger and M.G. Madden, Classification of a target analyte in solid mixtures using principal component analysis, support vector machines, and Raman spectroscopy, *Proc. SPIE-Int. Soc. Opt. Eng.*, **5826**, 340–350 (2005).
114. J.S. Day, H.G.M. Edwards, S.A. Dobrowski and A.M. Voice, The detection of drugs of abuse in fingerprints using Raman spectroscopy II: Cyanoacrylate-fumed fingerprints, *Spectrochim. Acta, Part A*, **60A**, 1725–1730 (2004).
115. M.J. West and M.J. Went, The spectroscopic detection of exogenous material in fingerprints after development with powders and recovery with adhesive lifters, *Forensic Sci. Int.*, **174**, 1–5 (2008).
116. V.S. Tumuluri, M.S. Kemper, A. Sheri, S.-R. Choi, I.R. Lewis, M.A. Avery and B.A. Avery, Use of Raman spectroscopy to characterize hydrogenation reactions, *Org. Process Res. Dev.*, **10**, 927–933 (2006).
117. J. Wiss and A. Zilian, Online spectroscopic investigations (FTIR/Raman) of industrial reactions: synthesis of tributyltin azide and hydrogenation of chloronitrobenzene, *Org. Process Res. Dev.*, **7**, 1059–1066 (2003).
118. G.M. Hamminga, G. Mul and J.A. Moulijn, Applicability of fiber-optic-based Raman probes for on-line reaction monitoring of high-pressure catalytic hydrogenation reactions, *Appl. Spectrosc.*, **61**, 470–478 (2007).
119. E. Venardou, E. Garcia-Verdugo, S.J. Barlow, Y.E. Gorbaty and M. Poliakoff, On-line monitoring of the hydrolysis of acetonitrile in near-critical water using Raman spectroscopy, *Vib. Spectrosc.*, **35**, 103–109 (2004).
120. E. Venardou, E. Garcia-Verdugo, S.J. Barlow, Y.E. Gorbaty and M. Poliakoff, On-line monitoring of the hydrolysis of acetonitrile in near-critical water using Raman spectroscopy. [Erratum to document cited in CA141:139935], *Vib. Spectrosc.*, **37**, 237 (2005).
121. E. Widjaja, Y.Y. Tan and M. Garland, Application of band-target entropy minimization to on-line Raman monitoring of an organic synthesis. An example of new technology for process analytical technology, *Org. Process Res. Dev.*, **11**, 98–103 (2007).
122. J. Griffiths, Raman spectroscopy for medical diagnosis, *Anal. Chem.*, **79**, 3975–3978 (2007).
123. C. Mello, E. Severi, L. Coelho, A. Marangoni, C. Dezuane, E. Ricci, D. Ribeiro and R.J. Poppi, Two-dimensional low resolution Raman spectroscopy applied to fast discrimination of microorganisms that cause pharyngitis: a whole-organism fingerprinting approach, *J. Mol. Struct.*, **883–884**, 61–65 (2008).
124. C. Krafft and V. Sergo, Biomedical applications of Raman and infrared spectroscopy to diagnose tissues, *Spectroscopy*, **20**, 195–218 (2006).
125. S.-Y. Lin, M.-J. Li and W.-T. Cheng, FT-IR and Raman vibrational microspectroscopies used for spectral biodiagnosis of human tissues, *Spectroscopy*, **21**, 1–30 (2007).
126. L.N. Rocha-Osornio, J.L. Pichardo-Molina, O. Barbosa-Garcia, C. Frausto-Reyes, C. Araujo-Andrade, R. Huerta-Franco and G. Gutiérrez-Juárez, Chemometric techniques on the analysis of Raman spectra of serum blood samples of breast cancer patients, *Proc. SPIE-Int. Soc. Opt. Eng.*, **6864**, 686416/1–686416/7 (2008).
127. I.H. Chou, M. Benford, H.T. Beier, *et al.*, Nanofluidic biosensing for β-amyloid detection using surface enhanced Raman spectroscopy, *Nano Lett.*, **8**, 1729–1735 (2008).
128. Massachusetts Institute of Technology, Raman spectroscopy for non-invasive glucose measurements. Inventors: M. Hunter, A.M. Enejder, T. Scecina, M.S. Feld and W.-C. Shih. 44 pp. (incl. 16 fig.). Appl. 27 Apr 2006. Int. Cl. A61B 5/00. US Patent Application Publication 2007/0060806 A1
129. R.E. Kast, G.K. Serhatkulu, A. Cao, *et al.*, Raman spectroscopy can differentiate malignant tumors from normal breast tissue and detect early neoplastic changes in a mouse model, *Biopolymers*, **89**, 235–241 (2008).
130. S. Koljenovic, T.C.B. Schut, R. Wolthuis, *et al.*, Raman spectroscopic characterization of porcine brain tissue using a single fiber-optic probe, *Anal. Chem.*, **79**, 557–564 (2007).
131. P. Matousek, Deep non-invasive Raman spectroscopy of living tissue and powders, *Chem. Soc. Rev.*, **36**, 1292–1304 (2007).
132. M.V. Schulmerich, S. Srinivasan, J. Kreider, J., *et al.*, Raman tomography of tissue phantoms and bone tissue, *Proc. SPIE-Int. Soc. Opt. Eng.*, **6853**, 68530V-68530V-7 (2008).
133. ChemImage Corporation, Method for Raman computer tomography imaging spectroscopy. Inventors: J.S. Maier and P.J. Treado. 9 pp. (incl. 5 fig.). Appl. 20 Jan 2006. Int. Cl. G01J 3/44; G01N 21/65. US Patent Application Publication 2006/0158645 A1
134. D.E. Pivonka and J.R. Empfield, Real-time *in situ* Raman analysis of microwave-assisted organic reactions, *Appl. Spectrosc.*, **58**, 41–46 (2004).
135. N.E. Leadbeater and R.J. Smith, Real-time monitoring of microwave-promoted Suzuki coupling reactions using in situ Raman spectroscopy, *Org. Lett.*, **8**, 4589–4591 (2006).
136. T.M. Barnard and N.E. Leadbeater, Real-time monitoring of microwave-promoted organometallic ligand-substitution reactions using *in situ* Raman spectroscopy, *Chem. Commun.*, 3615–3616 (2006).
137. N.E. Leadbeater and R.J. Smith, *In situ* Raman spectroscopy as a probe for the effect of power on microwave-promoted Suzuki coupling reactions, *Org. Biomol. Chem.*, **5**, 2770–2774 (2007).

138. N.E. Leadbeater, R.J. Smith and T.M. Barnard, Using *in situ* Raman monitoring as a tool for rapid optimisation and scale-up of microwave-promoted organic synthesis: Esterification as an example, *Org. Biomol. Chem.*, **5**, 822–825 (2007).
139. N.E. Leadbeater and J.R. Schmink, Use of Raman spectroscopy as a tool for *in situ* monitoring of microwave-promoted reactions, *Nat. Protoc.*, **3**, 1–7 (2008).
140. S. Vaucher, J.-M. Catala-Civera, A. Sarua, J. Pomeroy and M. Kuball, Phase selectivity of microwave heating evidenced by Raman spectroscopy, *J. Appl. Phys.*, **99**, 113505/1–113505/5 (2006).
141. S. Vaucher, P. Unifantowicz, C. Ricard, *et al.*, On-line tools for microscopic and macroscopic monitoring of microwave processing, *Physica B*, **398**, 191–195 (2007).
142. Ahura Scientific, available http://www.ahurascientific.com accessed 16 November 2009.
143. J.D. Santillán, C.D. Brown and W. Jalenak, Advances in Raman spectroscopy for explosive identification in aviation security, *Proc. SPIE-Int. Soc. Opt. Eng.*, **6540**, 65400P/1–65400P/8 (2007).
144. C. Eliasson, N.A. Macleod and P. Matousek, Noninvasive detection of concealed liquid explosives using Raman spectroscopy, *Anal. Chem.*, **79**, 8185–8189 (2007).
145. C.W. Wright, S.D. Harvey and B.W. Wright, Detection of hazardous chemicals using field-portable Raman spectroscopy, *Proc. SPIE-Int. Soc. Opt. Eng.*, **5048**, 107–118 (2003).
146. R.L. Green and C.D. Brown, Raw-material authentication using a handheld Raman spectrometer, *Pharm. Technol.*, **32**, 148–162 (2008).
147. W. Smith and S. Farquharson, A portable fuel analyzer, *Proc. SPIE-Int. Soc. Opt. Eng.*, **6377**, 63770E/1–63770E/8 (2006).
148. C. Eliasson, N.A. Macleod and P. Matousek, Non-invasive detection of powders concealed within diffusely scattering plastic containers, *Vib. Spectrosc.*, **48**, 8–11 (2008).
149. A. Kudelski, Analytical applications of Raman spectroscopy, *Talanta*, **76**, 1–8 (2008).
150. H.G.M. Edwards, S.E.J. Villar, L.F.C. de Oliveira and M. Le Hyaric, Analytical Raman spectroscopic study of cacao seeds and their chemical extracts, *Anal. Chim. Acta*, **538**, 175–180 (2005).
151. A.B. Rubayiza and M. Meurens, Chemical discrimination of arabica and robusta coffees by Fourier transform Raman spectroscopy, *J. Agric. Food Chem.*, **53**, 4654–4659 (2005).
152. H.G.M. Edwards, T. Munshi and K. Page, Analytical discrimination between sources of ginseng using Raman spectroscopy, *Anal. Bioanal. Chem.*, **389**, 2203–2215 (2007).
153. H. Schulz and M. Baranska, Identification and quantification of valuable plant substances by IR and Raman spectroscopy, *Vib. Spectrosc.*, **43**, 13–25 (2007).
154. M. Baranska, H. Schulz, A. Walter, P. Rösch, R. Quilitzsch, G. Lösing and J. Popp, Investigation of eucalyptus essential oil by using vibrational spectroscopy methods, *Vib. Spectrosc.*, **42**, 341–345 (2006).
155. M. Baranska, H. Schulz, S. Reitzenstein, *et al.*, Vibrational spectroscopic studies to acquire a quality control method of *Eucalyptus* essential oils, *Biopolymers*, **78**, 237–248 (2005).
156. A.N. Batsoulis, N.G. Siatis, A.C. Kimbaris, *et al.*, FT-Raman spectroscopic simultaneous determination of fructose and glucose in honey, *J. Agric. Food Chem.*, **53**, 207–210 (2005).
157. M. Meurens, V. Baeten, S.H. Yan, E. Mignolet and Y. Larondelle, Determination of the conjugated linoleic acids in cow's milk fat by Fourier transform Raman spectroscopy, *J. Agric. Food Chem.*, **53**, 5831–5835 (2005).
158. H. Schulz, M. Baranska, R. Quilitzsch, W. Schütze and G. Lösing, Characterization of peppercorn, pepper oil, and pepper oleoresin by vibrational spectroscopy methods, *J. Agric. Food Chem.*, **53**, 3358–3363 (2005).
159. T. Salomonsen, H.M. Jensen, D. Stenbæk and S.B. Engelsen, Chemometric prediction of alginate monomer compositions: A comparative spectroscopic study using IR, Raman, NIR, and NMR, *Carbohydr. Polym.*, **72**, 730–739 (2008).
160. M. Baranska, H. Schulz, R. Siuda, *et al.*, Quality control of *Harpagophytum procumbens* and its related phytopharmaceutical products by means of NIR-FT-Raman spectroscopy, *Biopolymers*, **77**, 1–8 (2005).
161. R. Kizil and J. Irudayaraj, Discrimination of irradiated starch gels Using FT-Raman spectroscopy and chemometrics, *J. Agric. Food Chem.*, **54**, 13–18 (2006).
162. H. Schulz, G. Özkan, M. Baranska, H. Krüger and M. Özcan, Characterization of essential oil plants from Turkey by IR and Raman spectroscopy, *Vib. Spectrosc.*, **39**, 249–256 (2005).
163. H. Schulz, M. Baranska, H.-H. Belz, P. Rösch, M.A. Strehle and J. Popp, Chemotaxonomic characterization of essential oil plants by vibrational spectroscopy measurements, *Vib. Spectrosc.*, **35**, 81–86 (2004).
164. M.A. Strehle, P. Rösch, M. Baranska, H. Schulz and J. Popp, On the way to a quality control of the essential oil of fennel by means of Raman spectroscopy, *Biopolymers*, **77**, 44–52 (2005).
165. A. Nose, T. Hamasaki, M. Hojo, R. Kato, K. Uehara and T. Ueda, Hydrogen bonding in alcoholic beverages (distilled spirits) and water-ethanol mixtures, *J. Agric. Food Chem.*, **53**, 7074–7081 (2005).

166. M. Baranska, H. Schulz, R. Baranski, T. Nothnagel and L.P. Christensen, In situ simultaneous analysis of polyacetylenes, carotenoids and polysaccharides in carrot roots, *J. Agric. Food Chem.*, **53**, 6565–6571 (2005).
167. N.K. Afseth, J.P. Wold and V.H. Segtnan, The potential of Raman spectroscopy for characterization of the fatty acid unsaturation of salmon, *Anal. Chim. Acta*, **572**, 85–92 (2006).
168. R.P. Paradkar, S.S. Sakhalkar, X. He and M.S. Ellison, On-line estimation of molecular orientation in polypropylene fibers using polarized Raman spectroscopy, *Appl. Spectrosc.*, **55**, 534–539 (2001).
169. R.P. Paradkar, S.S. Sakhalkar, X. He and M.S. Ellison, Estimating crystallinity in high density polyethylene fibers using online Raman spectroscopy, *J. Appl. Polym. Sci.*, **88**, 545–549 (2003).
170. R.P. Paradkar, R.M. Patel, E. Knickerbocker and A. Doufas, Raman spectroscopy for spinline crystallinity measurements. I. Experimental studies, *J. Appl. Polym. Sci.*, **109**, 3413–3420 (2008).
171. R.M. Patel, A.K. Doufas and R.P. Paradkar, Raman spectroscopy for spinline crystallinity measurements. II. Validation of fundamental fiber-spinning models, *J. Appl. Polym. Sci.*, **109**, 3398–3412 (2008).
172. K. Kong and S.J. Eichhorn, Crystalline and amorphous deformation of process-controlled cellulose-II fibres, *Polymer*, **46**, 6380–6390 (2005).
173. G. Giriprasath and A.A. Ogale, Online orientation during polyethylene blown film extrusion using polarized Raman spectroscopy. *Proceedings, 65th Annual Technical Conference – Society of Plastics Engineers (ANTEC), Cincinnati OH, USA, May 6–11, 2007, Society of Plastics Engineers, Brookfield, CT, 1601–1605*.
174. S.S. Cherukupalli, S.E. Gottlieb and A.A. Ogale, Real-time Raman spectroscopic measurement of crystallization kinetics and its effect on the morphology and properties of polyolefin blown films, *J. Appl. Polym. Sci.*, **98**, 1740–1747 (2005).
175. S.S. Cherukupalli and A.A. Ogale, Online measurements of crystallinity using Raman spectroscopy during blown film extrusion of a linear low-density polyethylene, *Polym. Eng. Sci.*, **44**, 1484–1490 (2004).
176. G. Giriprasath and A.A. Ogale, Processing–structure relationship in bilayer polyolefin blown films. *Proceedings, 64th Annual Technical Conference – Society of Plastics Engineers (ANTEC), Charlotte, NC, USA, May 7–11, 2006, Society of Plastics Engineers, Brookfield, CT, 830–834*.
177. M. Tanaka and R.J. Young, Molecular orientation distributions in uniaxially oriented poly(L-lactic acid) films determined by polarized Raman spectroscopy, *Macromolecules*, **39**, 3312–3321 (2006).
178. T. Gough, S.E. Barnes, H.G.M. Edwards and P.D. Coates, Application of Raman spectroscopy and X-ray scattering for in-line analysis of polymer orientation during processing. *Proceedings, 63rd Annual Technical Conference – Society of Plastics Engineers (ANTEC), Boston, MA, USA, May 1–5, 2005, Society of Plastics Engineers, Brookfield, CT, 3655–3659*.
179. C. Galiotis and J. Parthenios, Stress/strain measurements in fibers and composites using Raman spectroscopy, in *Vibrational Spectroscopy of Biological and Polymeric Materials*, V.G. Gregoriou and M.S. Braiman (Eds), CRC Press, Boca Raton, 2006.
180. M. Tanaka and R.J. Young, Review: Polarised Raman spectroscopy for the study of molecular orientation distributions in polymers, *J. Mater. Sci.*, **41**, 963–991 (2006).
181. Advanced Micro Devices Inc., *Method and apparatus for monitoring wafer stress*. Inventor: M.I. Wright. 21 Jan 2003. 15 pp. (incl. 7 fig.). Appl. 11 Apr 2001. Int. Cl. H01L 21/66; G01R 31/26. US Patent Specificiation 6,509,201 B1
182. Hitachi High Technologies Corporation, *Device and method for measurement of strain in silicon substrates for semiconductor devices (stress measurement device for semiconductor device)*. Inventors: S. Suzuki, Y. Nakano, T. Konno and K. Umemura. 14 pp. (incl. 17 fig.). Appl. 11 Nov 2005. Int. Cl. H01L21/66; G01N21/65. JP Patent Application Publication 2007-134531 (Applic. 2005–326834)
183. I.D. Wolf, Semiconductors, in *Analytical Applications of Raman Spectroscopy*, M.J. Pelletier (Ed), Blackwell Science Ltd, Malden, 1999.
184. M.J. Pelletier, M.L. Fabilli and B. Moon, On-line analysis of a continuous-flow ozonolysis reaction using Raman spectroscopy, *Appl. Spectrosc.*, **61**, 1107–1115 (2007).
185. T.W. Francisco, S. Yembrick and K.W. Leffew, Semi-batch polymerization analysis by multi-point Raman spectroscopy, *PAT – J. Process Anal. Technol.*, **3**, 13–19 (2006).
186. H.G.M. Edwards, K.S. Johal and A.F. Johnson, FT-Raman spectroscopic monitoring of the group-transfer polymerisation of methyl methacrylate, *Vib. Spectrosc.*, **41**, 160–169 (2006).
187. BASF Aktiengesellschaft, Determination of the amount of polymer deposited from (meth)acrylic acid and/or (meth)acrylic esters. Inventors: F. Hofer, S. Haremza, G. Wagenblast, V. Schliephake, U. Jager, J. Schroder, H. Keller, C. Dieleman and R. Diehl. 11 pp. (incl. 4 fig.). Appl. 2 Sept 2003. Int. Cl. C08F 2/00. US Patent Application Publication 2004/0082737 A1

188. Chevron Phillips Chemical Company LP, Raman spectrometry for monitoring and control of slurry processes for polymerizing olefins. Inventor: D.R. Battiste. 20 Apr 2004. 11 pp. (incl. 1 fig.). Appl. 3 Nov 2000. Int. Cl. C08F 2/14. US Patent Specification 6,723,804 B1
189. Chevron Phillips Chemical Company LP, Spectroscopic method and apparatus for monitoring polyolefin production. Inventor: D.R. Battiste. 1 Jan 2008. 47 pp. (incl. 31 fig.). Appl. 24 Sep 2004. Int. Cl. G01N 21/65; G01J 3/44. US Patent Specification 7,315,369 B2
190. Chevron Phillips Chemical Company LP, Monitoring and control of processes for making 1-hexene. Inventor: D.R. Battiste. 8 Jul 2008. 9 pp. (incl. 0 fig.). Appl. 3 Nov 2000. Int. Cl. C07C 5/00; C07C 5/08. US Patent Specification 7,396,970
191. ExxonMobil Chemical Company, On-line analysis of polymer properties for control of a solution phase reaction system. Inventors: D.G. Marrow, B.E. Englehorn and D.A. Yahn. 10 pp. (incl. 2 fig.). Appl. 23 Jun 2006. Int. Cl. G01J 3/44; G01N 21/65. US Patent Application Publication 2007/019191 A1
192. ExxonMobil Chemical Company, On-line measurement and control of polymer properties by Raman spectroscopy. Inventors: R.L. Long and D.G. Marrow. 38 pp. (incl. 9 fig.). Appl. 8 May 2003. Int. Cl. G06F 19/00. US Patent Application Publication 2006/0136149 A1
193. ExxonMobil Chemical Patents Inc, On-line measurement and control of polymer properties by Raman spectroscopy. Inventors: R.L. Long and D.G. Marrow. 3 Oct 2006. 40 pp. (incl. 11 fig.). Appl. 15 Oct 2002. Int. Cl. G01J 3/44. US Patent Specification 7,116,414 B2
194. ExxonMobil Chemical Patents Inc, On-line measurement and control of polymer product properties by Raman spectroscopy. Inventors: D.G. Marrow, A.M. Cochran and S.T. Roger. 12 Sep 2006. 17 pp. (incl. 4 fig.). Appl. 9 Dec 2003. Int. Cl. G01J 3/44; G01N 21/65. US Patent Specification 7,106,437
195. ExxonMobil Chemical Company, On-line Raman analysis and control of a high pressure reaction system. Inventors: D.G. Marrow and D.A. Yahn. 11 pp. (incl. 3 fig.). Appl. 22 Jul 2005. Int. Cl. C08G 73/00. US Patent Application Publication 2007/0021586 A1
196. M. van den Brink, M. Pepers and A.M. van Herk, Raman spectroscopy of polymer latexes, *J. Raman Spectrosc.*, **33**, 264–272 (2002).
197. Bayer Polymers LLC, Process for the production of graft polymers. Inventors: E. Wenz, V. Buchholz, H. Eichenauer, U. Wolf, R.-J. Born, U. Jansen and W. Dietz. 11 pp. (incl. 1 fig.). Appl. 28 Oct 2002. Int. Cl. C08F 279/00; C08F 255/06. US Patent Application Publication 2003/0130433 A1
198. J.C. Santos, M.M. Reis, R.A.F. Machado, A. Bolzan, C. Sayer, R. Giudici and P.H.H. Araújo, Online monitoring of suspension polymerization reactions using Raman spectroscopy, *Ind. Eng. Chem. Res.*, **43**, 7282–7289 (2004).
199. J.C. Santos, C.N. Lopes, M.M. Reis, R. Giudici, C. Sayer, R.A.F. Machado and P.H.H. Araújo, Comparison of techniques for the determination of conversion during suspension polymerization reactions, *Braz. J. Chem. Eng.*, **25**, 399–407 (2008).
200. M.M. Reis, P.H.H. Araújo, C. Sayer and R. Giudici, Evidences of correlation between polymer particle size and Raman scattering, *Polymer*, **44**, 6123–6128 (2003).
201. C. Bauer, B. Amram, M. Agnely, D. Charmot, J. Sawatzki, M. Dupuy and J.-P. Huvenne, On-line monitoring of a latex emulsion polymerization by fiber-optic FT-Raman spectroscopy. Part I: Calibration, *Appl. Spectrosc.*, **54**, 528–535 (2000).
202. O. Elizalde, J.M. Asua and J.R. Leiza, Monitoring of high solids content starved-semi-batch emulsion copolymerization reactions by Fourier transform Raman spectroscopy, *Appl. Spectrosc.*, **59**, 1270–1279 (2005).
203. O. Elizalde, J.R. Leiza and J.M. Asua, On-line monitoring of all-acrylic emulsion polymerization reactors by Raman spectroscopy, *Macromol. Symp.*, **206**, 135–148 (2004).
204. O. Elizalde, M. Azpeitia, M.M. Reis, J.M. Asua and J.R. Leiza, Monitoring emulsion polymerization reactors: calorimetry versus Raman spectroscopy, *Ind. Eng. Chem. Res.*, **44**, 7200–7207 (2005).
205. O. Elizalde, J.M. Asua and J.R. Leiza, Monitoring of emulsion polymerization reactors by Raman spectroscopy: Calibration model maintenance, *Appl. Spectrosc.*, **59**, 1280–1285 (2005).
206. T.R. McCaffery and Y.G. Durant, Application of low-resolution Raman spectroscopy to online monitoring of miniemulsion polymerization, *J. Appl. Polym. Sci.*, **86**, 1507–1515 (2002).
207. M.M. Reis, M. Uliana, C. Sayer, P.H.H. Araújo and R. Giudici, Monitoring emulsion homopolymerization reactions using FT-Raman spectroscopy, *Braz. J. Chem. Eng.*, **22**, 61–74 (2005).
208. M.M. Reis, P.H.H. Araújo, C. Sayer and R. Giudici, Development of calibration models for estimation of monomer concentration by Raman spectroscopy during emulsion polymerization: facing the medium heterogeneity, *J. Appl. Polym. Sci.*, **93**, 1136–1150 (2004).
209. M.M. Reis, P.H.H. Araújo, C. Sayer and R. Giudici, Spectroscopic on-line monitoring of reactions in dispersed medium: chemometric challenges, *Anal. Chim. Acta*, **595**, 257–265 (2007).

210. D. Dong and T.J. McAvoy, Nonlinear principal component analysis – based on principal curves and neural networks, *Comput. Chem. Eng.*, **20**, 65–78 (1996).
211. J.-H. Jiang, Y. Ozaki, M. Kleimann and H.W. Siesler, Resolution of two-way data from on-line Fourier-transform Raman spectroscopic monitoring of the anionic dispersion polymerization of styrene and 1,3-butadiene by parallel vector analysis (PVA) and window factor analysis (WFA), *Chemom. Intell. Lab. Syst.*, **70**, 83–92 (2004).
212. D. Schaubroeck, S. Brughmans, C. Vercaemst, J. Schaubroeck and F. Verpoort, Qualitative FT-Raman investigation of the ring opening metathesis polymerization of dicyclopentadiene, *J. Mol. Catal. A: Chem.*, **254**, 180–185 (2006).
213. M. Gnyba, M. Keränen, A. Maaninen, J. Suhonen, M. Jedrzejewska-Szczerska, B.B. Kosmowski and P. Wierzba, Raman system for on-line monitoring and optimisation of hybrid polymer gelation, *Opto-Electron. Rev.*, **13**, 9–17 (2005).
214. A. Mank, E. Verstegen and H. Kloosterboer, In-situ curing analysis of photoreplicated objective lenses using Raman and IR spectroscopy, *J. Appl. Polym. Sci.*, **99**, 1287–1295 (2006).
215. K.C. Cole, Vibrational spectroscopy of polymer composites, in *Vibrational Spectroscopy of Polymers*, N.J. Everall, J.M. Chalmers, and P.R. Griffiths (Eds), John Wiley & Sons, Hoboken, 2007.
216. J.C. Cruz, T.A. Osswald and M. Kemper, Precise curing analysis of accelerated unsaturated polyester reactions under Raman spectroscopy. *Proceedings, 64th Annual Technical Conference – Society of Plastics Engineers (ANTEC), Charlotte, NC, USA, May 7–11, 2006, Society of Plastics Engineers, Brookfield, CT*, 2828–2832.
217. J.C. Cruz, T.A. Osswald and M. Kemper, Monitoring curing reactions of thermosets under high pressure by use of Raman spectroscopy. *Proceedings, 65th Annual Technical Conference – Society of Plastics Engineers (ANTEC), Cincinnati, OH, USA, May 6–11, 2007, Society of Plastics Engineers, Brookfield, CT*, 2129–2133.
218. S.E. Barnes, E.C. Brown, N. Corrigan, P.D. Coates, E. Harkin-Jones and H.G.M. Edwards, Raman spectroscopic studies of the cure of dicyclopentadiene (DCPD), *Spectrochim. Acta, Part A*, **61A**, 2946–2952 (2005).
219. J. Monni, P. Niemelä, L. Alvila and T.T. Pakkanen, Online monitoring of synthesis and curing of phenol-formaldehyde resol resins by Raman spectroscopy, *Polymer*, **49**, 3865–3874 (2008).
220. P. Musto, M. Abbate, G. Ragosta and G. Scarinzi, A study by Raman, near-infrared and dynamic-mechanical spectroscopies on the curing behaviour, molecular structure and viscoelastic properties of epoxy/anhydride networks, *Polymer*, **48**, 3703–3716 (2007).
221. C. Saal, Quantification of polymorphic forms by Raman spectroscopy, *Am. Pharm. Rev.*, **9**, 76–81 (2006).
222. T.L. Threlfall and J.M. Chalmers, Vibrational spectroscopy of solid-state forms – introduction, principles and overview, in *Applications of Vibrational Spectroscopy in Pharmaceutical Research and Development*, D.E. Pivonka, J.M. Chalmers, and P.R. Griffiths (Eds), Wiley-Interscience, New York, 2007.
223. D.H. Igo and P. Chen, Vibrational spectroscopy of solid-state forms – applications and examples, in *Applications of Vibrational Spectroscopy in Pharmaceutical Research and Development*, D.E. Pivonka, J.M. Chalmers and P.R. Griffiths (Eds), Wiley-Interscience, New York, 2007.
224. G. Zhou, R. Guenard and Z. Ge, Infrared and Raman spectroscopy for process development, in *Applications of Vibrational Spectroscopy in Pharmaceutical Research and Development*, D.E. Pivonka, J.M. Chalmers and P.R. Griffiths (Eds), Wiley-Interscience, New York, 2007.
225. F. Laplant and A. De Paepe, Raman spectroscopy for identifying polymorphs, in *Pharmaceutical Applications of Raman Spectroscopy*, S. Šašic (Ed), Technology for the Pharmaceutical Industry Series, John Wiley & Sons, Ltd, Hoboken, 2008.
226. J.M. Chalmers and G. Dent, Vibrational spectroscopic methods in pharmaceutical solid-state characterization, *Polymorphism*, 95–138 (2006).
227. G. Févotte, In situ Raman spectroscopy for in-line control of pharmaceutical crystallization and solids elaboration processes: a review, *Chem. Eng. Res. Des.*, **85**, 906–920 (2007).
228. R. Hilfiker, J. Berghausen, F. Blatter, *et al.*, Polymorphism – integrated approach from high-throughput screening to crystallization optimization, *J. Therm. Anal. Calorim.*, **73**, 429–440 (2003).
229. R. Hilfiker, J. Berghausen, F. Blatter, S.M. De Paul, M. Szelagiewicz and M. Von Raumer, High-throughput screening for polymorphism, *Chim. Oggi*, **21**, 75–76, 78–79 (2003).
230. C. Anderton, A valuable technique for polymorph screening, *Am. Pharm. Rev.*, **10**, 34–40 (2007).
231. T. Kojima, S. Onoue, N. Murase, F. Katoh, T. Mano and Y. Matsuda, Crystalline form information from multiwell plate salt screening by use of Raman microscopy, *Pharm. Res.*, **23**, 806–812 (2006).
232. S.M. Mehrens, U.J. Kale and X. Qu, Statistical analysis of differences in the Raman spectra of polymorphs, *J. Pharm. Sci.*, **94**, 1354–1367 (2005).

233. B. De Spiegeleer, D. Seghers, R. Wieme, J. Schaubroeck, F. Verpoort, G. Slegers and L. Van Vooren, Determination of the relative amounts of three crystal forms of a benzimidazole drug in complex finished formulations by FT-Raman spectroscopy, *J. Pharm. Biomed. Anal.*, **39**, 275–280 (2005).
234. R. Vehring, Red-excitation dispersive Raman spectroscopy is a suitable technique for solid-state analysis of respirable pharmaceutical powders, *Appl. Spectrosc.*, **59**, 286–292 (2005).
235. A. Caillet, F. Puel and G. Févotte, Quantitative in situ monitoring of citric acid phase transition in water using Raman spectroscopy, *Chem. Eng. Process.*, **47**, 377–382 (2008).
236. G. Févotte, C. Alexandre and S.O. Nida, A population balance model of the solution-mediated phase transition of citric acid, *AIChE J.*, **53**, 2578–2589 (2007).
237. F. Tian, J.A. Zeitler, C.J. Strachan, D.J. Saville, K.C. Gordon and T. Rades, Characterizing the conversion kinetics of carbamazepine polymorphs to the dihydrate in aqueous suspension using Raman spectroscopy, *J. Pharm. Biomed. Anal.*, **40**, 271–280 (2006).
238. H. Qu, M. Louhi-Kultanen, J. Rantanen and J. Kallas, Solvent-mediated phase transformation kinetics of an anhydrate/hydrate system, *Cryst. Growth Des.*, **6**, 2053–2060 (2006).
239. H. Qu, J. Kohonen, M. Louhi-Kultanen, S.-P. Reinikainen and J. Kallas, Spectroscopic monitoring of carbamazepine crystallization and phase transformation in ethanol-water solution, *Ind. Eng. Chem. Res.*, **47**, 6991–6998 (2008).
240. A. Jayasankar, D.J. Good and N. Rodríguez-Hornedo, Mechanisms by which moisture generates cocrystals, *Mol. Pharmaceutics*, **4**, 360–372 (2007).
241. A. Vega-González, P. Marteau and P. Subra-Paternault, Monitoring a crystallization induced by compressed CO_2 with Raman spectroscopy, *AIChE J.*, **52**, 1308–1317 (2006).
242. J.A. Falcon and K.A. Berglund, In situ monitoring of antisolvent addition crystallization with principal components analysis of Raman spectra, *Cryst. Growth Des.*, **4**, 457–463 (2004).
243. P.A. Anquetil, C.J.H. Brenan, C. Marcolli and I.W. Hunter, Laser Raman spectroscopic analysis of polymorphic forms in microliter fluid volumes, *J. Pharm. Sci.*, **92**, 149–160 (2003).
244. L.E. O'Brien, P. Timmins, A.C. Williams and P. York, Use of in situ FT-Raman spectroscopy to study the kinetics of the transformation of carbamazepine polymorphs, *J. Pharm. Biomed. Anal.*, **36**, 335–340 (2004).
245. Y. Hu, J.K. Liang, A.S. Myerson and L.S. Taylor, Crystallization monitoring by Raman spectroscopy: Simultaneous measurement of desupersaturation profile and polymorphic form in flufenamic acid systems, *Ind. Eng. Chem. Res.*, **44**, 1233–1240 (2005).
246. Y. Hu, H. Wikström, S.R. Byrn and L.S. Taylor, Estimation of the transition temperature for an enantiotropic polymorphic system from the transformation kinetics monitored using Raman spectroscopy, *J. Pharm. Biomed. Anal.*, **45**, 546–551 (2007).
247. X. Yang, J. Lu, X. Wang and C.B. Ching, *In situ* monitoring of the solution-mediated polymorphic transformation of glycine: Characterization of the polymorphs and observation of the transformation rate using Raman spectroscopy and microscopy, *J. Raman Spectrosc.*, **39**, 1433–1439 (2008).
248. T. Ono, J.H. ter Horst and P.J. Jansens, Quantitative measurement of the polymorphic transformation of L-glutamic acid using in-situ Raman spectroscopy, *Cryst. Growth Des.*, **4**, 465–469 (2004).
249. J. Cornel, C. Lindenberg and M. Mazzotti, Quantitative application of in situ ATR-FTIR and Raman spectroscopy in crystallization processes, *Ind. Eng. Chem. Res.*, **47**, 4870–4882 (2008).
250. H.R.H. Ali, H.G.M. Edwards, M.D. Hargreaves, T. Munshi, I.J. Scowen and R.J. Telford, Vibrational spectroscopic characterisation of salmeterol xinafoate polymorphs and a preliminary investigation of their transformation using simultaneous in situ portable Raman spectroscopy and differential scanning calorimetry, *Anal. Chim. Acta*, **620**, 103–112 (2008).
251. I. Miroshnyk, L. Khriachtchev, S. Mirza, J. Rantanen, J. Heinämäki and J. Yliruusi, Insight into thermally induced phase transformations of erythromycin A dihydrate, *Cryst. Growth Des.*, **6**, 369–374 (2006).
252. S.-Y. Lin, W.-T. Cheng and S.-L. Wang, Thermal micro-Raman spectroscopic study of polymorphic transformation of famotidine under different compression pressures, *J. Raman Spectrosc.*, **38**, 39–43 (2007).
253. W.-T. Cheng, S.-Y. Lin and M.-J. Li, Raman microspectroscopic mapping or thermal system used to investigate milling-induced solid-state conversion of famotidine polymorphs, *J. Raman Spectrosc.*, **38**, 1595–1601 (2007).
254. B.H. Lohse, A. Calka and D. Wexler, Raman spectroscopy as a tool to study TiC formation during controlled ball milling, *J. Appl. Phys.*, **97**, 114912/1–114912/7 (2005).
255. X. Yang, W. Ke, P. Zi, F. Liu and L. Yu, Detecting and identifying the complexation of nimodipine with hydroxypropyl-β-cyclodextrin present in tablets by Raman spectroscopy, *J. Pharm. Sci.*, **97**, 2702–2719 (2008).

256. C. Starbuck, A. Spartalis, L. Wai, J. Wang, P. Fernandez, C.M. Lindemann, G.X. Zhou and Z.H. Ge, Process optimization of a complex pharmaceutical polymorphic system via in situ Raman spectroscopy, *Cryst. Growth Des.*, **2**, 515–522 (2002).
257. S. Barnes, J. Gillian, A. Diederich, D. Burton and D. Ertl, In process monitoring of polymorphic form conversion by Raman spectroscopy and turbidity measurements, *Am. Pharm. Rev.*, **11**, 80, 82–85 (2008).
258. J. Aaltonen, K. Kogermann, C.J. Strachan and J. Rantanen, In-line monitoring of solid-state transitions during fluidisation, *Chem. Eng. Sci.*, **62**, 408–415 (2006).
259. N. Sandler, J. Rantanen, J. Heinämäki, M. Römer, M. Marvola and J. Yliruusi, Pellet manufacturing by extrusion-spheronization using process analytical technology, *AAPS PharmSciTech*, **6**, E174–83 (2005).
260. A.C. Williams, V.B. Cooper, L. Thomas, L.J. Griffith, C.R. Petts and S.W. Booth, Evaluation of drug physical form during granulation, tabletting and storage, *Int. J. Pharm.*, **275**, 29–39 (2004).
261. A.M. Amado, M.M. Nolasco and P.J.A. Ribeiro-Claro, Probing pseudopolymorphic transitions in pharmaceutical solids using Raman spectroscopy: Hydration and dehydration of theophylline, *J. Pharm. Sci.*, **96**, 1366–1379 (2007).
262. Basell Polyolefine GmbH, Determination of mechanical properties of polymers by Raman spectroscopy. Inventors: D. Lilge and C. Gabriel. 22 pp. (incl. 3 fig.). Appl. 11 Aug 2003. Int. Cl. G01N 21/65; G01N 21/63. WO Patent Application Publication 2004/017056 (A1)
263. Akzo Nobel NV, Technique for measuring properties of polymeric fibres. Inventors: R.J. Van Wijk, A.P. De Weijer, D.A. Klarenberg, R. De Jonge and G.J. Jongerden. 23 Jul 2003. 8 pp. (incl. 2 fig.). Appl. 1 May 2000. Int. Cl. G01N 21/65; G01N 21/89. US Patent Specification 6,423,262
264. K. Ito, T. Kato and T. Ona, Rapid viscosity determination of waterborne automotive paint emulsion system by FT-Raman spectroscopy, *Vib. Spectrosc.*, **35**, 159–163 (2004).
265. Marathon Ashland Petroleum, Gasoline RFG analysis by a spectrometer. Inventors: W.T. Welch, R.R. Bledsoe, B.K. Wilt, and M.B. Sumner. 31 Oct 2000. 37 pp. (incl. 13 fig.). Appl. 19 Dec 1997. Int. Cl. G01J 3/42. US Patent Specification 6,140,647
266. Ashland Inc. and Old Dominion University Research Foundation, Process and apparatus for octane numbers and Reid vapor pressure by Raman spectroscopy. Inventors: J.B. Cooper, R.R. Bledsoe Jr, K.L. Wise, M.B. Sumner, W.T. Welch and B.K. Wilt. 6 Apr 1999. 19 pp. (incl. 10 fig.). Appl. 30 Sep 1996. Int. Cl. G01N 21/35. US Patent Specification 5,892,228
267. R. Müller, U. Künecke, A. Thuaire, M. Mermoux, M. Pons and P. Wellmann, Investigation of the charge carrier concentration in highly aluminum doped SiC using Raman scattering, *Phys. Status Solidi C*, **3**, 558–561 (2006).
268. C. Frausto-Reyes, J.R. Molina-Contreras, C. Medina-Gutiérrez and S. Calixto, CdTe surface roughness by Raman spectroscopy using the 830nm wavelength, *Spectrochim. Acta, Part A*, **65A**, 51–55 (2006).
269. E.D. Lipp and R.L. Grosse, On-line monitoring of chlorosilane streams by Raman spectroscopy, *Appl. Spectrosc.*, **52**, 42–46 (1998).
270. Institut Francais Du Petrole, Method for optimization of operation of a xylene separation unit by a simulated countercurrent. Inventors: N. Couenne and L. Wolff. 5 Mar 2004. 21 pp. (incl. 3 fig.). Appl. 28 Aug 2002. Int. Cl. B01D 15/18, C07C 15/18, 7/12. FR Patent Specification 2,843,893
271. Institut Francais Du Petrole, Method for regulating a process for the separation of isomers of aromatic hydrocarbons having from 8 to 10 carbon atoms. Inventors: F. Cansell, G. Hotier, P. Marteau, and N. Zanier. 29 Oct 1996. 16 pp. (incl. 7 fig.). Appl. 29 Jun 1994. Int. Cl. C07C 700. US Patent Specification 5,569,808
272. M. Louhi-Kultanen, H. Qu, N. Salmela, J. Kallas and L. Nyström, Filter cake washing: Partial dissolution of organic particles and real-time monitoring based on Raman spectroscopy, *Sep. Purif. Technol.*, **59**, 270–276 (2008).
273. J.K. Liang, S. Byrn, A. Newman, J. Stults and V. Smolenskayay, On-line Raman spectroscopy in pharmaceutical process development: Application and future prospects, *Am. Pharm. Rev.*, **10**, 45–51 (2007).
274. G. Walker, S.E.J. Bell, M. Vann, D.S. Jones,and G. Andrews, Fluidised bed characterization using Raman spectroscopy: Applications to pharmaceutical processing, *Chem. Eng. Sci.*, **62**, 3832–3838 (2007).
275. D.S. Hausman, R.T. Cambron and A. Sakr, Application of on-line Raman spectroscopy for characterizing relationships between drug hydration state and tablet physical stability, *Int. J. Pharm.*, **299**, 19–33 (2005).
276. T.R.M. De Beer, M. Alleso, F. Goethals, *et al.*, Implementation of a process analytical technology system in a freeze-drying process using Raman spectroscopy for in-line process monitoring, *Anal. Chem.*, **79**, 7992–8003 (2007).
277. W. Schabel, I. Ludwig and M. Kind, Measurements of concentration profiles in polymeric solvent coatings by means of an inverse confocal micro Raman spectrometer – Initial results, *Drying Technol.*, **22**, 285–294 (2004).

278. Advanced Micro Devices Inc., Method of in situ monitoring of thickness and composition of deposited films using Raman spectroscopy. Inventor: E. Adem. 23 Dec 2003. 5 pp. (incl. 1 fig.). Appl. 4 Feb 2002. Int. Cl. C23C 16/52. US Patent Specification 6,667,070
279. V.A. Maroni, J.L. Reeves and G. Schwab, On-line characterization of YBCO coated conductors using Raman spectroscopy methods, *Appl. Spectrosc.*, **61**, 359–366 (2007).
280. M. Mermoux, B. Marcus, L. Abello, N. Rosman and G. Lucazeau, *In situ* Raman monitoring of the growth of CVD diamond films, *J. Raman Spectrosc.*, **34**, 505–514 (2003).
281. W. Qiu, Y.K. Vohra and S.T. Weir, Role of nitrogen in the homoepitaxial growth on diamond anvils by microwave plasma chemical vapor deposition, *J. Mater. Res.*, **22**, 1112–1117 (2007).
282. D. Hill, T. Jawhari, J.G. Cespedes, J.A. Garcia and E. Bertran, In-situ monitoring of laser annealing by micro-Raman spectroscopy for hydrogenated silicon nanoparticles produced in radio frequency glow discharge, *Phys. Status Solidi A*, **203**, 1296–1300 (2006).
283. J. Palm, S. Jost, R. Hock and V. Probst, Raman spectroscopy for quality control and process optimization of chalcopyrite thin films and devices, *Thin Solid Films*, **515**, 5913–5916 (2007).
284. V. Terzieva, L. Souriau, F. Clemente, A. Benedetti, M. Caymax and M. Meuris, Ge substrates made by Ge-condensation technique. Challenges and current understanding, *Mater. Sci. Semicond. Process.*, **9**, 449–453 (2006).
285. Y. Huang, X.D. Chen, S. Fung, C.D. Beling and C.C. Ling, Spatial characterization of a 2 inch GaN wafer by Raman spectroscopy and capacitance-voltage measurements, *J. Phys. D: Appl. Phys.*, **37**, 2814–2818 (2004).
286. S. Romero-Torres, J.D. Pérez-Ramos, K.R. Morris and E.R. Grant, Raman spectroscopic measurement of tablet-to-tablet coating variability, *J. Pharm. Biomed. Anal.*, **38**, 270–274 (2005).

8

Near-Infrared Chemical Imaging for Product and Process Understanding

E. Neil Lewis[1], Joseph W. Schoppelrei[2], Lisa Makein[3], Linda H. Kidder[1], and Eunah Lee[4]

[1] *Malvern Instruments, Columbia, MD, USA*
[2] *National Geospatial-Intelligence Agency, Reston, VA, USA*
[3] *Malvern Instruments, Malvern, UK*
[4] *Horiba Jobin Yvon, Edison, NJ, USA*

8.1 The PAT Initiative

Conventional near-infrared (NIR) spectroscopy is a very rugged and flexible technique that can be generally adopted for use in a wide range of chemical analyses in many different research and industrial (process) applications. NIR spectroscopic imaging, through the additional information provided by its spatial perspective, offers greater understanding and therefore control of the manufacturing process of complex composite materials and products. While the focus of this work is on pharmaceuticals and their manufacture, the reader can easily incorporate these ideas for other industries and products including, general chemical, polymer, agricultural, food, cosmetic, biological, or any other manufacturing environments where organic materials are processed to form a finished product.

The process analytical technology (PAT) initiative of the US Food & Drug Administration (FDA) strongly advocates the use and development of new measurement technologies for pharmaceutical manufacturing.[1] Manufacturing processes that rely on testing only after the production process is complete are understandably inefficient, and one of the chief goals of the PAT initiative is to replace this antiquated quality assurance (QA) system. Recently, the reach of this new approach to pharmaceutical products has been extended beyond manufacturing to explicitly encompass formulation development, as described in the International Conference of Harmonisation (ICH) Pharmaceutical Development guideline (Q8 (R1)).[2,3] Based on the overall philosophy of quality by design (QbD), the goal is to develop a final product that will have the appropriate performance characteristics across a range of manufacturing operating values, making product quality testing a check rather than a primary means of control. It relies on identifying 'critical to quality attributes' (CQAs) of both formulation and manufacturing processes. In practice CQAs are identified by systematically changing a

Process Analytical Technology 2e Edited by Katherine Bakeev
© 2010 John Wiley & Sons, Ltd.

broad range of variables and identifying those which have the most significant impact on final product quality. These include the overall formulation, as well as manufacturing operations such as blending, drying, granulation, particle size, milling, etc. With a focus on end product performance, this process promotes an understanding of the entire formulation and manufacturing process. These steps and the knowledge that they impart allow the determination of a 'processing window', a design and manufacturing space within which product quality is assured, even when specific inputs vary. In this respect PAT can be viewed as a subset of QbD, as the requirement remains for process control. The route to success in this area is by increased product understanding and control, accomplished through the identification of CQAs and the subsequent implementation of analytical technologies that can monitor them.

The successful adoption of QbD, which incorporates PAT, will result in more reliable and efficient manufacturing processes and end products. Perhaps a less obvious but equally important benefit is the promise of a more flexible regulatory approach.

8.2 The Role of Near-Infrared Chemical Imaging (NIR-CI) in the Pharmaceutical Industry

8.2.1 Characterization of solid dosage forms

Solid dosage forms (tablets or capsules) are heterogeneous, including one or more active pharmaceutical ingredients (APIs), fillers, binders, disintegrants, lubricants, and other materials. The goal of formulation development is to ensure that the final product has desirable and consistent characteristics for manufacturing, storage, handling and release of the therapeutic agent. The compositions, sizes, spatial relationships, and consistency of localized domains in a formulation reflect the individual unit operations used to produce the product, and can have a significant impact on physical properties and performance characteristics. Often poor product performance can be caused by differences in the spatial distribution of sample components. Even a relatively simple formulation may produce widely varying therapeutic performance if the manufacturing process is unstable, as a result of poor *product* and/or *process* understanding. Formulations that require the distribution of sample components to be extremely well characterized and controlled, such as low dose or controlled release products, further increase the demands on product uniformity.

Current QA testing requirements for potency, purity, and bioavailability (determining gross composition, verifying the absence of contaminants, and dissolution profiling) do not capture the spatial distribution of sample components that can lead to out-of-specification issues. Additionally the analytical techniques used (high performance liquid chromatography (HPLC), mass spectrometry, UV-vis, dissolution testing, etc.) are destructive, so can only be applied to a small subset of a standard batch. And being based on wet chemistry, these traditional QA techniques for pharmaceutical release testing are laboratory based and time consuming, making it difficult or impossible to quickly trace and correct the sources of variations or failures. A number of spectroscopic techniques such as mid-infrared (MIR), near-infrared (NIR) and Raman enable rapid, nondestructive sample analysis, and these methods can be employed at a number of points in the pharmaceutical development and manufacturing process. In particular, NIR spectroscopy is quickly becoming a workhorse technique for the industry due to its high information content and flexibility of implementation.[4–6] Currently, it is widely used for the rapid characterization of raw materials,[7,8] and has also been used in applications such as blend homogeneity,[9–11] moisture measurement, and the analysis of intact tablets.[12,13]

However, spectroscopic techniques, although nondestructive, do not provide the spatial information that can be critical in solving performance issues, or in developing true product and process understanding. Only techniques that combine both spatial and chemical identification information provide a complete set of tools to determine both product and process understanding.

8.2.2 'A picture is worth a thousand words' (attributed to Fred R. Barnard)

This need is addressed in part by NIR-CI, in that it offers the ability to obtain high fidelity, spatially resolved pictures of the chemistry of the sample. The ability to visualize and assess the compositional heterogeneity and structure of the end product is invaluable for both the development and manufacture of solid dosage forms.[14] NIR chemical images can be used to determine content uniformity, particle sizes and distributions of all the sample components, polymorph distributions, moisture content and location, contaminations, coating and layer thickness, and a host of other structural details.[15–19]

The underlying tenant of QbD and the PAT initiative is to develop a better product through understanding and control of the manufacturing process. NIR-CI rapidly and nondestructively delivers data that can be used to identify CQAs. The technique is fast and nondestructive and can be used independently, or in concert with other techniques such as dissolution analysis, to rapidly diagnose potential production problems. Incorporating NIR-CI analysis into the preformulation and formulation development phases can ultimately improve testing turn-around, limit scale-up difficulties, and reduce the time-to-market. This presents a more robust process, imparting significant economic benefits to the pharmaceutical industry.

NIR-CI instrumentation is rugged and flexible, suitable for both the laboratory and manufacturing environment. Therefore analysis methods developed in the laboratory can often be tailored for implementation near-line or at-line. NIR-CI is also a massively parallel approach to NIR spectroscopy, making the technique well suited for high throughput applications.

8.3 Evolution of NIR Imaging Instrumentation

8.3.1 Spatially resolved spectroscopy – mapping

The development of spatially resolved molecular spectroscopy has its beginnings in 1949 with a publication in *Nature* that described the use of a microscope coupled with an IR spectrometer, collecting the first spatially resolved IR spectra.[20] While this was not an imaging instrument, since it only collected single-point IR spectra, it proved the feasibility of collecting spectral information from a spatially defined location. Taking this idea one step further, the first chemical maps were published in 1988 by Harthcock and Atkin.[21] In this work spectra were collected from an inhomogeneous sample using a Fourier transform infrared (FTIR) microscope fitted with a moving stage. This experiment consisted of collecting a series of spectra from adjacent locations by moving the sample at regular intervals between each measurement. The resulting 'pictures', while somewhat crude by today's standards, demonstrated the concept of spatially localizing chemical species within the sample using infrared spectroscopy. The value of this approach was swiftly realized in the analytical community.

Since that time, improvements in optical design and advancements in computers, software, and automation technology have enhanced the speed, performance, and utility of spectroscopic mapping instruments. Subsequently, FTIR and Raman microscope mapping systems are now found in many research laboratories. However, these instruments still utilize a moving sample stage and the 'step-and-acquire' acquisition mode, and because of this limitation, data collection progresses relatively slowly. Additionally, the fact that moving parts are not optimal for at-line or on-line instrumentation limits the usefulness of this type of system for PAT applications.

8.3.2 The infrared focal-plane array

Unlike mapping, the detector used for infrared chemical imaging is a two-dimensional detector array. The infrared focal-plane array (FPA) detector, the long-wavelength analog of a typical digital camera, was

originally developed for military applications such as missile detection, targeting and night-vision devices. It is only within the last decade or so that they have been 'discovered' and used by the general infrared spectroscopy community.[22]Originally, only small format arrays (64 × 64 pixels) were available and were constructed from material such as indium antimonide (InSb) or mercury–cadmium–telluride (MCT). They were prone to mechanical failure, required cryogenic cooling, and were quite expensive (>$10 per pixel). More recently, as the technology has begun to shift from military applications to higher volume commercial uses, more economical, large format (320×256 pixels) FPAs made from InSb, MCT, or indium–gallium–arsenide (InGaAs) have entered the market at much lower cost (<$1 per pixel). Many of these cameras operate uncooled or incorporate solid-state cooling, and exhibit much improved operability and durability. More recently the infrared camera market has also seen the emergence of the uncooled microbolometer array using low-cost complementary metal oxide semiconductor CMOS technologies.[23]

8.3.3 Wavelength selection

While the two dimensional format of an FPA provides the ability to spatially analyze a sample, a spectroscopic imaging instrument must still incorporate a method of wavelength discrimination. Several approaches have been described in the literature[15,24–28] and, in some cases, subsequently implemented by instrument manufacturers. Each technology has its own benefits and limitations but it is beyond the scope of this work to explore each of these in detail. MIR and some NIR imaging instruments utilize the Fourier transform (FT) scheme and as a result, function in much the same way as traditional FTIR mapping systems; however, if a large-format FPA is employed, the need for moving the sample with an automated stage is eliminated. NIR and Raman imaging instruments have been constructed using solid-state tunable filter technologies, which offer several distinct advantages for industrial applications, including mechanical simplicity and flexibility for industrial applications.

8.3.4 The benefits of NIR spectroscopy

As stated in the introduction to this chapter, NIR spectroscopy has been shown to be an excellent tool applied to the solution of a wide range of pharmaceutical applications.[4–6] The high information content coupled with ease and reliability of sampling is a primary attraction of the technique over other spectroscopic approaches. While it is arguably true that fundamental modes observed in the MIR may offer higher specificity and sensitivity than the overtone and combination bands measured in the NIR region, one must balance the value of this 'additional' information against both the increased sampling constraints of the MIR experiment and the actual needs of the application at hand.

The high sensitivity of the MIR technique arises from the highly absorbing nature of most organic materials in this spectral region. Samples must be prepared in such a way as to limit the amount of material that interacts with the probing radiation (e.g. microtomed sections, or KBr dilutions). The use of reflection techniques, such a specular (SR) and diffuse reflectance (DRIFTS) in the MIR is highly dependent on the optical properties and morphology of the surface of the sample.[29] The data from these measurements often require care in interpretation, significant post collection corrections, and can be difficult to acquire reproducibly.[30,31] Even attenuated total reflectance (ATR) techniques require an extremely flat or deformable sample surface to achieve intimate contact with the internal reflection element. Subsequently, the quality and even the information content of MIR data will typically have at least some dependence on the technique and skill of the operator and the analyst. These sampling constraints become even more significant for a chemical imaging experiment. Preparing a sample suitable for large field interrogation without compromising its original spatial integrity is a formidable task, greatly limiting the applicability and speed of the MIR

experiment. However, some very elegant ATR imaging studies of pharmaceuticals in the MIR have been performed.[32,33]

On the other hand, the overall absorbance of NIR radiation exhibited by most organic materials is reduced by a factor of 10 to 100 compared to that of MIR light. Therefore the constraints placed upon sample presentation are significantly relaxed, even for optically 'thick' materials. The majority of solid samples can be measured noninvasively with little or no preparation using diffuse reflectance techniques. While the optical properties and physical uniformity of the sample surface will still influence the spectral data, their overall impact is less significant and more easily separated from the spectral information.[34] Therefore, NIR methods significantly reduce the sample handling burden and the dependence on operator technique to gather high quality, reproducible data. Relaxed sampling requirements increase the efficiency and widen the applicability of the technique, especially for nonlaboratory based measurements. For imaging, this also means that the native spatial integrity of the sample need not be disturbed.

The trade-off is that the broader overtone and combination bands observed in the NIR spectrum can be more difficult to assign to discrete vibrations or functionalities of the interrogated material. This is the source of the (largely undeserved) reputation of inferior specificity that has been assigned to the NIR technique. While interpretation problems may arise if one has no *a priori* information as to the chemical nature of the material being analyzed, this is seldom the case for pharmaceutical applications. In these samples, the component materials are known and generally well characterized. The pure components are readily available and can be used to acquire the independent spectral references, if necessary. The real information content coupled with enhanced reproducibility has proved NIR spectroscopy to be sensitive to subtle compositional changes, even in complex samples, making it preferable for a wide range of both qualitative and quantitative pharmaceutical applications.[4–6] This aspect of NIR spectroscopy makes it especially well suited for use in chemical imaging analysis where the added value is largely realized through the comparison of a large number of spatially resolved measurements rather than the examination of individual spectra.

8.3.5 NIR imaging instrumentation

In addition to the analytical and sampling benefits of NIR spectroscopy, implementing this spectral range in a chemical instrument also poses several distinct advantages. First, as described earlier, the most dramatic improvements in commercial FPA technology have been seen for detectors that operate in this region, spurred by applications such as night-vision devices and telecom diagnostic tools. The spectroscopy communities benefit from these advances in the wider choice of reliable, larger format, and more reasonably priced 'cameras' for the chemical imaging application. The breadth of illumination source options for work in the NIR region is also excellent. DC-powered quartz–halogen lamps are typically used as sources, and are widely available in a range of powers and configurations. These sources are relatively inexpensive and flexible, and do not pose any significant safety concerns.

An ideal PAT imaging analyzer should be flexible, robust and amenable to the manufacturing environment. Figure 8.1 depicts a generalized schematic of a NIR imaging instrument incorporating a large-format FPA as a detector and a liquid crystal tunable filter (LCTF) for wavelength discrimination. This relatively simple experimental configuration offers many advantages over other chemical imaging techniques. One tremendous benefit is the more relaxed demands for optics relative to MIR and Raman techniques. Working in the NIR permits the use of relatively simple refractive achromatic optics with long working distances. A standard laboratory system can be quickly configured by the operator for microscopic applications (a magnification of ~10 μm/pixel) or macroscopic applications (a magnification of ~500 μm/pixel). The result is a chemical image data set comprising tens of thousands of spectra of a single pharmaceutical granule or an entire blister pack of tablets collected rapidly from the same instrument. Instruments that image even larger

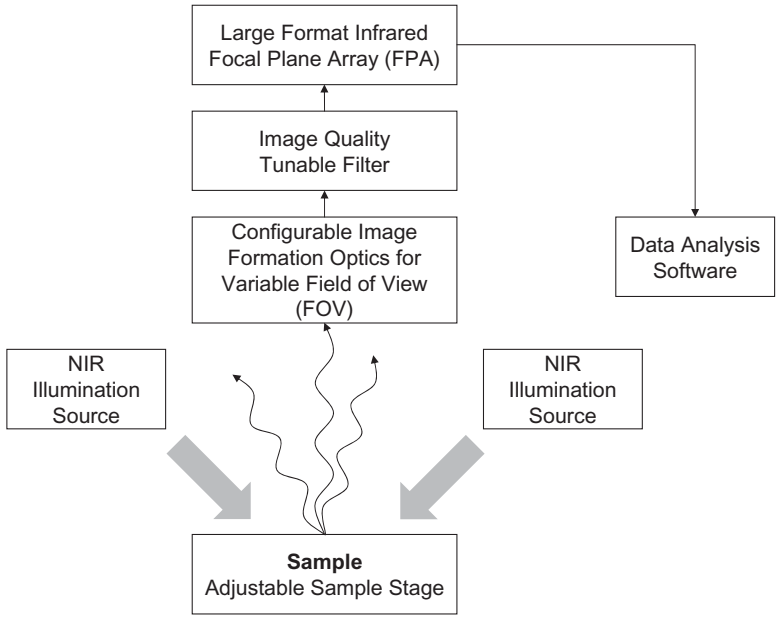


Figure 8.1 *Schematic representation of NIR chemical imaging instrument operating in diffuse reflectance mode. Radiation from the illumination source interacts with the sample. Light reflected off of the sample is focused onto a NIR sensitive 2D detector after passing through a solid-state tunable wavelength selection filter.*

areas can be purpose-built for process, high throughput, or other specialized applications. Because the illumination source used for the experimental configuration is a readily available quartz-halogen lamp, the area to be illuminated can be simply and cost effectively scaled up.

The tunable filter technology also has distinct PAT advantages. Rapid tuning of discrete wavelengths through software control and no moving parts enables the collection of data sets comprising approximately 80 000 spectra in a matter of a couple of minutes. However, finished process applications may only require data collection over a narrow spectral range or even use just a few analytically relevant wavelengths. Under these circumstances the use of a random access tunable filter offers compelling advantages by minimizing the size of the data sets, and dramatically speeding up the data collection process, as single wavelength images can be collected in less than a second. The basic design also permits the camera, LCTF, and imaging optics to be unified into a single compact wavelength-selectable chemical imaging module which can also be mounted in side-looking or inverted orientations for more specialized applications. Further, once a data collection method has been streamlined and validated, it can be readily 'bundled' such that collection and analysis is integrated into a turnkey module, and accomplished in 'real-time'. These chemical imaging protocols can proceed with little or no operator intervention.

In summary, the technology readily enables a broad selection of optical configurations to be used, providing ease of switching between highly disparate fields of view and the flexibility to chemically image irregular and uneven samples. Furthermore, this is achieved with a system that can collect >80 000 moderate resolution NIR spectra in several minutes or >80 000 lower resolution spectra in a few seconds, all with no-moving parts. This suggests that the generalized instrument design depicted by Figure 8.1 represents a highly rugged and robust chemical imaging platform.

8.4 Chemical Imaging Principles

Conventional digital imaging is the process of reproducing the spatial information of a scene onto a two-dimensional optical detector. Typical visible grayscale or color images are collected over a broad range of optical wavelengths. On the other hand, a single *spectral* image is usually collected over a narrow wavelength range. In the infrared or NIR spectral region it can be used to reveal the chemical composition of the sample through absorption by one or more chemical species within the sample at a particular 'diagnostic' wavelength. The result is a spatially resolved chemical map of the sample. Chemical (also called *hyperspectral*) imaging is the acquisition of images over a larger, usually contiguous, series of narrow spectral bands comparable to traditional (single-point) spectroscopic techniques. This is the fundamental concept of chemical imaging – a rapid analytical method that simultaneously delivers spatial, analytical (chemical), structural, and functional information.

8.4.1 The hypercube

Chemical image data sets are 'visualized' as a three-dimensional cube – called a hypercube – spanning one wavelength and two spatial dimensions (Figure 8.2). Each element within the cube contains the intensity response information measured at that spatial and spectral index. The hypercube can be treated as a series of spatially resolved spectra (called pixels) or, alternatively, as a series of spectrally resolved images (called image planes or channels). Selecting a single pixel will yield the spectrum recorded at that particular spatial location in the sample. Similarly, selecting a single image plane will show the intensity response (absorbance) of the scene at that particular wavelength.

8.4.2 Data analysis

Hypercubes collected from a chemical imaging experiment can be quite large. For example, a moderate-format NIR FPA (320×256) collects data using 81 920 individual detector elements (pixels). A full-range spectrum may contain more than 100 wavelength channels resulting in a cube with more than 8 000 000 individual elements. The corresponding information content of the entire cube is enormous. It is impractical (and not particularly useful) to attempt to manually interpret each element independently. Therefore the analysis software used to mine and portray the salient information from the hypercube becomes an integral part of the chemical imaging experiment.

There are four basic steps that are customarily used for the analysis of chemical imaging data: spectral correction, spectral preprocessing, classification, and statistical image analysis. The first two steps are somewhat universal for all comparative spectroscopic analyses. Classification techniques are also common in spectral analyses, but their application in imaging may differ from more familiar methods. A statistical analysis considers the data to be simply a collection of multiple spectra, without consideration of their spatial location on the sample, whereas image analysis serves as the departure from 'spectral' thinking and is used to evaluate and quantify the heterogeneity (contrast) developed in the image from the first three steps. There are no firm requirements to use all of the steps- they merely serve as a guideline for the general flow of data work-up. Later in this chapter we will present two 'case studies', showing some typical data work-ups. For each case study, the desired and derived information is different, and thus illustrates strategies for data processing.

As is true in approaching any large set of analytical data, efficient analysis of chemical images requires goals with regards to the information being sought. For poorly characterized samples, it is often necessary to use an iterative approach, using the initial results as a guide to develop more appropriate processing and analysis methods. Also because typically there is more than one processing path that can be used to elucidate

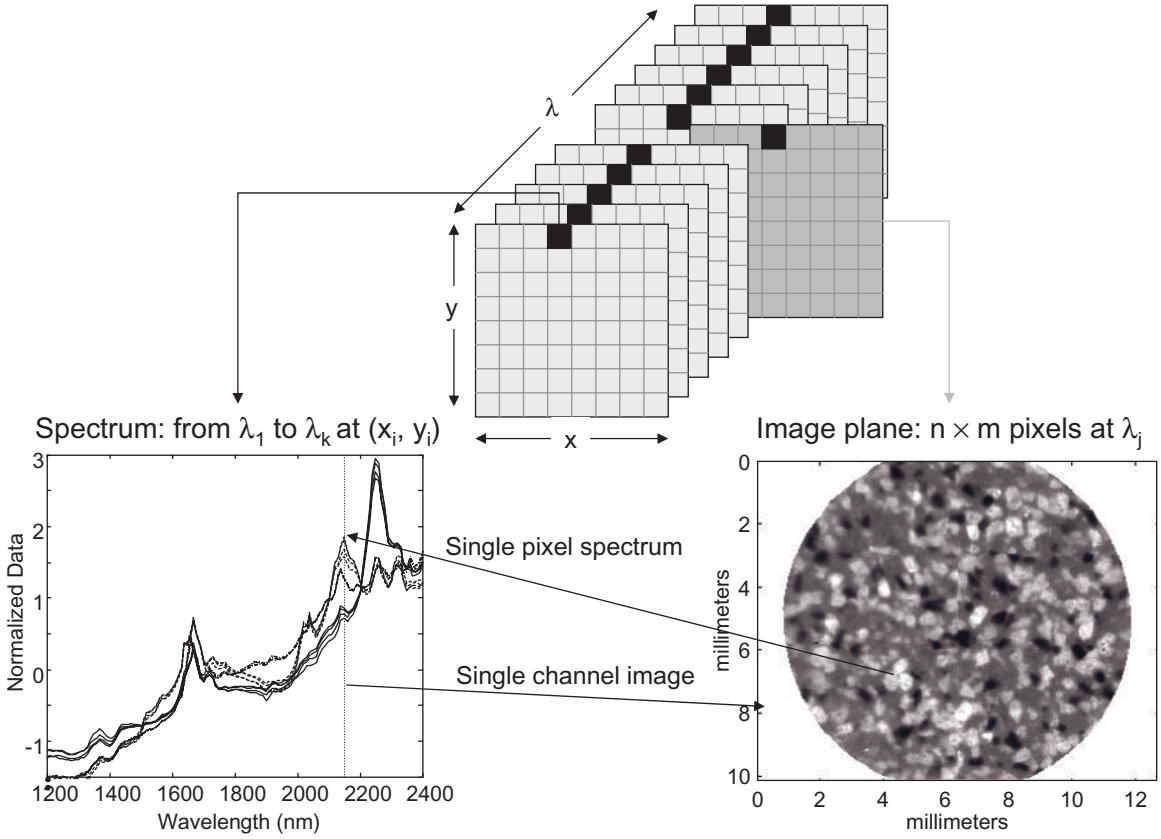


Figure 8.2 *Schematic representation of spectral imaging hypercube showing the relationship between spatial and spectral dimensions. The image displayed is a single-wavelength image, and its contrast is based on the intensity of IR absorption at that wavelength. Pixels containing sample components with a strong absorption at that wavelength will appear bright. The spectra shown demonstrate how the spectral features of different sample components can be used to derive image contrast.*

particular details, getting the same result with a different method can be used as a tool to confirm results and conclusions.

While this is not the place for a complete description of the theory, uses, and ramifications of all possible processing treatments that can be applied to chemical imaging data, the majority of NIR chemical images recorded to date have been obtained from solid or powder samples, measured predominantly by diffuse reflectance. As such, the following discussions may serve as a general outline of the more common procedures and their applications to these types of imaging data.

8.4.3 Spectral correction

As for all absorbance-based spectral measurements, the intensity data represented in a raw (single-beam) chemical image are a combination of the spectral response of both the instrument and the sample. In order to remove the instrument response component, it is necessary to ratio the data to a background reference. For reflectance measurements, the background is a separate data cube typically acquired from a uniform,

high-reflectance standard such as Spectralon (Labsphere, Inc., North Sutton, NH, USA) or white ceramic. Since the filter-based NIR-CI instruments operate in a true 'staring' or DC mode, it is also necessary to subtract the 'dark' camera response from the image cube. The correction to reflectance (R) is therefore:

$$R = (\text{Sample} - \text{Dark})/(\text{Background} - \text{Dark})$$

Further processing is also usually performed to transform the reflectance image cube to its $\log_{10}(1/R)$ form, which is effectively the sample 'absorbance'. This results in chemical images in which brightness linearly maps to the analyte concentration, and is generally more useful for comparative as well as quantitative purposes. Note that for NIR measurements of undiluted solids the use of more rigorous functions such as those described by Kubelka and Munk[35]are usually not required.[34]

8.4.4 Spectral preprocessing

The primary objective of preprocessing treatments is to remove the nonchemical biases from the spectral information. Scattering effects induced by particle size or surface roughness may lead to offsets or other more complex baseline distortions. Differences in sample density or the angle of presentation may induce overall intensity variations as well as additional baseline distortions. Most samples are not perfectly uniform on the microscopic scale, and these effects may dominate the initial contrast observed in an un-processed chemical image. In some cases this contrast may provide useful information to the analyst; however, it is generally removed in order to focus on the chemical information.

For physically uniform samples, simple techniques such as offset or polynomial baseline correction may suffice to minimize the physical/optical contribution from the spectral data, but there are also a number of more advanced techniques available that are quite effective. For example, it is common to convert the spectral data to a Savitzky–Golay[36] first- or second-order derivative form. This conversion minimizes the baseline and random noise contributions while highlighting subtle band shape and wavelength position information, often aiding in the resolution of overlapping bands. It also has the benefit of preserving the relative band intensity information necessary for further quantitative treatments of the data. For powder samples and tablets, multiplicative scatter correction (MSC)[37] can be very effective in correcting baseline as well as any intensity artifacts due to scattering. It may be used by itself, or as a preliminary step to taking spectral derivatives.

Spectral normalization techniques, such as mean-centering and/or normalizing the spectral intensities to unit variance, are used to scale spectral intensities to a common range. Such treatment will virtually eliminate baseline offsets and physical intensity variations, which is useful in comparing spectra with widely varying overall absorbance properties. However, rescaling will also distort the quantitative information contained in the spectral intensities; therefore this preprocessing approach should be used with caution if quantitative assessments of the individual spectra will be required during further analysis.

8.4.5 Classification

Classification is the process of using analytical techniques to identify and group similar chemical species within a sample. More on this topic can be found in Chapters 12 and 14 of this book. The result is a chemical image that visualizes regions of spectral and therefore chemical similarity within the sample. In some cases a single image plane at the wavelength of a characteristic absorption band will readily display the spatial distribution of a particular species of interest. This is called a *univariate* approach and is the simplest and most direct way of probing the sample data. Although fast and intuitive, this method requires that the spectra of the individual sample components be readily differentiable at a single wavelength. Univariate

analysis often falls short of discerning finer distribution details in more complex samples, and examination of just a few disassociated image planes does not take advantage of the tremendous information content of the complete hypercube. Alternatively, *multivariate* approaches utilize all or distinctive regions of the spectral range in order to better distinguish and deconvolve overlapping spectral responses. This strategy takes advantage of a significantly higher proportion of the imaging data, and can successfully isolate individual sample components, even in very complex samples.

Multivariate analysis of single point NIR spectra has become a mainstay for a wide range of pharmaceutical applications.[5,6] Single point methods are generally based on a relatively small number of individually collected reference spectra, with the quality of a method dependent on how well the reference spectra model the unknown sample. With a robust method, highly accurate concentration estimates of sample components can be extracted from unknown mixture spectra.

Multivariate analysis tools are available in many commercial software packages, and in general, one does not have to be an 'expert' to utilize them effectively. However, a general familiarity with the basic assumptions and limitations of the algorithms at hand is useful in both the selection of appropriate methods and the interpretation of the results. Many good texts are available which discuss these methods from the fundamental mathematical origins,[38,39] through general spectroscopic,[40–43] and imaging applications.[44]

The multivariate tools typically used for the NIR-CI analysis of pharmaceutical products fall into two main categories: pattern recognition techniques and factor-based chemometric analysis methods. Pattern recognition algorithms such as spectral correlation or Euclidian distance calculations basically determine the similarity of a sample spectrum to a reference spectrum. These tools are especially useful for images where the individual pixels yield relatively 'unmixed' spectra. These techniques can be used to quickly define spatial distributions of known materials based on external reference spectra. Alternatively, they can be used with internal references, to locate and classify regions with similar spectral response.

Factor-based chemometric methods are often more appropriate for the analysis of images where there is significant spectral mixing in the individual pixels, as these algorithms are specifically designed to account for the varying spectral contributions of multiple materials in individual (pixel) measurements. Principal components analysis (PCA) is an 'unsupervised' algorithm meaning that it does not use a predefined set of references or a spectral library to interrogate the sample spectra set. Instead, this method defines the significant sources of variance (spectral and physical) within the sample set as a series of ranked components or factors and assigns each spectrum (pixel) a score (brightness) based on the relative contribution of each factor. PCA analysis is often a very efficient method for highlighting subtle spectral variations between different objects or regions in a sample scene. It can also serve as an efficient preliminary analysis step by pointing the user to significant spectral and spatial regions that may be further analyzed by other methods.

Directed chemometric algorithms such as principal component regression (PCR) and partial least squares (PLS) analyses utilize a predefined set of reference spectra (a library or calibration set) to discriminate discrete spectral contributions attributable to each of the materials. Traditional quantitative NIR applications employing these methods are typically calibrated using a set of reference samples with a range of varying component concentrations to encompass that expected from the unknown samples. However, constructing reliable calibration data sets for chemical imaging is much more difficult, since each pixel must see exactly the same composition. Achieving homogeneity at the distance scale of NIR imaging (30 μm) is challenging, and analysts can find themselves in the unenviable position of needing to produce highly homogeneous calibration samples. Since producing highly homogeneous samples is often the problem in the first place, this can be an unsolvable task.

As an alternative, qualitative 'classification' models are developed from a library composed of pure component spectra. It is considerably easier to generate an imaging data set of a pure component, and these data enjoy a certain 'statistical robustness' due to the large number of individual pixels representing the sample. The resulting models are then applied to the sample image data to produce score predictions that

relate to the abundance of each of the materials in the individual measurements. The intensity of each pixel is determined by how much of that component is predicted to be present at that spatial location based on the reference spectra. So for a score image based on the API library component, the brighter the pixel, the more API content predicted at that location. Simple observation of images may show obvious qualitative differences between samples, however as with other analytical methods, there needs to be a way to compare the data in a quantitative manner to allow further correlations and possible performance predictions to be evaluated. A statistical analysis of the distribution of these pixel values provides an objective and reproducible measure of sample heterogeneity.

In addition to this statistical approach where the spatial distribution of these data points is ignored, the same scores images can also be evaluated morphologically, in which the spatial information is analyzed. The following two sections provide strategies for deriving quantitative and reproducible metrics from imaging data, the first without consideration to the spatial location of the spectra, and the second using morphological analysis of segmented images.

8.4.6 Image processing – statistical

The pixel-to-pixel contrast in a single image plane provides information on the spatial distribution of chemical components in a sample. Contrast in an image can be based on any number of parameters: absorbance, score value, correlation coefficient, etc., and although an image is a highly intuitive representation of complex information, it is not the best format for objective and reproducible analysis. Figure 8.3 shows a schematic of an image portrayed as a histogram, in which the inherent pixel-to-pixel variation in the image is recorded along the *x*-axis, with the *y*-axis containing the number of pixels that have a particular range of values (bin). It is interesting to note that the spatial location information of the individual spectra is ignored, that is, the histogram is 'blind' to where the spectra come from on the sample. What matters is simply that there are tens of thousands of spectra, and a statistical analysis of their distribution can be used to describe physical characteristics of a sample. This representation enables reproducible and objective analysis of images.

Standard statistical values (mean, standard deviation, skew and kurtosis) of chemical images can be related to physical characteristics of the sample. The mean value of the distribution is equivalent to what can be generated through traditional single point NIR spectroscopy, where the spectral data across the sample is averaged into one spectrum. In the example of a classification PLS score image this mean value can provide an estimate of the contribution or abundance of each component over the image. Although an interesting parameter for analysis, the novel value of imaging data can be found in the variation of the tens of thousands of spectra around the mean, not the mean itself.

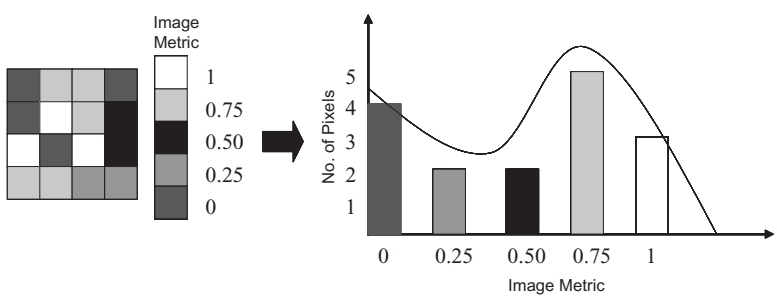


Figure 8.3 *Schematic of an image, showing the relationship between pixel intensity and frequency distribution (histogram.)*

The shape of a distribution can be described by the standard deviation, skew and kurtosis values. For any type of image based on sample abundance, such as an absorbance or PLS score image, these values can be indicative of sample heterogeneity. The standard deviation describes the spread about the mean value, and highly heterogeneous samples would be expected to have high standard deviations. The standard deviation though, is not sensitive to small populations of outliers. Skew and kurtosis describe deviation from a normal distribution, and can give insight into the characteristics of relatively small pixel populations at the extremes of the distribution. The skew value describes any asymmetry in the tailing of the distribution. A negative skew indicates an increase of lower value pixels, and can indicate the presence of 'holes' in the image where there might be none of that particular component. A positive skew value indicates an increase in the population tail at the high end of the scale, and can be indicative of 'hot spots,' which are regions with a very high abundance of that particular component. The kurtosis value describes the symmetry of distribution tailing, commonly described as 'peakedness' of the distribution. A positive kurtosis value relates to a sharper peak with longer tails (Lorentzian shape) and a negative kurtosis value relates to a flatter peak with smaller tails (mousehole shape). A negative kurtosis value means the distribution does not have any tailing, whereas a positive kurtosis shows that it does.

The simplified histograms shown in Figure 8.4 give examples of where there is a homogeneous distribution (Figure 8.4a) and a heterogeneous distribution (Figure 8.4b). The mean values are very similar for the two distributions, and these samples would be equivalent when analyzed by single point NIR spectroscopy or HPLC, methods with no spatial resolution. Obviously though, these distributions are significantly different, and these differences can be described statistically. The standard deviation for Figure 8.4b is higher than Figure 8.4a, and as there is a significant tail to the right, its skew value would also be higher. The kurtosis for Figure 8.4b is negative, as the shape of the histogram distribution is flatter and spread out, showing that regions of pixels in the Figure 8.4b image have higher contribution to the image metric (higher component concentration, for example, if this was a PLS score image) than those in Figure 8.4a. This is an idealized situation, where the images are significantly different. For most real samples though, differences are much more subtle, and looking at an image may not provide a clear cut answer. The ability to represent images as a histogram enables statistical values to be related to physical sample characteristics, and is often useful for defining subtle differences between similar samples.[17] Images (and therefore samples) can be

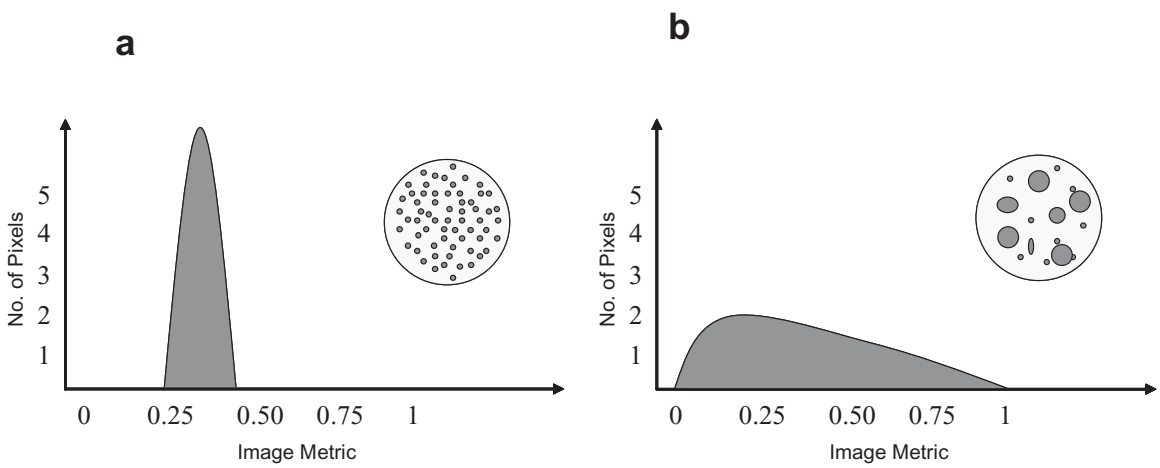


Figure 8.4 *Schematic histograms demonstrating (a) homogeneous chemical distribution, (b) heterogeneous chemical distribution. Areas of different gray tones represent areas of different chemical composition.*

compared and 'rated' based on these relatively simple but powerful parameters. Examples of this type of statistical analysis will be shown in the applications sections of this chapter.

8.4.7 Image processing – morphology

One of the goals of image analysis is to produce a 'picture' depicting the distribution of one or more components. After the chemical contrast between the individual pixels has been revealed using the classification techniques described previously, it can be displayed more effectively through the choice of an appropriate color map and intensity scaling of that map. For example a gray-scale map can be used to simply convey a range of infrared absorbances, or a pseudocolor scheme may aid the eye in associating different regions of similar intensities. The visualization quality of the image may also be improved by masking or truncation to remove irrelevant regions of the image. Conversely, concatenation of multiple images can assist in the direct comparison of samples collected in different measurements. Another useful technique for the visualization of the distribution of multiple components simultaneously is to produce a red–green–blue (RGB) composite image. In this method each color of the map represents the contrast developed for a separate component.

Another image processing method that is especially applicable to many pharmaceutical sample images is particle statistics analysis. This technique works well when the individual component distribution appears in the form of discrete domains in the processed image. Such domains may represent actual particles of material, but more often are indicative of agglomerations of the components either intentionally or inadvertently produced by the manufacturing process. The particle analysis tools permit the analyst to quickly characterize the statistical distribution of the domain sizes based on particle area or diameter. Knowing the original particle size of blend ingredients, for example, an analysis of the size of corresponding domains within a tablet can provide insights into the agglomeration process during blending. By looking at a histogram of domain sizes, it is often possible to characterize the agglomeration of agglomerates. In other words an analysis of domain sizes may indicate three distinct populations, with the larger domain sizes being integer multiples of the smallest domain sizes.

Moving beyond a simple analysis of domain size, the location of domains in a tablet can be even more critical in determining sample performance. For example, whether or how excipients are associated with an API can impact dissolution properties of a sample. NIR-CI is able to characterize 'colocalization' of sample domains in a tablet, and can therefore be a powerful tool in correlating this characteristic with product performance. Coatings can be considered an extended or continuous agglomeration, and NIR-CI can be used to determine *a priori* coating thickness values, as will be presented in Section 8.5.4. And interestingly, NIR-CI is a primary (not secondary) analytical technique with respect to characterizing domain shapes, orientations, and relative separation distances. This information can be vital in the elucidation of product performance differences between samples that have identical chemical assays and can be used to relate the impact of the different unit operations on the composition of the final dosage form.

8.5 PAT Applications

In this section the authors would like to focus on the attributes of NIR-CI that make it such a powerful tool for PAT applications. Chemical imaging approaches provide tremendous capabilities due to the high throughput nature of the technique. For example, 'self-calibrating' content uniformity measurements can be performed rapidly, counterfeit products can be identified amongst real products, intact blister packs can be screened for quality assurance, and trace amounts of contaminants, including different polymorphic forms can be identified and visualized within finished products. This is in addition to performing 'typical' imaging

applications such as investigating the spatial and chemical structure of dosage forms for root cause analysis of manufacturing problems and determining CQAs during formulation development.

8.5.1 Content uniformity measurements – 'self calibrating'

In this example, the field of view (FOV) of the experiment encompasses a large area of ~9 × 7 cm. This large FOV enables a matrix of samples (20 tablets and 3 pure components) to be simultaneously evaluated. The goal for this example is intersample comparison, using the pure component samples as a built-in calibration set. The capability to simultaneously view both samples and pure components is unique to array detector-based chemical imaging instrumentation. Data were collected using a Sapphire (Malvern Instruments Inc., Westborough, MD, USA) chemical imaging system in the spectral range of 1100–2450 nm at 10 nm increments with 16 frames co-added. Each pixel (81 920 in total) samples an area of ~275 × 275 μm, so the entire FOV covers ~9 × 7 cm.

To demonstrate the ability to evaluate intersample variations, an over-the-counter (OTC) pain relief medication from two different manufacturers was compared. The samples contain three APIs each: acetaminophen, aspirin and caffeine. Pure acetaminophen, aspirin and caffeine samples are obtained in either tablet form or powder compact and included within the same FOV as the tablets to provide simultaneous reference materials for the tablet samples. The tablets and pure components were arranged as shown in Plate 8.1a. Measurements on all samples were collected simultaneously. Tablet A samples from one manufacturer have a reported label concentration of 37%, 37%, and 10%, for the three API components, respectively. Tablet B samples from the second manufacturer contain the same three APIs, at label concentrations of 39%, 39%, and 10 %, respectively. In addition to these samples, tablet C samples are included in the array of tablets. These samples contain only acetaminophen as the API with a reported label concentration of 79%, and are made by the manufacturer who produces tablet A. The remaining mass of all three tablet types represents the excipient (binder, disintegrant, and lubricant) materials.

The raw images were dark and background corrected as described previously (Section 8.4.3) using Spectralon (Labsphere, Inc., North Sutton, NH, USA) as the reference, and converted to $\log_{10}$ ($1/R$). A Savitsky–Golay second derivative was performed to eliminate contrast due to baseline offset and slope, and the resulting spectra mean-centered and normalized to unit variance. A spectral reference library for the experiment was created by selecting pixels from the pure component materials. A PLS2 model was constructed from the reference library and used to generate PLS scores images for each of the library components.

Plate 8.1b shows the RBG image created from the PLS scores images, with acetaminophen assigned to the red channel, aspirin to blue and caffeine to green. The channels are scaled independently, and only represent qualitative information about the component distributions. This image highlights the advantages of having reference samples in the same FOV. The results can immediately be visually confirmed – the acetaminophen pure sample is solid red – indicating that the PLS score is accurately identifying that component. The same holds true for aspirin (blue), and caffeine (green). Additional confirmation is that the samples labeled tablet C show all red with no blue or green, verifying that the model is correctly classifying acetaminophen. Creating single data sets that contain both the sample and the reference materials simplifies many aspects of the experiment. The data set is 'internally' consistent because the reference and sample data are measured simultaneously by the same instrument – temporal variations in instrument response are automatically accounted for. Also since the relevant data always remain together in a single image data file, the data and reference samples are always processed together in identical ways, and cannot be separated. The reference samples may be left in place when new samples are introduced, and therefore, a new experiment simultaneously results in new reference spectra. Any changes that effect instrument response or data collection parameters result in those changes being reflected in the reference sample spectra. This obviates

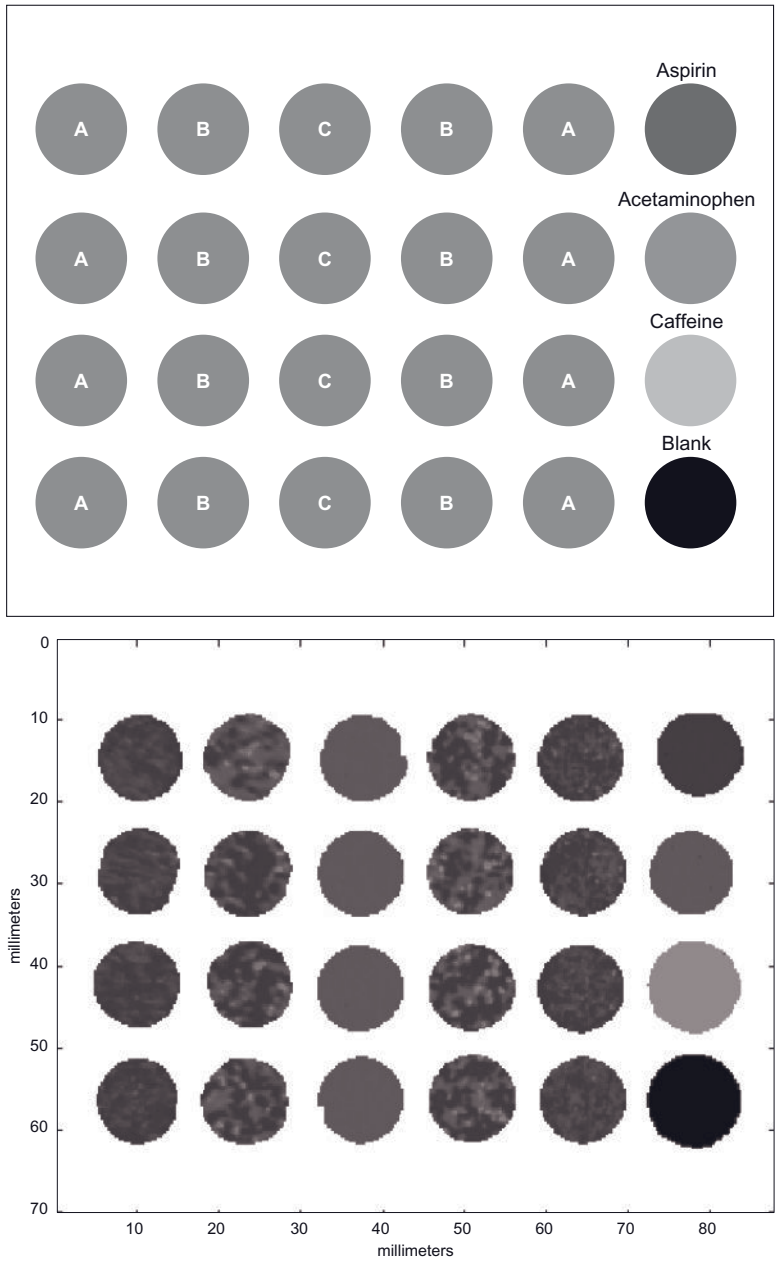


Plate 8.1 *(a) Key to placement of tablet matrix and pure component reference samples for high-throughput inter-tablet comparison. Tablet groups A and B are all tablets of an OTC pain relief medication, with A and B from two different manufactures. The tablets contain three APIs: acetaminophen, caffeine, and aspirin. Group C is composed of tablets that have only one of these APIs. The API pure components are positioned in the right column of the tablet matrix. (b) Summary image in which the distribution of the three APIs is shown simultaneously. Acetaminophen is shown in red, caffeine in green, aspirin in blue, and the three color channels are overlaid.*

the need to apply transfer of calibration procedures to analyze new samples. Moving beyond qualitative information, quantitative information can be gleaned by examining the statistical properties of these tens of thousands of spectra collected in a single data set, to allow many inter-tablet and intra-tablet quantitative comparisons to be made. Even using low (high throughput) magnification it is still possible to glean intra-tablet uniformity information since approximately 1100 NIR spectra are recorded for each of the 24 tablets.

8.5.2 Quality assurance – imaging an intact blister pack

Figure 8.5 demonstrates another example of utilizing NIR-CI as a 'high throughput' technology for use in the pharmaceutical manufacturing process. In this example the sample is a sealed blister pack approximately 7×11 cm in size containing 10 visually identical white tablets. One of the tablets was removed and replaced with one of identical form factor containing a different API. Figure 8.5A is a visible image of the adulterated package. The NIR spectral image of the complete sample scene was collected without disturbing the samples or the packaging. The PCA score image of the arrangement in Figure 8.5B clearly identifies the rogue tablet. While the result of this measurement is an image, the data can also be considered as simply a series of NIR spectra taken in parallel. We can therefore interpret the data as a quality control measurement performed on all ten tablets simultaneously, with the number of 'objects' that can be simultaneously addressed in this manner limited only by magnification and the number of pixels on the infrared array. Simple, bulk quantitative information can also be derived: 90% of the objects are correct for that blister pack, 10% of the objects do not belong.

There are numerous applications where such a high throughput system could simply replace single-point or multiplexed NIR spectrometers, by providing significant advantages in sample throughput and handling. These advantages are operative over a range of size scales and sample shapes that are difficult or impossible for current single-point instrumentation to address. One such example is the validation of the filling of clinical trial blisters. The intent is to provide either the dosage forms or a placebo in such a manner that the patient is unable to know which they are taking. In the case of double-blind trials, the dispensing physician is also unaware of the difference. In order for this to be feasible, the physical size and shape of all types of dosage forms must be indistinguishable. Errors can void a clinical trial, imposing significant costs in both immediate resources and lost time to market for a new drug. The ability to rapidly verify the accuracy of

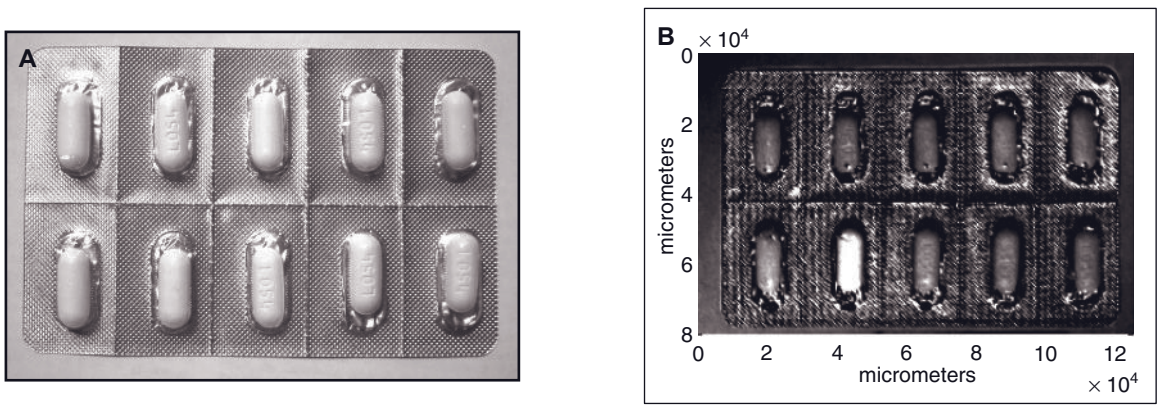


Figure 8.5 *Adulterated pharmaceutical blister pack containing a single 'rogue' tablet. (A) Visible image showing that visually the tablets are virtually indistinguishable; and (B) NIR PCA score image highlighting the single tablet that has a different chemical signature.*

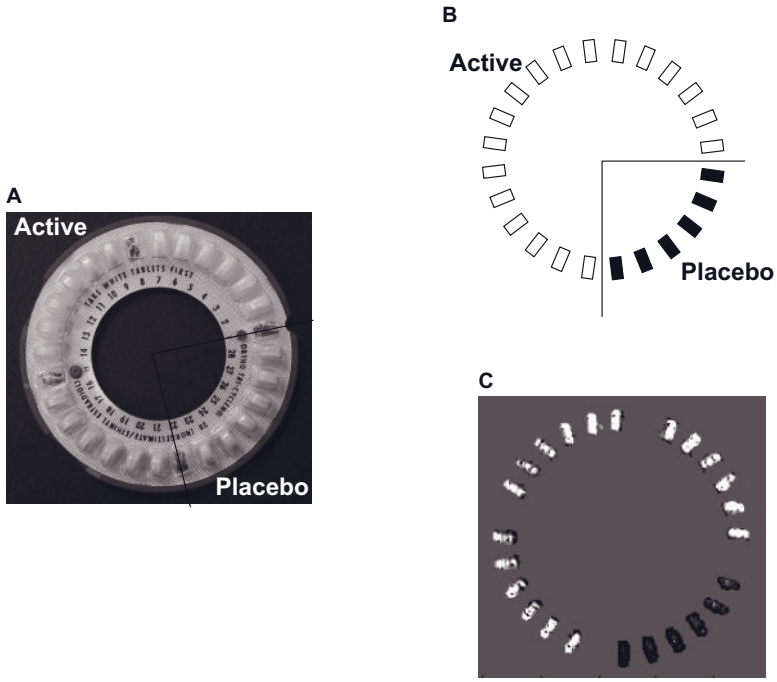


Figure 8.6 *Contraceptive contained in the original package, demonstrating high-throughput classification possibilities. (A) Visible image; (B) description of sample positioning; and (C) NIR PCA score image showing contrast between the placebo and the active-containing tablets.*

the packaging process after the product is sealed therefore offers significant benefit. In many cases this array based approach can be optimized so that only a few carefully selected wavelengths are needed for sample identification and confirmation, yielding an inspection time for a complete sealed package on the order of seconds. The ability to perform this measurement in close to real time is critical if 100% inspection is the goal. In addition, the large, flexible FOV of an array based approach enables a variety of package shapes and orientations to be analyzed with little or no reconfiguration. Figure 8.6 shows the ease with which difficult packaging problems are handled with an imaging approach. Unlike a single point or multiplexed system, the redundancy in the number of detector elements makes the system relatively insensitive to complex spatial patterns or sample placement reproducibility.

8.5.3 Contaminant detection

Chemical imaging can also be used to detect low concentrations or weakly absorbing species. It contrast to bulk or macro-scale spectroscopy where the spectral signature of a contaminant is lost in the overall matrix spectral contribution, microscopic approaches can provide highly localized spectra, effectively increasing the relative concentration (and thus, the detectability) of a scarce component in the sampled area. For mapping approaches, this requires the selection of the 'correct' location for detection. Therefore, one must either have prior knowledge of the location of the impurity, or it must be found by time-consuming mapping techniques. NIR-CI collects spatially organized microspectra from large areas of the sample in a single experiment. The full area of a typical 1 cm diameter tablet sample can be interrogated with a spatial resolution of about $40 \times 40\,\mu$m in only a few minutes, providing both enhanced sensitivity and location information.

Even if a rogue material is detected but is unidentifiable by its NIR spectrum alone, the location information derived from the imaging experiment can greatly assist in the rapid analysis of the suspect region by other analytical methods.

To illustrate this point, two examples are presented: the first identifying a low level impurity in a tablet, and the second, detecting low levels of a particular polymorphic form. Figure 8.7 summarizes the NIR-CI analysis of a tablet that contains a single API (furosemide) in a mixture of excipients. Furosemide has been shown to degrade with exposure to light,[45] resulting in a finished product that also contains a low-level contaminant – the degradation product. A PLS2 model was constructed using library spectra of pure components, acquired separately. The resulting PLS score image for furosemide is shown in Figure 8.7A. The

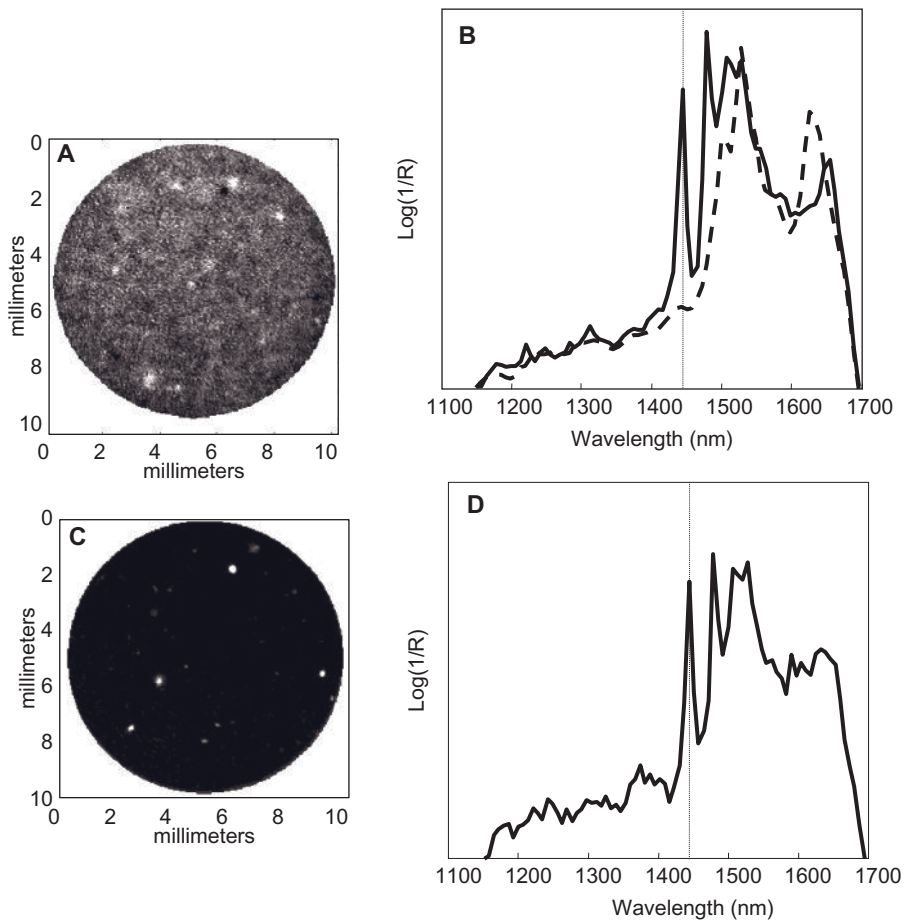


Figure 8.7 *Trace contaminant analysis of furosemide and its breakdown product in a tablet. (A) PLS score image highlighting furosemide-rich areas in the tablet; (B) dashed line is the pure API spectrum. Solid line is spectrum of the degradation product. The dotted vertical line on the graph shows that the largest spectral difference occurs at 1444 nm; (C) normalized and scaled single channel image at 1444 nm highlighting impurity rich areas. The impurity concentration derived from sample statistics is ~0.61%. (D) Solid line spectrum is a residual microspectrum observed at the center of one of the bright spots after subtracting the variance due to the predominant chemical component (the excipients) from the entire data set. This spectrum compares well to the spectrum of the degradation product in (B).*

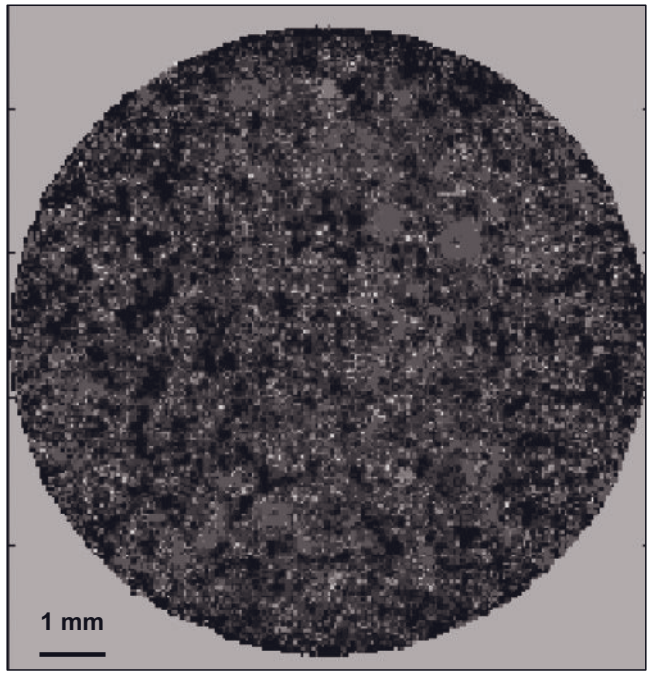


Plate 8.2 *Trace analysis of polymorphic form 1 of ranitidine. The RGB summary image highlights the distribution of sample components: polymorphic form 2 of ranitidine in red, polymorphic form 1 of ranitidine in green and the excipients in blue.*

NIR spectra of the API and the degradation product are overlaid in Figure 8.7B. The latter exhibits a prominent spectral feature at 1444 nm, which readily distinguishes it from both the API and excipient materials. Since the impurity is expected to occur in very low concentrations, the spectral contribution from the drug material was enhanced in the image through a scaled subtraction of the spectrum of the bulk (excipient) material from each pixel in the original hypercube. The corrected single-channel image at 1444 nm of the residual data clearly shows several small areas of increased intensity distributed across the sample area (Figure 8.7C). Single pixel residual spectra from these regions closely resemble the known spectrum of the breakdown product (Figure 8.7D) and verify its presence in the sample tablet. The overall abundance of the impurity in the sample was calculated through pixel-counting statistics and estimated to be approximately 0.6% of total tablet area.

In other experiments, we have also used this approach with some success to determine the existence and location of low levels of different polymorphic forms within finished products. Plate 8.2 shows an image of a sample containing both polymorphic forms (1 and 2) of the API ranitidine. The overall composition of the tablet is 75% ranitidine (both forms) and 25% excipients. Plate 8.2 highlights the form 1 domains in green, whose abundance is estimated to be ~1.2% of the total tablet weight. This information is derived from the analysis of a single tablet.

8.5.4 Imaging of coatings – advanced dosage delivery systems

Coatings are used for a number of difference purposes. For example seeds may be coated with pesticides to increase the yield of a crop, tissues may be coated with lotion to decrease irritation for the user and

pharmaceutical solid dosage forms have coatings to perform a variety of functions. These range from aesthetic to the control of the drug release into the body. Variations in the coating integrity and/or thickness could compromise these controlled delivery and bioavailability mechanisms. Current methods of coating assessment are typically destructive to the sample so if any discrepancy were found in the results, the sample would not be available for further analysis.

The NIR-CI analysis of an extended-release dosage form is shown in Figure 8.8. A sample of the microspheres was removed from a capsule and imaged with the Sapphire NIR chemical imaging system. The corrected single channel image at 2130 nm (Figure 8.8A) indicates that there are two 'types' of spheres

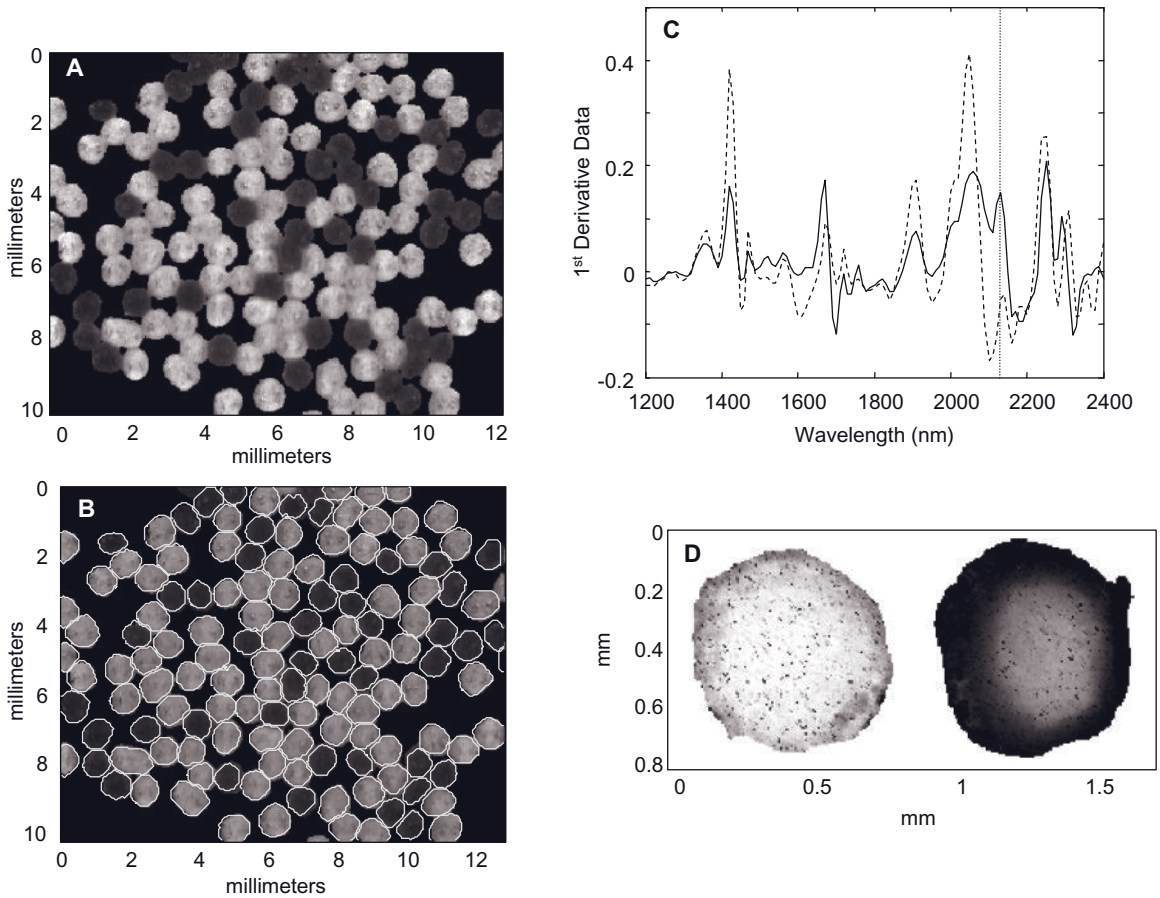


Figure 8.8 *Microbeads from extended-release dosage form. Whole beads at low magnification showing two distinct "types." The contrast is derived from differences in their NIR spectra. (A) Single-channel image at 2130 nm highlights differences between beads; (B) single-pixel microspectra of the two bead 'types.' The dashed line spectrum is from a "dark" bead and the solid line spectrum is from a 'bright' bead. The dotted vertical line on the graph shows that the largest spectral difference occurs at 2130 nm; (C) image showing intensity-based particle contours from which bead size statistics may be determined; and (D) high-magnification chemical images showing spatial difference in the composition of the two 'types' of beads identified in (A). The left- and right-hand beads correspond respectively to the 'dark' and the 'bright' beads from (A).*

present in the sample, each having a distinctly different spectral response (Figure 8.8B). The solid line spectrum is from a 'bright' bead, and the dashed line spectrum is from a 'dark' bead. Quantitative information about the dosage form was obtained from the image by simply setting the appropriate intensity level so as to define each of the observed spheres as discrete particles (Figure 8.8C). Automated image analysis indicated that there were twice as many 'bright' (91) as 'dark' (45) spheres in this particular FOV, and that the 'bright' spheres are, on average, approximately 5% larger in diameter.

The release mechanism of the form was elucidated through further microanalysis of the individual sphere types. One of each 'type' of sphere was cut in half and the cut faces were chemically imaged at higher magnification. The resulting images (Figure 8.8D) reveal that the 'bright' sphere is made up of a coating that surrounds a core composed of the same material found in the 'dark' spheres, accounting for the increased size of these spheres.

Image analysis during the development and scale-up of the process that produces the coated spheres would provide a clear picture of this component and allow the 'critical control parameters' to be quickly identified, improving quality and reducing the risk of costly performance failures. During manufacturing, chemical imaging could be applied to monitor the mean and standard deviation of coating thickness uniformity, the mean and standard deviation of the microsphere diameter, as well as the relative composition of mixtures of different types of microspheres. These data could be used to enable the desired composition to be optimized and maintained.

A slightly different approach to coating characterization was taken to investigate an OTC product designed to release the active component in a certain region of the gastrointestinal tract. Polarizers on the illumination sources of the Sapphire NIR-CI system were used to minimize glare from the curved, shiny sample surfaces. Images, spectra and histogram representations of the images are presented in Figure 8.9. The image of the two tablets (Figure 8.9) is at 2070 nm, where the coating has a strong absorbance band. The tablet on the right side of this image has a region of decreased brightness, and spectra from this low intensity region are superimposed over spectra from outside of this region, and displayed in Figure 8.9b. Since the absorbance values of this spectral feature correspond to the amount of coating material, brighter pixels correspond to regions with increased coating thickness. Variation in pixel intensity in the image corresponds to coating thickness variation, and the low intensity area is a region on the tablet with significantly thinner coating. Figure 8.9c and d present the image intensities of the two tablets as histograms, so that a statistical comparison of the two samples can be made, presented in Table 8.1.

As discussed in Section 8.4.6, the standard deviation (SD) is a measure of the spread of the data about the mean and provides an indication of the overall uniformity. The SD is slightly lower for the left tablet than the right (0.046 versus 0.071), indicating that the left tablet coating is more homogeneous, which corresponds to the visual appearance of the two tablets. A significant difference is seen in the skew values (–0.4 versus –1.5), with the right tablet having a significant negative value, indicative of area(s) of lower concentration (thinner coating). Pixels having significantly lower absorbance values at 2070 nm are circled in Figure 8.9d. This statistical analysis of the data provides quantitative metrics that describe the intuitive and qualitative observations available through the NIR chemical images.

Table 8.1 *Statistical comparison of the analysis of the two tablets shown in Figure 8.9*

	Mean	Standard deviation	Skew
Left tablet	2.3	0.046	–0.42
Right tablet	2.4	0.071	–1.49

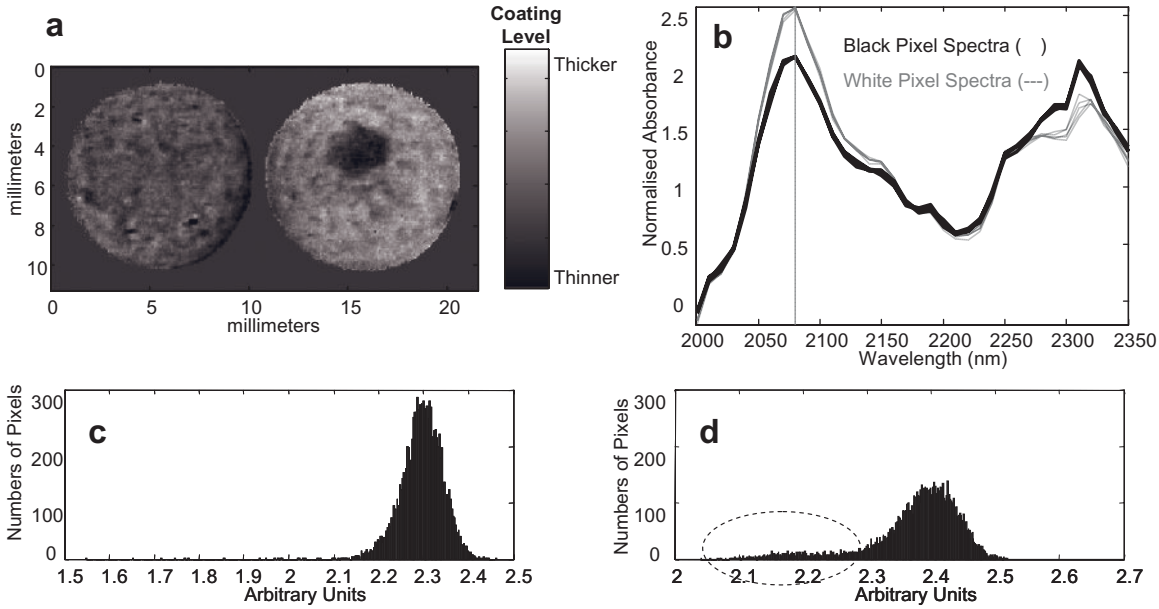


Figure 8.9 *(a) Single wavelength NIR chemical image at 2070 nm showing tablet coating differences between two of the same type of OTC tablets. The spectra in (b) are from the high intensity/thick coating area (in white) and the low intensity/thin coating (black) regions of the image. (c) and (d) are the histogram distributions for the two tablet images in (a), and present the distribution of the pixel intensities. The low-value population (circled) of pixels in (d) highlights the area of thin coating in the tablet on the right in (a).*

NIR-CI can also be utilized to directly measure coating thickness and uniformity in a single sample or multiple samples in the same image. The same coated sample was cross-sectioned to allow measurement of the coating thickness around the tablet. Figure 8.10 summarizes the results. Because the core and coating have significantly different spectra, and the coating is highly localized, a single wavelength NIR image at 2070 nm where the coating material has a strong absorbance, clearly highlights the coating (Figure 8.10a). In order to perform a morphological analysis on the coating alone, the region of interest was chemically segmented through the creation of a binary image, shown in Figure 8.10b. The coating can be symmetrically segmented into single pixel lines, which are measured to produce a thickness image and the subsequent thickness distribution (Figure 8.10c and d, respectively). The mean of the distribution in Figure 8.10d, 0.26 mm, represents the mean coating thickness; whereas the standard deviation shows the variation in coating thickness across the whole sample. The distribution also provides the minimum and maximum values of coating thickness, and that can be very valuable information as well.

It is clear that the spatial element of NIR-CI combined with the flexible and forgiving nature of the optics to nonflat surfaces is invaluable in providing information for both the manufacture and development of coated products. In the examples shown it was possible to identify defects in the coating as well as gaining a direct measurement of the thickness range for a complex dosage form. This information allows a better understanding of the coating on the product and allows identification of critical to quality parameters that require control and hence minimize the risk of performance failures. In this case the skew or standard deviation could be monitored for the intact tablets to help identify any defectively coated product to maintain the quality and performance.

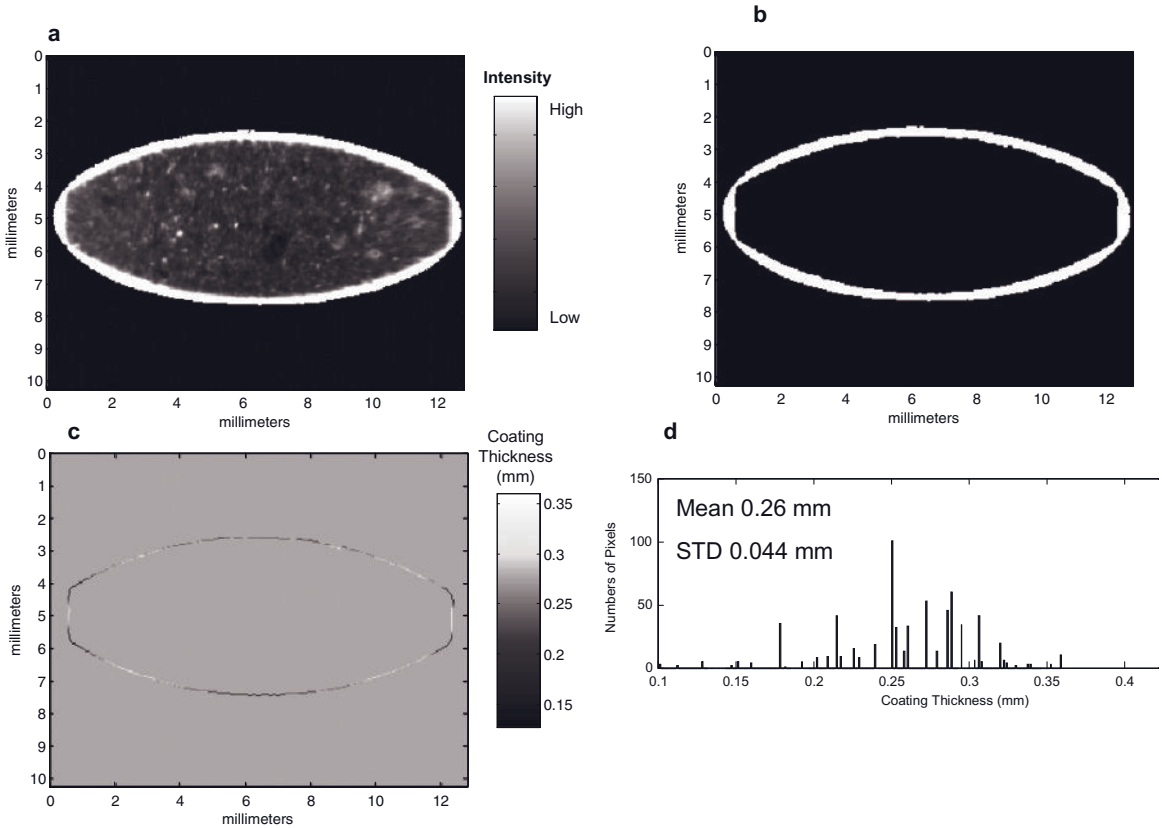


Figure 8.10 *Single wavelength NIR chemical image (a) of the cross-sectioned tablet at a wavelength that highlights the coating component; (b) corresponding binary image; (c) thickness image calculated from the binary image; and (d) the thickness distribution histogram from the thickness image.*

8.6 Processing Case Study: Estimating 'Abundance' of Sample Components

One of the unique capabilities of chemical imaging is the ability to estimate the 'abundance' (an approximation of concentration) of sample components that comprise a heterogeneous pharmaceutical sample *without* developing a standard concentration calibration model. NIR-CI can sample an entire tablet surface with >80 000 spectral data points. For a spatially heterogeneous sample, a precise calibration model may not be necessary, *because* of this density of sampling. The classification of 80 000 spectra across the sample surface provides a statistically robust measurement of the distribution of sample components. For tablets that are extremely heterogeneous relative to the experimental magnification (typically ~40 μm/pixel), sample components appear to aggregate in domains of pure components. A single pixel spectrum can be examined, and classified as belonging to one species or another. For pharmaceutical samples that are well blended, a better approach is to use a multivariate algorithm to determine the mixture of components for each pixel in the data set. When summarized across all pixels, the contribution of each sample component can be determined. In this example, the abundance of three APIs is calculated for an OTC pain relief medication. A multivariate analysis gave the best results when comparing the estimated composition to the label concentrations.

8.6.1 Experimental

The imaging data were collected over the spectral range of 1200–2400 nm at 10 nm increments using a Sapphire NIR-CI system (Spectral Dimensions, Inc., Olney, MD, USA). The size of FOV was 13×10 mm, each pixel sampling an area of approximately $40 \times 40\,\mu$m.

The concentrations of the active ingredients as reported from the manufacturer's label are 37% acetaminophen, 37% aspirin, and 10 % caffeine. The remainder of the tablet mass represents the excipient (binder, disintegrant, and lubricant) materials. Pure acetaminophen, aspirin and caffeine samples are obtained in either tablet form or powder compact and used to obtain reference spectra of pure components.

8.6.2 Spectral correction and preprocessing

The data cube was converted to $\log_{10}(1/R)$ as described in Section 8.4.3 on data analysis, using background and dark cube correction. As described above, preprocessing steps minimize effects due to optical or physical differences in the sample. Differences in packing density or particle sizes create baseline offset and slope differences, just as for single point spectroscopy. In chemical imaging techniques, other physical issues such as nonuniform lighting also can influence the data. For chemical imaging, physical contributions to spectra actually impact the contrast in resulting images. The naïve user that does not perform appropriate preprocessing can be fooled into thinking that image contrast is chemical in nature, when actually it is physical. In this example, to eliminate baseline offset and slope effects, a series of preprocessing algorithms was applied. First, a Savitsky–Golay second derivative was taken, then the data were mean centered. Finally, the spectra were normalized to unit variance. It is worth noting that the strong NIR positive-going bands in $\log_{10}(1/R)$ spectra become strong negative-going bands in second derivative data.

8.6.3 Analysis

Traditional macroscale NIR spectroscopy requires a 'calibration' set, made of the same chemical components as the target sample, but with varying concentrations that are chosen to span the range of concentrations possible in the sample. A concentration matrix is made from the known concentrations of each component. The PLS algorithm is used to create a model that best describes the mathematical relationship between the reference sample data and the concentration matrix. The model is applied to the 'unknown' data from the target sample to estimate the concentration of sample components. This is called 'concentration mode PLS'.

These same analysis techniques can be applied to chemical imaging data. Additionally, because of the huge number of spectra contained within a chemical imaging data set, and the power of statistical sampling, the PLS algorithm can also be applied in what is called 'classification mode' as described in Section 8.4.5. When the model is applied to data from the sample, each spectrum is scored relative to its membership to a particular class (i.e. degree of purity relative to a chemical component). Higher scores indicate more similarity to the pure component spectra. While these scores are not indicative of the absolute concentration of a chemical component, the relative abundance between the components is maintained, and can be calculated. If all sample components are accounted for, the scores for each component can be normalized to unity, and a statistical assessment of the relative abundance of the components made.

A partial least square type two (PLS2) analysis was employed, based on a library of the three API pure components. Applying the model in 'classification mode' to the sample data set results in PLS score images that show the spatial distribution of the three API components.

The PLS scores images for acetaminophen, aspirin and caffeine highlighting high concentration areas for each chemical components are shown in Figure 8.11A–C. Contrast in these score images is based on the

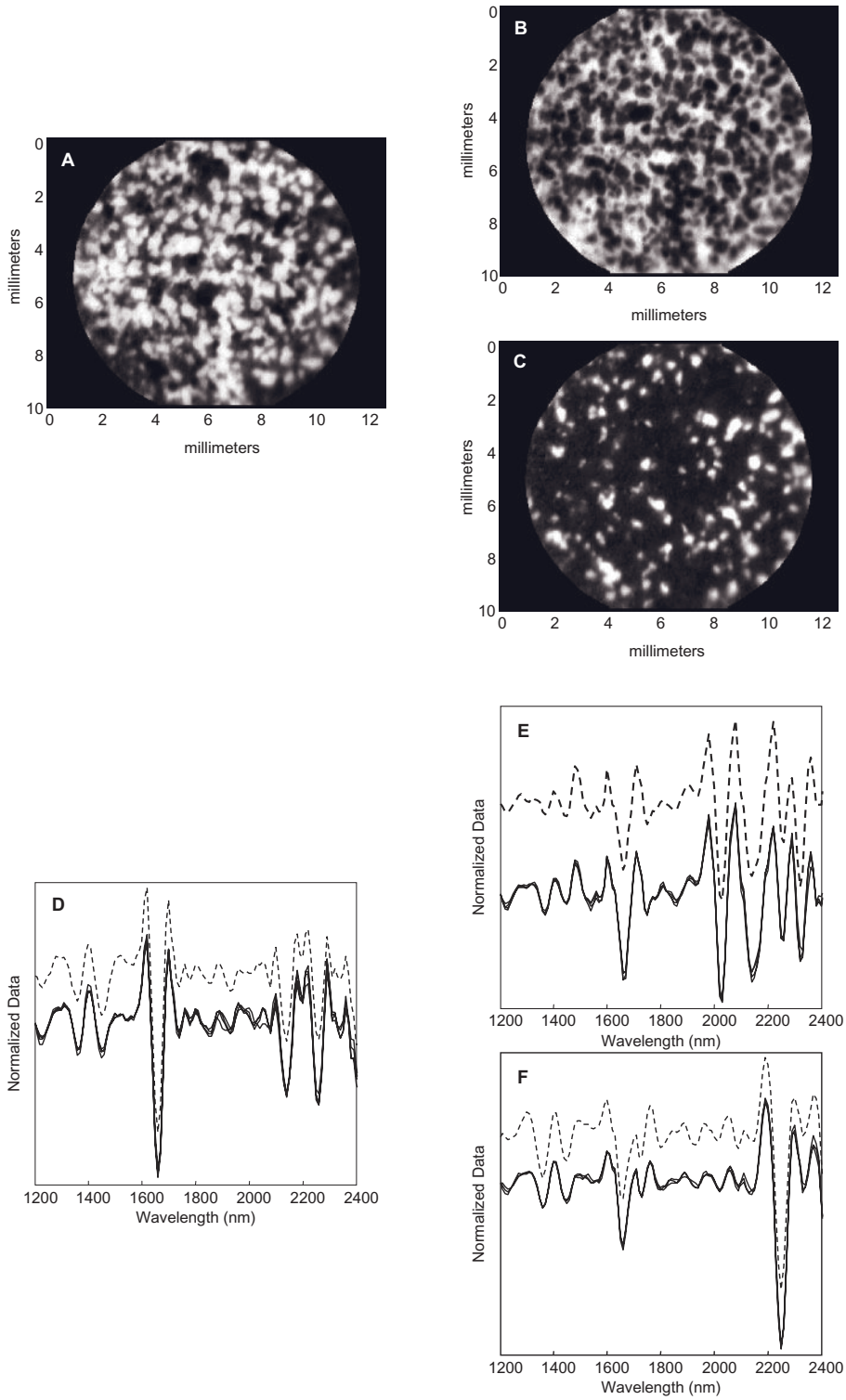


Figure 8.11 *An OTC analgesic tablet (Excedrin) with three APIs. Spatial distribution of each API was obtained using PLS analysis. (A–C) PLS score images of acetaminophen, aspirin, and caffeine, respectively; and (D–F) single-pixel microspectra (solid line) compared to pure component acetaminophen, aspirin, and caffeine spectra (dashed line), respectively.*

value of the PLS score of each pixel relative to the respective component. A pixel with high aspirin concentrations will have a high aspirin score, and will appear bright in the aspirin score image. To verify that the model is working in a qualitative sense, a comparison is made between pure component spectra of the APIs and single pixel spectra from acetaminophen, aspirin and caffeine rich areas shown in their respective score images (Figure 8.11D–F).

Qualitatively, acetaminophen and aspirin appear to be more abundant than caffeine, which concurs with the label concentration values. Caffeine shows very distinctive areas of localized high concentration domains while aspirin is relatively evenly distributed on the spatial scale of the image. Acetaminophen appears to be somewhat in the middle, showing up as large domains that blend into one another.

To quickly see how the three components overlap spatially, it is possible to create a three-color RGB image in which each component is assigned to a separate color channel. This is a useful technique to correlate the spatial distributions of components. Often pharmaceuticals are formulated so that particular components or combinations of components form spatial aggregates. Imaging can verify that this association of specific components is taking place as designed. Acetaminophen is assigned to the red channel, aspirin to blue and caffeine to green. To create the composite, each of the individual color channels is scaled independently, so any quantitative information is lost. When overlapped, they form the RGB image shown in Plate 8.3, in which none of the three components show overlapping or associative spatial distributions.

It is extremely useful to move beyond a subjective and qualitative analysis of the spatial distribution of sample components, and to begin to explore the quantitative information contained within chemical imaging data sets. One of the most powerful statistical representations of an image does not even maintain spatial information. A chemical image can be represented as a histogram, with intensity along the *x*-axis, and the number of pixels with that intensity along the *y*-axis. This is a statistical and quantitative representation of a chemical image. Since each image contains tens of thousands of pixels from across the sample, the statistical sampling is quite reliable.

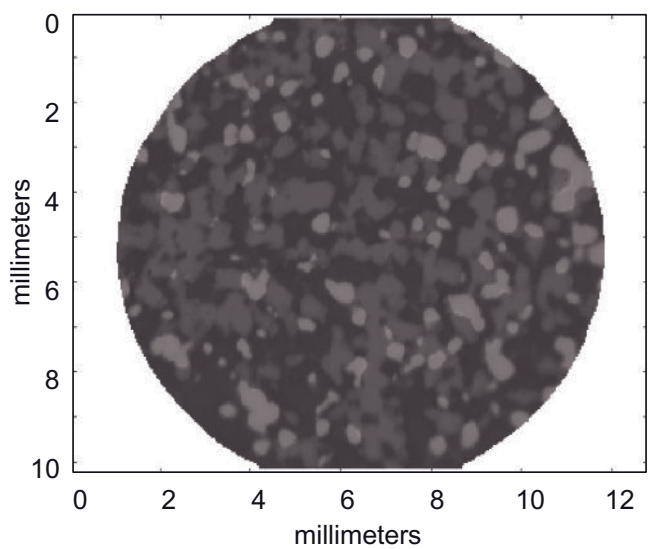


Plate 8.3 *Red–green–blue image created from PLS score images of three APIs of an OTC analgesic tablet (Excedrin). Red highlights acetaminophen-rich areas, green highlights caffeine-rich areas, and blue highlights aspirin-rich areas.*

For the acetaminophen and aspirin PLS score images, histograms show distinctive bi-modal normal distribution curves (Figures 8.12A and B, respectively). The bright pixels in the image are represented in the high intensity populations in the histograms. A threshold can be applied below which pixels are estimated to not belong to the class of interest. These thresholds are set at the inflection point between the bi-modal distributions. The threshold in the acetaminophen score image was set at a score value of 0.43. By counting the pixels above this threshold, 22 813 pixels were classified as acetaminophen. The threshold for aspirin in its corresponding score image was 0.36, and 23 058 pixels were classified as aspirin. The number of pixels that are classified as acetaminophen or aspirin ratioed against the total number of pixels in the tablet yields an estimation of abundance.

The histogram representation of the caffeine score image (Figure 8.12C) does not clearly show a bimodal distribution of pixels, even though the image shows the clearly distinctive caffeine domains. In other words, morphologically it is clear which pixels are contained within caffeine domains and which are not, but these populations are not clearly separated in the histogram. Obviously, a different approach is needed to estimate caffeine abundance. Returning to the image, caffeine domains can be selected by connecting pixels that have the same PLS score. Rather than choosing a threshold based on a statistical distribution, morphology in the image is used as the metric. Using standard image processing procedures, polygons are drawn around the caffeine domains (Figure 8.13). A less subjective threshold is now determined by looking at the image, and iteratively choosing a PLS score value so that the bright areas in the image are included in caffeine domains, and the darker pixels are excluded. In this example, pixels with a score above 0.45 are classified as caffeine containing, totaling 6755 pixels. From further investigation of the image statistics for the particles, one can determine the following information: there are 112 individual caffeine particles with a mean area of $0.097\,mm^2$, and a mean nearest neighbor distance of 0.728 mm.

In addition, taking the ratio of pixels calculated to belong to each component to the total number of pixels yields the following abundance estimates: 41% for acetaminophen, 42% for aspirin and 12% for caffeine.

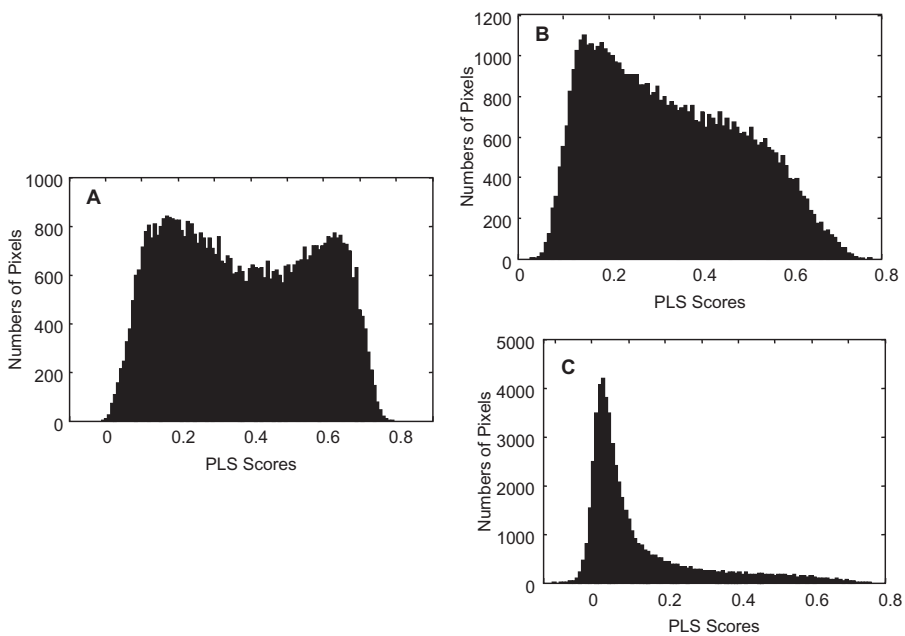


Figure 8.12 *Histograms of PLS score images. (A) acetaminophen; (B) aspirin; and (C) caffeine.*

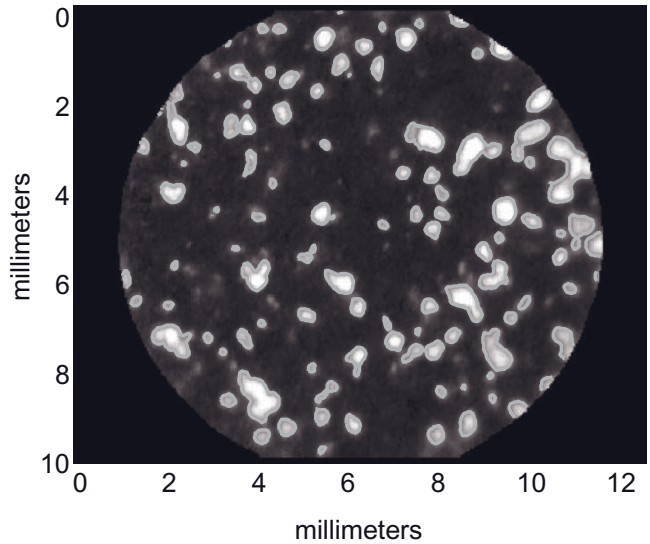


Figure 8.13 *Chemical image analysis tools use domain morphology to highlight caffeine domains in the tablet.*

There are other components in the tablet, composing the remaining 16% of the tablet. These excipients were not included in the initial model, so the results for the three API components need to be normalized to their expected contribution to the whole tablet. When normalized to a total contribution of 84%, the abundance estimations are 36% for acetaminophen, 37% for aspirin and 11% for caffeine. These values compare favorably with the manufacturers' label claim of 37% for acetaminophen, 37% for aspirin, and 10 % for caffeine.

The ability to determine the domain size of the individual components comprising a finished pharmaceutical product as shown above is not readily attainable by other analytical techniques. It is reasonable to assume that many of the physical properties of a tablet such as dissolution and hardness will be directly or indirectly impacted by how the sample components are packed – the size of a typical component domain, and how these domains interact with each other. Much of the 'art' of formulation is a function of tailoring the interactions of the sample components to bring about the desired physical characteristics of the finished product. NIR-CI can easily track the effects of changed manufacturing conditions on domain sizes and inter- and intracomponent affinities. In our experience in examining finished pharmaceutical tablets, the typical domain size of the individual components is often much larger than the manufacturers generally assume them to be. This does not imply that the tablets do not have the desired physical characteristics, but rather that the final domain sizes that give the desired properties are larger than would be estimated by knowing the size of particles of the pure components before blending.

A final 'quantitative' approach to estimating absolute abundance is through normalized mean PLS scores. This approach is totally objective, unlike the previous examples in which thresholds need to be determined. While classification mode PLS model is not designed to estimate absolute concentration, the relative abundance of the components is maintained: the higher the PLS scores, the higher the concentration. As with the abundance estimations performed using histogram thresholds, if the PLS scores are normalized, they give objective estimates of absolute abundance. In this example, as has already been shown, the three API components account for 84% of the total tablet. The abundance estimations based on normalized PLS scores are 33% for acetaminophen, 38% aspirin and 12% caffeine, compared to the true value of 37%, 37% and

10%, respectively. If the complete set of ingredients was used for the PLS calculation, the normalization step would be unnecessary.

8.6.4 Conclusions

This example shows a variety of processing approaches that can be used to estimate the abundance of the three API components using a single tablet. The statistics available when collecting over 80 000 spectra per imaging sample enable not only a 'standard' analytical assay to be performed, but also provide real information on domain size, morphology, and distribution of the individual components comprising the finished product. By statistically characterizing domain sizes through morphological data processing, even more insight can be gained into what sample characteristics determine tablet physical properties, and ultimately product performance.

8.7 Processing Case Study: Determining Blend Homogeneity Through Statistical Analysis

As discussed in the previous case study, the physical properties of finished pharmaceutical forms may be affected by the interactions and domains sizes of the individual sample components. Because the blending process can determine the formation of these component domains and component interactions, how the blending proceeds may have a significant impact on the performance of the finished form. For example, improperly blended powders may result in a final product where the release of therapeutic materials occurs at less than optimal rates. Determination of the blend quality in both powder blend samples and finished forms has posed a particular problem to the analyst. Traditional spectroscopic and chromatographic methods can only yield the bulk content relationship of the components, and dissolution testing on the finished form can only indicate that a problem exists, but cannot provide direct information as to its source. NIR-CI is well suited to rapid and direct determination of the blend uniformity for all of the chemical species comprising the formulation, both for powder blends and finished tablets. This example will focus on examining blend homogeneity of finished forms.

8.7.1 Experimental

In this application a series of six tablets representing a wide range of blending homogeneity are analyzed. Visually, and in terms of total chemical composition, all of the tablets are identical, composed of exactly the same proportion of materials: 80 mg furosemide (API) and 240 mg excipient mix (99.7% Avicel PH 102 and 0.33% magnesium stearate). Five of the samples were tableted in the laboratory with homogeneity controlled by varying the blending time. These tablets range from nearly unmixed to fully blended. The sixth tablet is a high-quality 'commercially' blended example. All of the tablets were formed by direct compression method; the laboratory examples with a Carver press, and the commercial tablet on an automated tableting press.

Each tablet was imaged with a MatrixNIR chemical imaging system (Spectral Dimensions, Inc., Olney, MD, USA) using a pixel resolution of $40 \times 40\,\mu$m over the wavelength range of 1000–1700 nm at 10 nm spectral resolution. The time required to acquire each image is approximately 2 min. The raw images were background corrected using a data cube collected from Spectralon (Labsphere, Inc., North Sutton, NH, USA) as the reference, converted to $\log_{10}(1/R)$, and baseline corrected. Images were also collected from samples of pure furosemide and avicel. These data were processed in the same manner as the tablet images and used to create a pure component reference library for the system. A PLS2 model was constructed from the

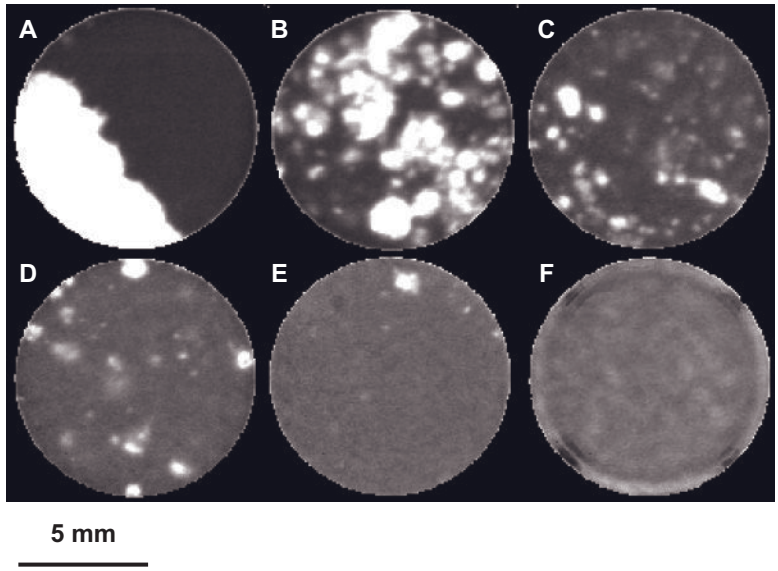


Figure 8.14 *(A–F) Chemical image analysis of six tablets composed of 80 mg of furosemide (API) mixed with 240 mg of an excipient mixture (99.7% Avicel PH 102 and 0.33% magnesium stearate.) Tablets A–E were created in the laboratory using increasing blending time, tablet F is an example of a commercial product. These PLS score images of the tablets highlight the location of the API component (bright pixels), and the gross blend-quality differences.*

reference library and used to generate PLS scores images from each of the tablet data cubes. The resulting scores images are shown in Figure 8.14 in which 'white' denotes the location of the API component.

8.7.2 Observing visual contrast in the image

The quality of the blending in each of these samples is readily apparent in the PLS scores images. In the poorly blended tablets (A and B) there is sharp contrast between the domains of nearly pure API against a black background of the nearly pure excipient materials. As the blend quality improves in tablets C and D, the contrast softens as the images start to become more 'gray', indicating more mixing of the white and dark (i.e. API and excipients). The API domains are still present, although they are smaller and less distinct. Tablet E is nearly uniform gray indicating good blending, although one small API-rich domain is still observed near the top of the image. The commercial tablet F is extremely well blended and shows almost no contrast at all when looking at the API distribution.

8.7.3 Statistical analysis of the image

While visual inspection of the chemical images gives a qualitative indication of the blend quality in the samples, more quantitative metrics are available through statistical analysis of the images using histogram representations (see Section 8.4.6). Figure 8.15 shows corresponding histograms for the six individual tablet score images. In these graphs the abscissa represents the range of intensities (in this case, the individual PLS score values) and the ordinate denotes the number of individual pixels within the image possessing that

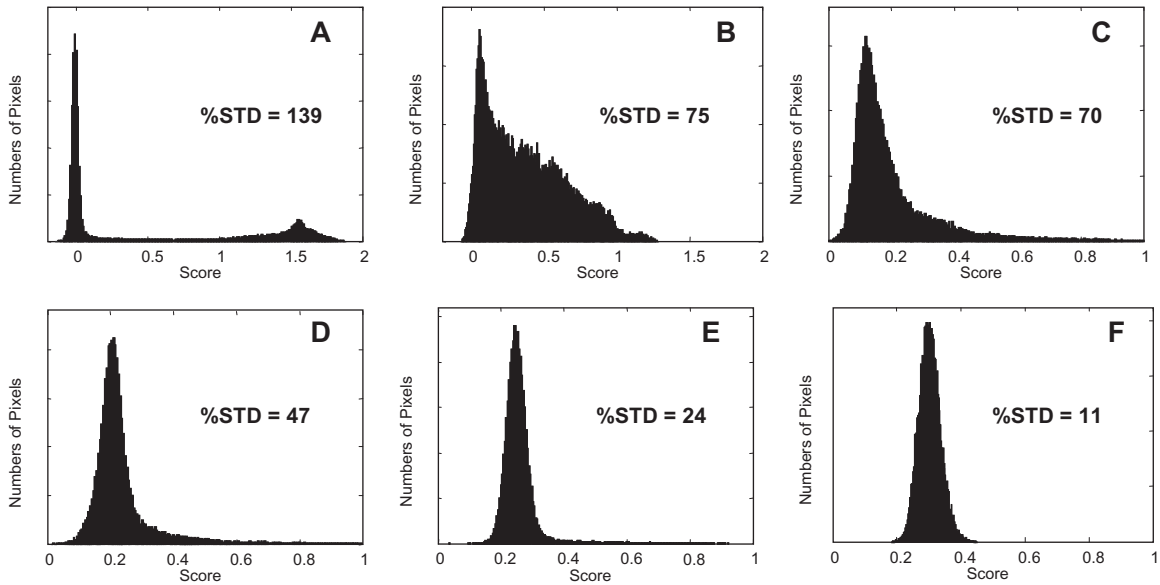


Figure 8.15 *(A–F): Histogram representations of the PLS score images showing the statistical distributions of the API class. Heterogeneity in the blend is indicated by deviations from a normal distribution and can be expressed as percent standard deviation (%SD), calculated by dividing the standard deviation by the mean.*

value. A highly homogeneous image will produce a narrow, Gaussian distribution such as that seen in the histogram of the commercial tablet (F). On the other hand, a starkly heterogeneous image will yield multimodal distributions. For example, the histogram for the unblended tablet (A) contains two distinct modes, the larger population at the left depicts the pure excipient area with near-zero PLS score values, and the smaller population on the right represents the pure API region with large score values. The histograms from the images of the intermediate tablets indicate heterogeneity as varying degrees of distortion from the normal distribution.

For a statistical analysis to adequately characterize the distribution of a component, the data should first be processed to obtain the optimum selectivity for that component. In this application the PLS model produces a score image that effectively separates the spectral response of the API from the excipients. Even though there is very little observable contrast in the images of the well blended samples, the results from the poorly blended samples convey confidence that the method is effective at tracking the API distribution.

It should be noted that the relationship between domain sizes of the components in the matrix and the magnification employed will influence the image statistics. For the well-blended samples, the domain sizes of the API and excipient materials are much smaller than the individual image resolution elements. Therefore, each pixel should correspond to a mixture spectrum denoting the relative amounts of the components at that location in the sample. The multivariate PLS modeling approach produces a measure of the relative API concentration at each individual pixel in the image. The resulting histogram representations of the PLS scores images represent the distribution of concentrations at the individual pixel locations. On the other hand, if the real domain sizes were larger than the pixel size, each pixel would record a relatively pure component spectrum, regardless of the actual blend quality. In this case, the distributions obtained from the

image may appear bi- or multimodal even for a highly homogeneous mixture (in the bulk sense). In these cases it may be advantageous to use other image analysis tools, such as particle statistics as shown in the previous example, to access the bulk uniformity of the blend. Alternatively, it is also possible to adjust the magnification of the NIR-CI instrument to increase the area interrogated by the individual pixels.

8.7.4 Blend uniformity measurement

For the blend uniformity application, the primary concern is overall heterogeneity within the sample. This is directly indicated by the width of the distribution and is quantitatively measured as the percent standard deviation relative to the mean (%SD). The %SD value obtained from the PLS scores distribution from each of the six imaged tablets is noted on the histograms in Figure 8.15(A–F). The value of %STD consistently decreases as the blending quality of the sample improves.

The range of blend quality represented by these six samples is much greater than would normally be encountered in an actual manufacturing process. However, the sensitivity of the %SD metric is demonstrated through the comparison of the results obtained from the commercial tablet (F) and the best-blended laboratory tablet (E). The general lack of contrast in the visual images suggests that both of these tablets are, on the whole, extremely well blended. The significantly larger %SD values obtained from tablet E likely stem from the isolated domain of increased concentration observed near the top edge of the sample in the PLS scores image. This hypothesis was tested by truncating the PLS scores image from these samples to a 100×100 pixel region from the center region of the tablets. The enlarged region is approximately 20% of the whole tablet area and is still large enough to be representative of the overall blend quality of the tablet. The truncated images and the corresponding histograms are shown in Figures 8.16a and b, respectively. In this analysis, the %SD value for the commercial tablet changes only very slightly while that of the laboratory tablet (E) decreases significantly to a value only slightly larger than that yielded by the commercial tablet. This indicates that, except for a single localized region, the general blend uniformity of both these tablets is quite similar.

This result also indicates the value of examining as much of the entire sample as possible when characterizing blend homogeneity. If only a small region is analyzed, it is possible that isolated 'hot-spots' would escape detection. NIR-CI has the advantage of the ready availability of large, high-density array detectors as well as flexible fields of view. This means that the configuration can be easily tailored such that the entire sample surface is analyzed rapidly and simultaneously. If one is concerned that the surface may not be representative of the entire thickness of the tablet, the sample can be turned over or easily scraped (e.g. with a razor-blade or microtome device) to reveal interior surfaces. Subsequent images of the revealed interior surfaces can then be collected and analyzed in the same manner to yield a more complete, volumetric analysis of the blend uniformity. In either case, the ability to rapidly generate statistically relevant blending data from single or multiple tablets is a critical aspect of NIR-CI's value to the PAT initiative.

8.7.5 Conclusions

The spatial aspect of NIR-CI provides the information necessary to develop a fast and accurate approach for the quantitative characterization of the homogeneity of a solid dosage form. This information cannot be directly accessed by any other traditional compendial or single-point spectroscopic technique. Access to such data is of great value in the development of the manufacturing process, monitoring finished products, and root cause analysis of manufacturing problems. For any given pharmaceutical product, the ultimate goal of the manufacturer may be to achieve a product that is as homogeneous as possible. In this case, the relative width (%SD) of the API distribution shown in the histogram of the chemical image provides an objective, accurate, and very sensitive metric for blend quality. Other components of a complex formulation could

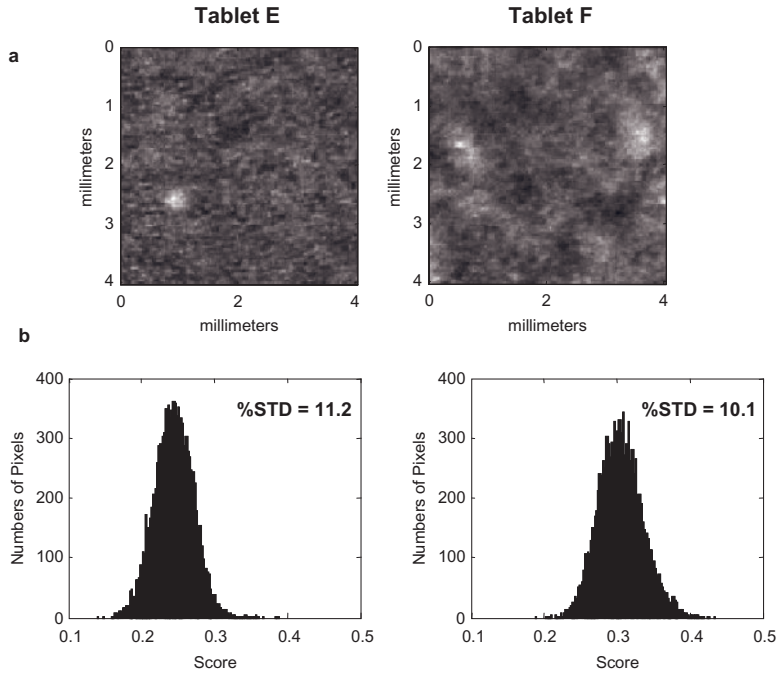


Figure 8.16 *Analysis of selected areas of the best laboratory blend (Tablet E) and the commercial blend (tablet F). (a) Truncated PLS score imaged for the API component (depicted as white) from the central region of the tablets showing that on a local scale, the blending in both samples is uniform; and (b) histograms and resulting %SD of the distribution from the truncated images. Note that both examples display nearly normal distributions.*

also be measured in the same manner. The data collection and processing can also be easily built into a 'turnkey' method so that the entire procedure can be automated.

8.8 Final Thoughts

NIR-CI provides analytical capabilities that are not achievable with standard single-point spectroscopic methods, or even other chemical imaging techniques. Deriving spatially resolved spectral information, such as domain size and component interactions, throughout the various pharmaceutical manufacturing steps is anticipated to provide much of the core understanding of critical process parameters that ultimately drive product performance. The ability to provide a simplified, turnkey data collection and processing solution also enables this application to move from an R&D to a manufacturing environment. In addition, the flexible FOV and built in redundancy of detector elements in the focal-plane array, make it possible to utilize 'built-in' calibration sets, potentially obviating the need for cumbersome transfer of calibration methods. In the particular hardware configuration emphasized in this chapter, a system with no-moving parts, flexible FOV and modular approach to data collection and processing can be integrated to provide high throughput capabilities for quality assurance processes. What and how much one may be able to learn about the complexities of pharmaceutical manufacturing from NIR-CI is still relatively unknown, but early results suggest that it is potentially a great deal.

Acknowledgements

The authors would like to thank Dr. Robbe C. Lyon of the FDA for providing the furosemide tablet samples.

References

1. FDA, Guidance for industry PAT – A framework for innovative pharmaceutical development, manufacturing and quality assurance, available http://www.fda.gov/cder/guidance/6419fnl.pdf
2. ICH, ICH harmonized tripartite guideline Q8(R1), Pharmaceutical development, available http://www.ich.org/LOB/media/MEDIA4986.pdf accessed 17 November 2009.
3. ICH, ICH harmonized tripartite guideline Q8(R1), Pharmaceutical development annex, available http://www.ich.org/LOB/media/MEDIA4986.pdf accessed 17 November 2009.
4. W.F. McClure, Near-infrared spectroscopy: the giant is running strong, *Anal. Chem.*, **66**, 43A-53A (1994).
5. K.A. Bakeev, Near infrared spectroscopy as a process analytical tool. part I: laboratory applications, *Spectroscopy*, **18**(11), 32–35 (2003).
6. E.W. Ciurczak and J.K. Drennen III (Eds), Pharmaceutical and medical applications of near-infrared spectroscopy, *Practical Spectroscopy Series Volume 31*, Marcel Dekker, New York, 2002.
7. R. Gimet and A.T. Luong, Quantitative determination of polymorphic forms in a formulation matrix using the near infrared reflectance analysis technique, *J. Pharm. Biomed. Anal.*, **5**, 205–211 (1987).
8. T. Norris, P.K. Aldridge and S.S. Sekulic, Determination of end-points for polymorph conversions of crystalline organic compounds using on-line near-infrared spectroscopy, *Analyst*, **122**, 549–552 (1997).
9. S.S. Sekulic, H.W. Ward, D.R. Brannegan, *et al.*, On-line monitoring of powder blend homogeneity by near-infrared spectroscopy, *Anal. Chem.*, **68**, 509 – 513 (1996).
10. P.A. Hailey, P. Doherty, P. Tapsell, P.T. Oliver and P.K. Aldridge, Automated system for the on-line monitoring of powder blending processes using near-infrared spectroscopy. part I. system development and control, *J. Pharm. Biomed. Anal.*, **14**, 551–559 (1996).
11. M. Warman, Using near infrared spectroscopy to unlock the pharmaceutical blending process, *Am Pharm. Rev.*, **7**(2), 54–57 (2004).
12. A.D. Trafford, R.D. Jee, A.C. Moffat and P. Graham, A rapid quantitative assay of intact paracetamol tablets by reflectance near-infrared spectroscopy, *Analyst*, **124**, 163–168 (1999).
13. C.T. Rhodes and K. Morissau, Tablet evaluation using near infrared spectroscopy, in *Encyclopedia Pharmaceutical Technology*, 2nd Edition, J. Swarbrick and J.C. Boylan (Eds), Marcel Dekker, New York, 2002.
14. F.C. Clarke, M.J. Jamieson, D.A. Clark, S.V. Hammond, R.D. Jee and A.C. Moffat, Chemical image fusion: the synergy of FT-NIR and Raman mapping microscopy to enable a more complete visualization of pharmaceutical formulations, *Anal. Chem.*, **73**(10), 2213–2220 (2001).
15. E.N. Lewis, J.E. Carroll and F.C. Clarke, A near infrared view of pharmaceutical formulation analysis, *NIR News*, **12**(3), 16–18 (2001).
16. R.C. Lyon, E.H. Jefferson, C.D. Ellison, L.F. Buhse, J.A. Spencer, M.M. Nasr and A.S. Hussain, Exploring pharmaceutical applications of near-infrared technology, *Am. Pharm. Rev.*, **6**(3), 62–70 (2003).
17. R.C. Lyon, D.S. Lester, E.N. Lewis, E. Lee, L.X. Yu, E.H. Jefferson and A. S., Hussain, Near-infrared spectral imaging for quality assurance of pharmaceutical products: analysis of tablets to assess powder blend homogeneity, *AAPS Pharm. Sci. Tech.*, **3**(3), 1–15 (2002).
18. F. Clarke, Extracting process-related information from pharmaceutical dosage forms using near infrared microscopy, *Vib. Spectr.*, **34**, 25–35 (2004).
19. F. Clarke and S. Hammond, NIR microscopy of pharmaceutical dosage forms, *Eur. Pharm. Rev.*, **8**(1), 41–50 (2003).
20. R. Barer, A.R.H. Cole and H.W. Thompson, Infra-red spectroscopy with the reflecting microscope in physics, chemistry and biology, *Nature*, **163**, 198–200 (1949).
21. M.A. Harthcock and S.C. Atkin, Imaging with functional group maps using infrared microspectroscopy, *Appl. Spectrosc.*, **42**(3), 449–455 (1988).
22. P.J. Treado, I.W. Levin and E.N. Lewis, Indium antimonide (InSb) focal plane array (FPA) detection for near-infrared imaging microscopy, *Appl. Spectrosc.*, **48**(5), 607–615, (1994).
23. D.S. Tezcan, S. Eminoglu and T. Akin, A low-cost uncooled infrared microbolometer detector in standard CMOS technology, *IEEE Trans. Electron. Dev.*, **50**(2), 494–502 (2003).

24. P.J. Treado, I.W. Levin and E.N. Lewis, Near-infrared acousto-optic spectroscopic microscopy: a solid-state approach to chemical imaging, *Appl. Spectrosc.*, **46**(4), 553–559 (1992).
25. E.N. Lewis and I.W. Levin, I.W., Real-time, mid-infrared spectroscopic imaging microscopy using indium antimonide focal-plane array detection, *Appl. Spectrosc.*, **49**(5), 672–678 (1995).
26. D. Wienke, W. van den Broek and L. Buydens, Identification of plastics among nonplastics in mixed waste by remote sensing near-infrared imaging spectroscopy. 2. Multivariate rank analysis for rapid classification, *Anal. Chem.*, **67**, 3760–3766 (1995).
27. M.K. Bellamy, A.N. Mortensen, R.M. Hammaker and W.G. Fateley, Chemical mapping in the mid- and near-IR spectral regions by Hadamard transform/FT-IR spectrometry, *Appl. Spectroc.*, **51**(4), 477–486 (1997).
28. E.N. Lewis, P.J. Treado, R.C. Reeder, G.M. Story, A.E. Dowrey, C. Marcott and I.W. Levin, Fourier transform step-scan imaging interferometry: high-definition chemical imaging in the infrared spectral region, *Anal. Chem.*, **67**, 3377–3381 (1995).
29. J.A. Reffner and P.A. Martoglio, Uniting microscopy and spectroscopy, *in Practical Guide to Infrared Microscopy*, H.J. Humecki (Ed), Practical Spectroscopy Series Volume 19, Marcel Dekker, New York, 1995.
30. M. Claybourn, External reflection spectroscopy, in *Handbook of Vibrational Spectroscopy*, J.M. Chalmers and P.R. Griffiths (Eds), Sampling Techniques Vol. **2**, John Wiley & Sons, Ltd, Chichester, 2002.
31. P.R. Griffiths and J.M. Olinger, Continuum theories in diffuse reflection, in *Handbook of Vibrational Spectroscopy*, J.M. Chalmers and P.R. Griffiths (Eds), Sampling Techniques Vol. **2**, John Wiley & Sons, Ltd, Chichester, 2002.
32. J. van der Weerd and S.G. Kazarian, Validation of macroscopic attenuated total reflection-Fourier transform infrared imaging to study dissolution of swelling pharmaceutical tablets, *Appl. Spectrosc.*, **58**, 1413–1419 (2004).
33. K.L.A. Chan, S.V. Hammond and S.G. Kazarian, Applications of attenuated total reflection infrared spectroscopic imaging to pharmaceutical formulations, *Anal. Chem.*, **75**, 2140–2146 (2003).
34. A.J. Eilert and D.J. Wetzel, Optics and sample handling for near-infrared diffuse reflection, in *Handbook of Vibrational Spectroscopy*, J.M. Chalmers and P.R. Griffiths (Eds), Sampling Techniques Vol. **2**, John Wiley & Sons, Ltd, Chichester, 2002.
35. P. Kubelka and F. Munk, Ein beitrag zur optik der farbastriche, *Zeits. f. tech. Physik*, **12**, 593–601 (1931).
36. A. Savitzky and M.J.E. Golay, Smoothing and differentiation of data by simplified least squares procedures, *Anal. Chem.*, **36**, 1627–1639 (1964).
37. P. Geladi, D. MacDougall and H. Martens, Linearization and scatter-correction for NIR reflectance spectra of meat, *Appl. Spectrosc.*, **39**, 491–500 (1985).
38. D.L. Massart, B.G.M. Vandeginste, L.M.C. Buydens, S. DeJong, P.J. Lewi and J. Smeyers-Verbeke, *Handbook of Chemometrics and Qualimetrics: Part A; Data Handling in Science and Technology*, Elsevier, Amsterdam, 1997.
39. D.L. Massart, B.G.M. Vandeginste, L.M.C. Buydens, S. DeJong, P.J. Lewi and J. Smeyers-Verbeke, *Handbook of Chemometrics and Qualimetrics: Part B; Data Handling in Science and Technology*, Elsevier, Amsterdam, 1998.
40. H. Martens and T. Naes, *Multivariate Calibration*, John Wiley & Sons, Ltd, Chichester, 1989.
41. H. Mark, *Principles and Practice of Spectroscopic Calibration*, John Wiley & Sons, Ltd, New York, 1991.
42. H. Mark and J. Workman Jr., *Statistics in Spectroscopy*, Academic Press, New York, 1991.
43. D.A. Burns and E.W. Ciurczak (Eds), *Handbook of Near-Infrared Analysis*, Practical Spectroscopy Series Volume 13, Marcel Dekker, New York, 1992.
44. P. Geladi and H. Grahn, *Multivariate Image Analysis*, John Wiley & Sons, Chichester, 1996.
45. S. Carda-Broch, J. Esteve-Romero and M.C. Garca-Alvarez-Coque, Furosemide assay in pharmaceuticals by micellar liquid chromatography: study of the stability of the drug, *J. Pharm. Biomed. Anal.*, **23**(5), 803–817 (2000).

9

Acoustic Chemometric Monitoring of Industrial Production Processes

Maths Halstensen[1] and Kim H. Esbensen[2]

[1] *ACRG (Applied Chemometrics Research Group), Telemark University College, Porsgrunn, Norway*
[2] *ACABS research group, Aalborg University, campus Esbjerg (AAUE), Esbjerg, Denmark*

9.1 What is Acoustic Chemometrics?

Acoustic chemometrics is an emerging method for on-line process monitoring based on characterization of system vibrations as generated by an industrial process such as a manufacturing process or transportation flow. Acoustic chemometrics is applicable for quantitative analysis of constituents for process monitoring and for physical characterization of the state of process equipment. Chemometric techniques such as principal component analysis (PCA) or partial least squares regression (PLS) [1] based on empirical input-response data are used to extract relevant information and for calibration of relevant acoustic signals. The PLS model can then be used to predict parameters of interest based on new independent acoustic spectra. Proper validation [2] of resulting regression models is critical in order to calibrate realistic prediction models which have been shown to be robust tools for process monitoring.

The main advantage of this method compared with many other on-line methods for process monitoring is the use of noninvasive, so-called 'clamp-on' sensors which can be easily mounted onto the process equipment (pipelines, reactors). The sensor, which often is a standard accelerometer, has no moving parts, and can, for example, withstand high temperatures, a dirty environment and is easy to maintain. The acoustic spectra measured with one sensor will often contain information about several process-relevant properties and/or analytes which makes it possible to predict several parameters of interest from the same acoustic spectrum.

Vibrational acoustic emission from industrial processes is often considered as audible random signals only, but it has recently been proven that within this 'noise' there is often a bonanza of hidden useful information [3–11], highly relevant for processes monitoring purposes. The fact that almost all processes produce some kind of acoustic emission opens up the potential for diverse applications which depend totally on sound, sensor-technological knowledge and chemometric multivariate calibration competencies.

Process Analytical Technology 2e Edited by Katherine Bakeev
© 2010 John Wiley & Sons, Ltd.

Several experiments covering multiphase fluids, system state of rotating machinery and powder characterization [3–11] have been conducted showing that this method is a promising *on-line* process analytical technology (PAT) approach. The objective of this chapter is to give an overview of the industrial on-line process monitoring potential, and to give several examples showing how to develop and implement acoustic chemometrics, for example in the context of industrial granulation and transportation flow.

Acoustic chemometrics has its greatest benefits in cases where traditional sensors and measurement techniques, such as flow, temperature and pressure transmitters cannot be used. In many processes it is preferable to use noninvasive sensors because invasive sensors may cause disturbances, for example fouling and clogging inside the process equipment such as pipelines, reactors cyclones, etc. In this chapter we concentrate mainly on new industrial applications for acoustic chemometrics, and only discuss the necessary elements of the more technical aspects of the enabling technology below – details can be found in the extensive background literature [3–5].

9.2 How Acoustic Chemometrics Works

Figure 9.1 shows schematically a general acoustic chemometric data path, from acoustic emission to the final multivariate calibration model, including future prediction capabilities.

9.2.1 Acoustic sensors

The acoustic sensor used for capturing process vibrations is often a standard *accelerometer*, covering a frequency range 0–50 kHz – or a so-called *acoustic emission sensor* (AE) often covering higher frequencies from 50 kHz up to several MHz. The 4396 and ATEX 5874 accelerometers from Brüel & Kjær are shown in Figure 9.2. The 4396 accelerometer is a standard accelerometer for industrial use and the ATEX 5874 is a certified accelerometer for use where explosion proof equipment is needed (EX certified).

Fluidized bed granulators, or dryers, often require high temperature sensors which are available from many different OEMs and suppliers; standard sensors operate only in the 0–70 °C interval. All accelerometers and AE sensors have small dimensions, which makes mounting easy. Figure 9.3 shows how to fix the sensor and cable properly in order to avoid nonrelevant vibrations from occurring and interfering with the measurements.

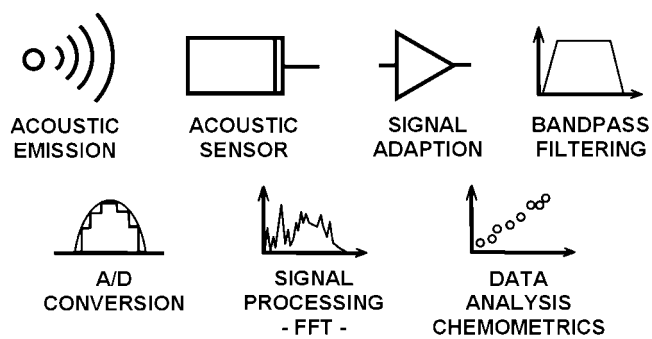


Figure 9.1 *Schematic data flow for acoustic chemometrics. Each step in the flow sheet constitutes a basic unit operation described in the text. Reprinted from [7]. Copyright 2006, with permission from Elsevier.*

Figure 9.2 *Left: 4396 standard accelerometer. Centre: ATEX 5874 Accelerometer which is certified for use where explosion proof sensors are required. Right: Acoustic Emission (AE) sensor from Kistler. Reprinted from [7]. Copyright 2006, with permission from Elsevier.*

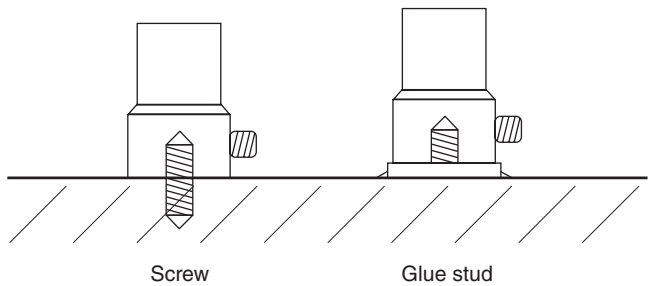


Figure 9.3 *Left: A screw is used to fix the sensor, Right: A glue stud is screwed into the sensor and glued to the measuring point.*

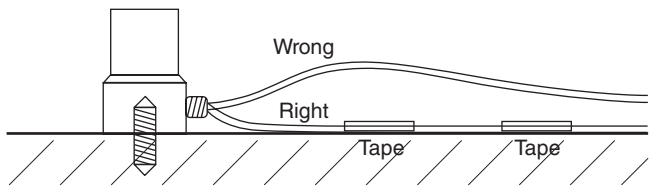


Figure 9.4 *The cable at the top is not mounted properly, while the cable at the bottom is correctly fixed to the same surface as the sensor, using tape, plastic strips or other fixing agents.*

9.2.2 Mounting acoustic sensors (accelerometers)

There are several different ways to mount the sensor: glue, screw, magnet, tape etc. The most used methods in the industry are to use a screw implement or to use a glue stud as indicated in Figure 9.3.

The preferred mounting method for long-term stability is the screw tap. The glue stud is often used for experimental use and when the sensor cannot be permanently fixed in one location. The cable from the sensor will cause problems if not fixed properly to the same surface as the sensor is mounted. Vibrations may cause the cable to add noise to the signal when constantly impacting the vibrating process equipment. The cable has to be fixed to the surface where the sensor is mounted to ensure a high quality signal as shown in Figure 9.4.

9.2.3 Signal processing

Time-domain signals acquired from the sensor are fed to a signal adaptor (Figure 9.1) in order to amplify and fit the signal to the desired input range of the analog to digital converter. Correct signal range from the amplification unit will ensure the highest possible digital resolution. The signal then passes through a band-pass filter in order to avoid unwanted frequencies, which would otherwise adversely influence the acoustic signal (e.g. stripping off mainline 50–60 Hz impacts). The filter is also necessary to avoid *aliasing* (false frequencies originating from inferior signal sampling). After this, the signal is converted to digital format and a fixed number of samples (time domain) are transformed by a window transformation function (for example a Hamming window) in order to reduce spectral leakage. The filtered signal is finally transformed from the time domain to the frequency domain by fast Fourier transformation (FFT) [12]. The role of the FFT is to take the acoustic signal into the frequency domain – the end result of which is termed an acoustic spectrum – or the acoustic signature. FFT spectra can now be analyzed by suitable data analytical methods (see below).

9.2.4 Chemometric data analysis

FFT spectra contain an abundance of potential information (as a function of frequency) related to the process/products characterized by the acoustic sensors. FFT spectra constitute fit for purpose input to any chemometric data analytical method deemed beneficial for making the final model relating the acoustic emissions (vibrations), X, to the salient functional properties (quality, quantity) of the products or process states, Y.

From the outset acoustic chemometrics is fully dependent upon the powerful ability of chemometric full spectrum data analysis to elucidate exactly where in the spectral range (which frequencies) the most influential information is found. The complete suite of chemometric approaches, for example PCA, PLS regression, SIMCA (classification/discrimination) are at the disposition of the acoustic spectral data analyst; there is no need here to delve further into this extremely well documented field. (See Chapter 12 for more detail.)

9.2.5 Acoustic chemometrics as a PAT tool

In this chapter we illustrate development of acoustic chemometrics for process monitoring in the chemical engineering realm with a feasibility survey of the application potential for industrial applications. Earlier studies [3–7] describe the initial small- to medium-scale pilot experiments behind this approach, collectively giving a broad introduction to the application of acoustic chemometric for liquid flow in pipelines [3,7], industrial granulation processes [7–9,11] and similar process technological regimens. Since a full introduction to acoustic chemometrics has been published elsewhere [3–7], only an operational summary is provided in this chapter.

As acoustic chemometrics is based on capturing *passive* system vibration characteristics, such as from two-phase systems (gas–solid/liquid–solid) the resulting power spectrum can be viewed as a process signature which carries *embedded* information about a whole range of system-relevant physical and chemical parameters: such as composition (earlier studies looked at oil, fat, ammonia, glycol, ethanol), mixing fractions, mixing progress, fiber length, flow, density, temperature – as well as system state.

Many earlier successful PLS prediction models (which in this chapter are presented as examples from industrial production processes) signify that acoustic chemometrics has matured into a proven on-line technology in the PAT domain. It is a salient point here that all were evaluated using test set validation [2].

9.3 Industrial Production Process Monitoring

The application potential of acoustic chemometrics is illustrated below by examples from the following types of industrial processes:

1. Fluidized bed granulation monitoring of a fertilizer product (urea) in a semi-industrial pilot plant (SIPP), illustrating the main features and benefits.
2. Process state monitoring for visualization of critical situations – early operator warnings.
3. Monitoring of ammonia concentration.

Acoustic chemometrics can be used for monitoring of both process state and product quality/quantity for better process control. Monitoring process states can provide early warnings which trigger the process operator to change relevant process parameters to prevent critical shutdown situations.

9.3.1 Fluidized bed granulation monitoring

The acoustic chemometric approach can also be used to monitor industrial production processes involving particles and powders and to provide a complementary tool for process operators for more efficient process control, or to monitor particle movement in a fluidized bed [7] for example. Below we illustrate the application potential by focusing on two applications: process monitoring of a granulation process and monitoring of ammonia concentration.

Granulation of urea [13] is a complex process that has to be controlled by experienced process operators in order to avoid critical shutdown situations. The parameters most often used for monitoring granulation processes are measured by classical univariate sensors, such as temperature, pressure and flow. However, these standard process measurements carry only little or no relevant information, or are only indirectly related to, for example particle size, clogging of the reactor, or the accumulation of a solids layer on the bottom plate. The response from these sensors often comes with quite a substantial delay time.

A sample of layering cake on the bottom perforated plate, extracted from a fluid bed reactor after several days in production as shown in Figure 9.5, is already a serious process impediment. When the layering cake

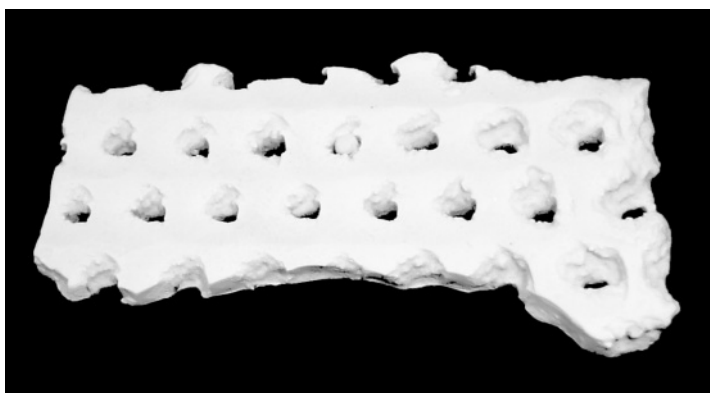


Figure 9.5 *Layering on perforated bottom plate in semi-industrial scale fluid bed granulator after several days of production. The thickness of the layer is approximately 5 mm.*

develops further, the perforated bottom plate of the reactor becomes increasingly clogged with a resultant fluidization airflow decrease. Decreased fluidization in turn leads to a situation with less agitation of the particles; the result is often generation of big lumps, which can quickly lead to a shutdown of the reactor, and a significant economic loss (reactor downtime and general production flow stoppage during cleanup).

A measurement system that can predict the thickness of the layering cake, average particle size, or give an early warning of lump formation is desirable. Similarly, an acoustic chemometrics prediction facility for general process state monitoring is of equally critical importance.

One of the major goals of the granulation feasibility study which was originally presented in [7] was to relate process state trends, presented as chemometric score plots, to specific conditions/qualities of the product inside the reactor. The process operators can then use this 'new' type of information to better operate the process, with an ultimate objective to significantly reduce costly shutdown situations. On-line measurements of process characteristics such as particle size and/or distribution together with properties of the liquid feed to the spray nozzles make it manifestly easier to control the process. The results presented below represent an experimental trial of 2 months, involving a suite of induced process pertubations and deviations of the general production process in order to learn as much as possible about the feasibility of acoustic chemometrics in monitoring granulation.

9.3.2 Pilot scale studies

The objectives of this particular experiment are divided into three major themes:

1. Optimal sensor positioning on a semi-industrial granulator.
2. Feasibility of acoustic chemometrics to predict:
   a. Crystallization point temperature of liquid urea feed to a granulator
   b. Moisture content in the granules
   c. Concentration of ammonia
   d. Average particle size.
3. Monitor the overall granulator process state, to detect critical situations and to visualize these situations as early warnings in an operator-friendly fashion (lump formation and clogging of the bottom plate are the most important mishaps in the industrial production setting).

9.3.2.1 Fluidized Bed Granulation Studies

The experiments were carried out on a semi-industrial fluidized bed reactor, illustrated in Figure 9.6, which shows four different sensor positions (A, B, C and D). Screw fittings were used to mount the sensors in order to secure optimal sensor pickup efficiency. Sensor location A is mounted onto an orifice plate on the main supply line of liquid urea to the reactor nozzles, following Esbensen *et al*. [5]. The sensors B, C and D are located on the wall of reactor chambers 1, 2 and 4, respectively.

The semi-industrial granulator is identical to an industrial full-sized granulator except for a volume scale-down, which results in a batch product reduction of roughly 1 : 10; all new formulation and process development is carried out using this pilot plant. The granulator is divided into five chambers: three injection (spraying) chambers and two cooling chambers. The injection chambers each have several nozzles where liquid urea is sprayed into the granulator at a certain process temperature. The bottom of the reactor is a perforated plate, which allows fluidization air to jet into the reactor to interact with the growing particles and keep all particles in the bed in vigorous agitation. The cooling chambers are used to cool down the granules before they exit as the final product: urea granules with a specified size and size range (important parameters for agroindustrial product use).

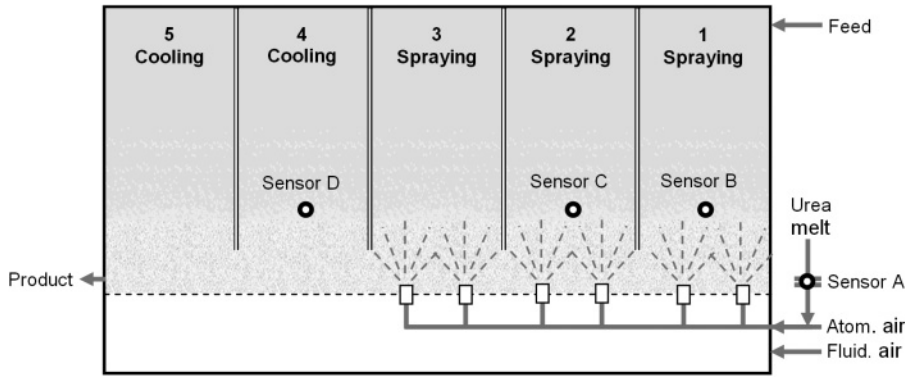


Figure 9.6 *Semi-industrial granulator with three spraying chambers and two cooling chambers. Sensor locations A, B, C and D are indicated.*

Sensor A is mounted onto an orifice plate inserted in the main supply pipeline for liquid urea. The orifice has a smaller hole diameter than the pipeline, which induces turbulence in the flowing urea downstream of the orifice. The vibrations produced by this turbulence will be detected by sensor A. Sensors B, C and D are mounted on the vertical wall on the granulator, about 30 cm above the perforated bottom plate to detect vibrations produced by the granules when they impact the reactor wall. Thus sensors B, C and D are used to monitor the process conditions inside the granulator, while sensor A is used to monitor the liquid supply of urea. The sensors used in this trial are all high temperature accelerometers.

The acoustic spectra were recorded simultaneously as other process experiments, in themselves not related to acoustic chemometrics, were carried out. This resulted in many days with stable conditions in the reactor, and no particular variations in the acoustic signals. Therefore there were only a limited number of days (hours) which displayed significant variation in process parameters, which are necessary for successful multivariate analysis and calibration.

9.3.2.2 Semi-industrial Reactor Experiments and Results

A model for crystallization point of the urea melt sprayed into the granulator was developed based on acoustic spectra recorded from sensor position A during a trial period of 24 hours. A flow sheet of the liquid urea feed process can be seen in Figure 9.7. Sensor A is mounted onto an orifice plate inserted in the main supply pipeline of liquid urea (see Figures 9.6 and 9.7). The reference values used to calibrate the model are the crystallization temperature (called the '*x* point'), as determined by the pilot plant laboratory (heat table visual nucleation/crystallization detection).

Physically the '*x* point' is correlated with flow characteristics of liquid urea (viscosity) which influence the acoustic spectral response and explain this particular indirect multivariate calibration.

A PLS regression model based on *X* (acoustic spectra from sensor A) and *Y* (crystallization temperature) was established. The **X** matrix contains 13 objects, each with 1024 variables (frequencies 0–25 kHz). An overview of the *X* data is shown in Figure 9.8, in which one can observe systematic changes in the acoustic signatures following the object (samples) succession.

The objects span a nominal urea concentration range of 85.5 to 91.4%. The PLS model will be established on experimental *Y* reference values (crystallization temperature), corresponding to 92–107 °C. A model for urea concentrations can also be established following appropriate laboratory data [14,15]. The crystallization temperature model (no outliers) is able to describe 87% of the *Y* variance with three PLS components.

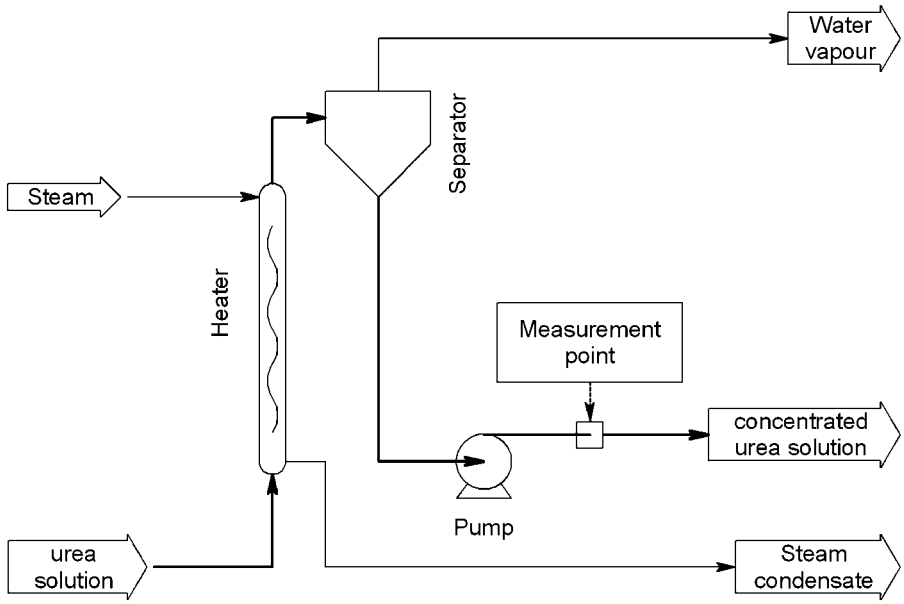


Figure 9.7 *Flow sheet of the liquid feed to the urea granulation process.*

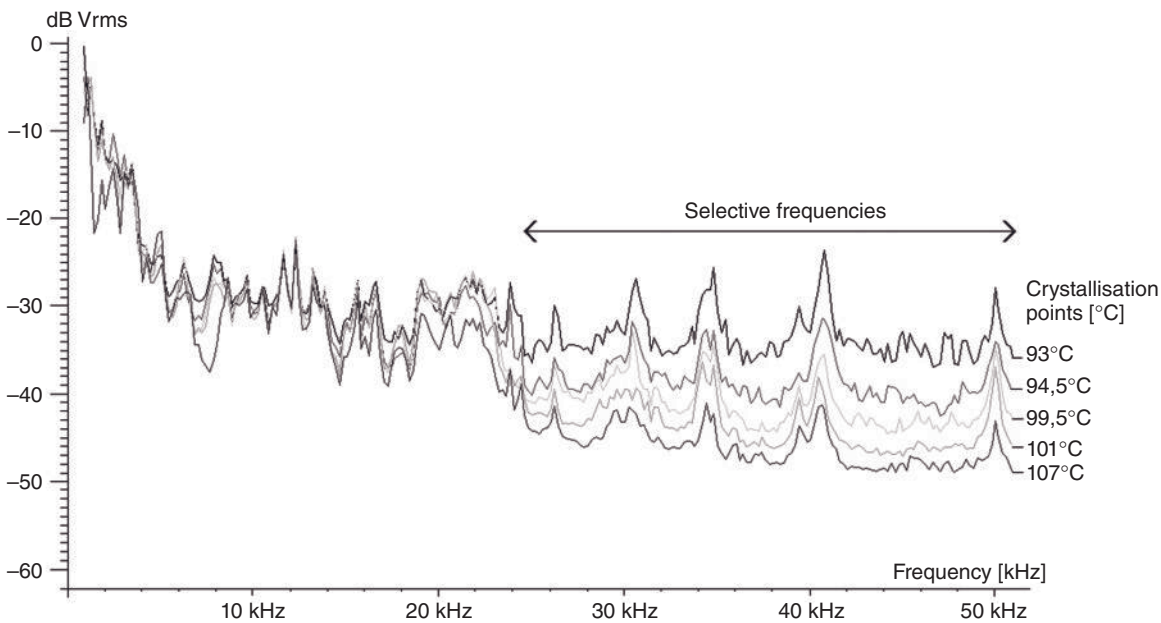


Figure 9.8 *X data matrix of acoustic spectra used in calibration of the urea crystallization temperature model. Reprinted from [7]. Copyright 2006, with permission from Elsevier.*

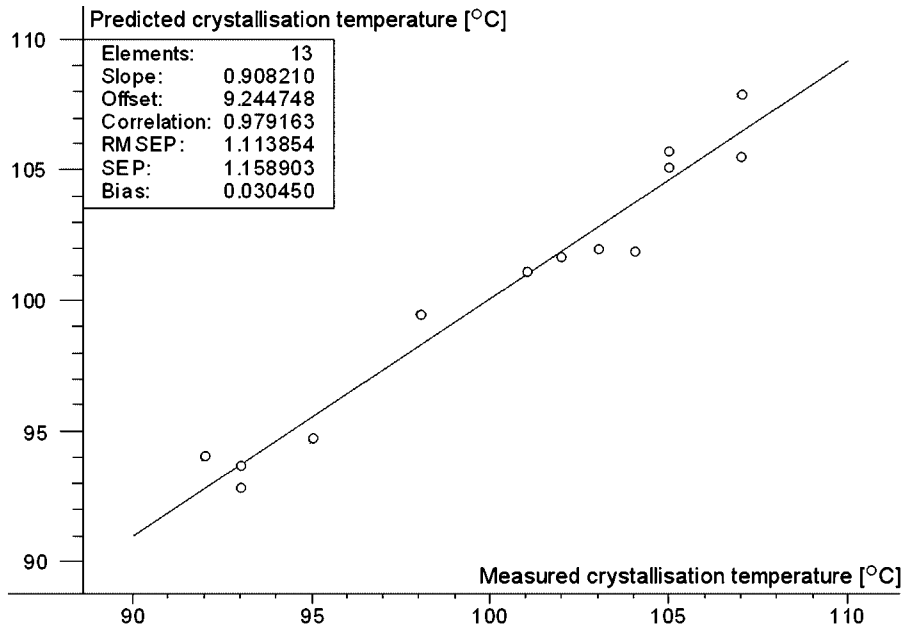


Figure 9.9 *Predicted versus measured plot for urea crystallization temperature.*

The predicted versus measured plot in Figure 9.9 shows the correlation between the acoustic signal and crystallization temperature, and hence its ability to predict it on-line.

The experimental crystallization temperature prediction model was evaluated using 13-segment cross-validation (full cross-validation); this can be considered acceptable for such 'small sample' data sets, albeit only furnishing indicative estimates [2].

The pilot study showed good prospects for predicting crystallization point temperatures directly from the acoustic signatures of the liquid feed into the granulator with an indicated prediction error (root mean square error of prediction) RMSEP = 1.1 °C.

Acoustic data were also calibrated against the average particle size as well as the particle size distribution, which were calculated from laboratory sieving measurements. Appropriate care is needed when trying to acquire *representative samples* in an industrial setting; full use of the Theory of Sampling is critical [14,15] (see also Chapter 3). The results for predicting average particle size were equally satisfactory (not shown here). The particle size distribution constitutes the prime industrial parameter of interest. Further development of the acoustic approach to monitor this parameter has the highest industrial priority.

9.3.2.3 Prediction of Average Particle Size

Particle sizes of the granules are often measured by manually taking out test samples at the granulator's product exit. These samples are sent to a central laboratory for sieving and measurement of particle size distribution and spread. One of the objectives in the feasibility project was to assess the potential of acoustic chemometrics for predicting the average particle size directly on-line at different locations in the granulator. This would provide the operators information about the particle sizes more frequently, and in several chambers simultaneously. The possibilities to correct for off-specification particle sizes are therefore significantly improved because the operator will get this critical information much earlier. The measurements recorded

to make the particle size model include both sensors C and D. (Sensor B was not used here because the particle sizes in chamber 1 are disturbed by the reflux of fine material from the crusher.) The reference used to calibrate the model is the average particle size (D50) calculated from sieve analysis in the laboratory.

The results displayed in Figure 9.10 show that it is possible to calibrate a model for the average particle size (slope 0.96, relative RMSEP = 3%), which would serve as a good indication to the process control operators to optimize the process to produce granules within product specifications. Test samples still have to be sent to the laboratory with some regularity to produce more accurate measures, but much less frequently

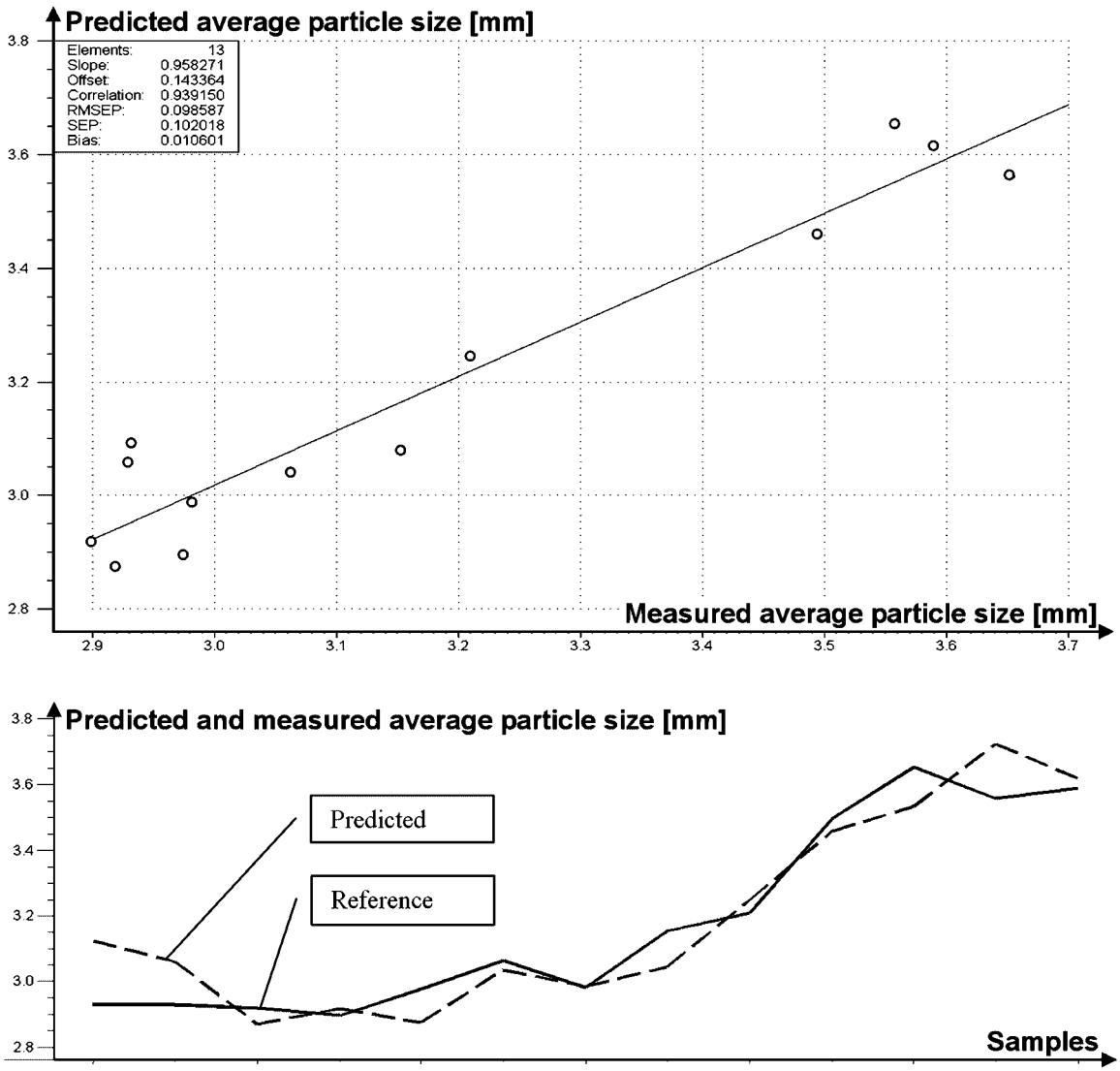


Figure 9.10 *PLS-1 model for the average particle size in chambers 2 and 3. Sensors C and D were used in this model based on four PLS components. The model was validated with segmented cross validation with 10 segments. Predicted versus measured (top) and predicted and measured (bottom).*

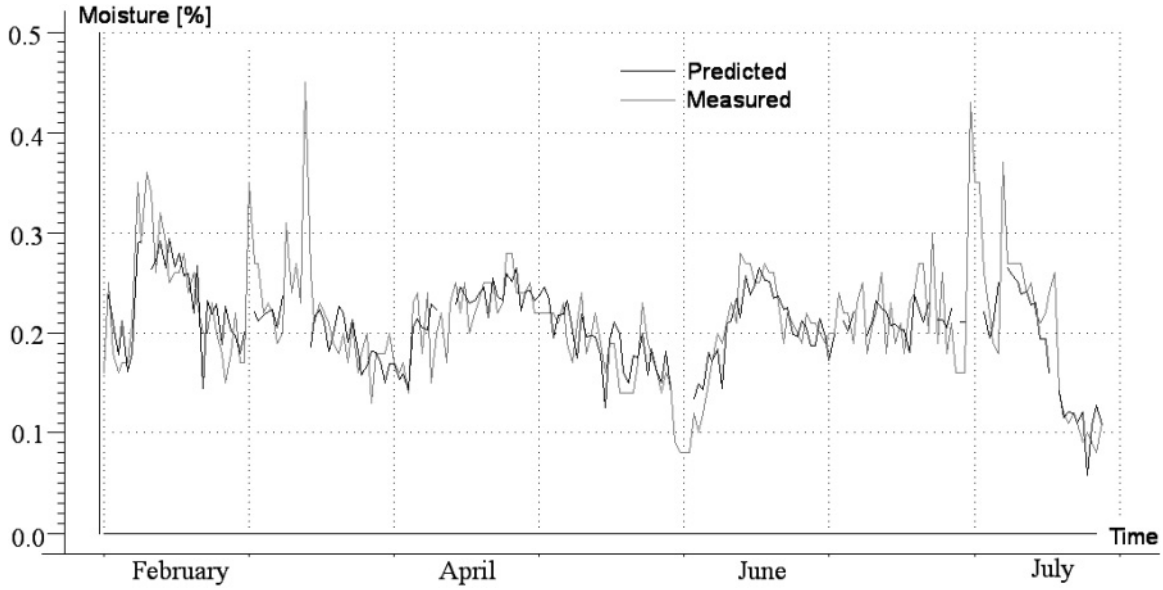


Figure 9.11 *Predicted and measured plot for granule moisture contents, calibrated on data from 5 months of production. Sensor B was used in this model (eight PLS components). Gray curve: measured, black curve: predicted.*

as when physical samples constituted the only basis for process monitoring. PAT predictive results are practically instantaneous and therefore real-time. Physical samples will only be needed for QC of the PAT system. It is highly likely that there are severe sampling uncertainties included in Y reference values as used today [14,15]. It is known that the current sampling procedures are not standardized, and probably not in statistical control (operator information). This is only of minor importance however, as all information, experience and equipment to achieve representative sampling in this and related industrial settings are now at hand by following the Theory of Sampling, see [14,15] and Chapter 3 with further references.

9.3.2.4 Monitoring of Urea Granule Moisture Content

For assessment of the potential to predict granule moisture content, a large 1032-object data set recorded during 5 months of urea production was used. The first 900 objects were used for calibration and the last 132 as a validation test set [2]. The data matrix was resampled to allow acoustic data to be calibrated against laboratory tests of moisture content, which were only available with a relatively low sampling rate; however, plenty of results were at hand to allow a full assessment of the multivariate prediction performance for granule moisture. The validated prediction results can be seen in Figure 9.11.

In spite of the fact that this model has a relatively large RMSEP of 14%, the general potential to pick up the important production trend *changes* is at an acceptable level for improved process monitoring.

9.3.3 Monitoring of a start-up sequence of a continuous fluidized bed granulator

A comparison of PCA score plots on both acoustic and conventional process data is presented to see if acoustic chemometrics can supply new information to the process operators in the critical start-up phase of

the reactor. The data set used is from 20 February 2001; a score plot for the acoustic measurements is shown in Figure 9.12.

The score plot in Figure 9.12 shows the whole start-up sequence: empty bed, bed filling, injection of liquid urea, and production. The score plot for the process data can be seen in Figure 9.13.

The two score plots in Figures 9.12 and 9.13 show similar trends for process data and acoustic data. A plot of score T1 for the acoustic model and T1 for the process data can be seen in Figure 9.14.

Figure 9.14 shows that the score T1 for the acoustic model and T1 for the process data have similar trend lines for the start-up procedure on the 20 February 2001, however the acoustic T1 has a significantly faster response than T1 for the process data when the fluidized bed is filled with granules.

The same comparison has also been made with data from 28 February 2001, and the results are shown in Figure 9.15 for the acoustic data, and Figure 9.16 for the process data.

After injection starts, but before the granulator reaches normal operating conditions, the airflow drops suddenly for approximately 10–15 min. The drop in the airflow can easily be seen in the score plot for the acoustic measurements. The score plot of the process data recorded simultaneously is displayed in Figure 9.16 and does not detect the change.

In general, these traditional process monitoring data are more variable than their smoother acoustic chemometric counterparts. The latter are in general also more informative; compare for example the score plot for the process data in Figure 9.16, in which the sudden drop in the airflow rate cannot be seen, to the score plot for the acoustic data (Figure 9.15). A plot of T2 for the acoustic model and T2 for the process data can be seen in Figure 9.17.

The plot in Figure 9.17 shows clearly that the acoustic data have a faster response to sudden changes in the airflow rate to the fluidized bed reactor. This means that monitoring the start-up sequence with acoustic chemometrics could indeed be an added value to the operators.

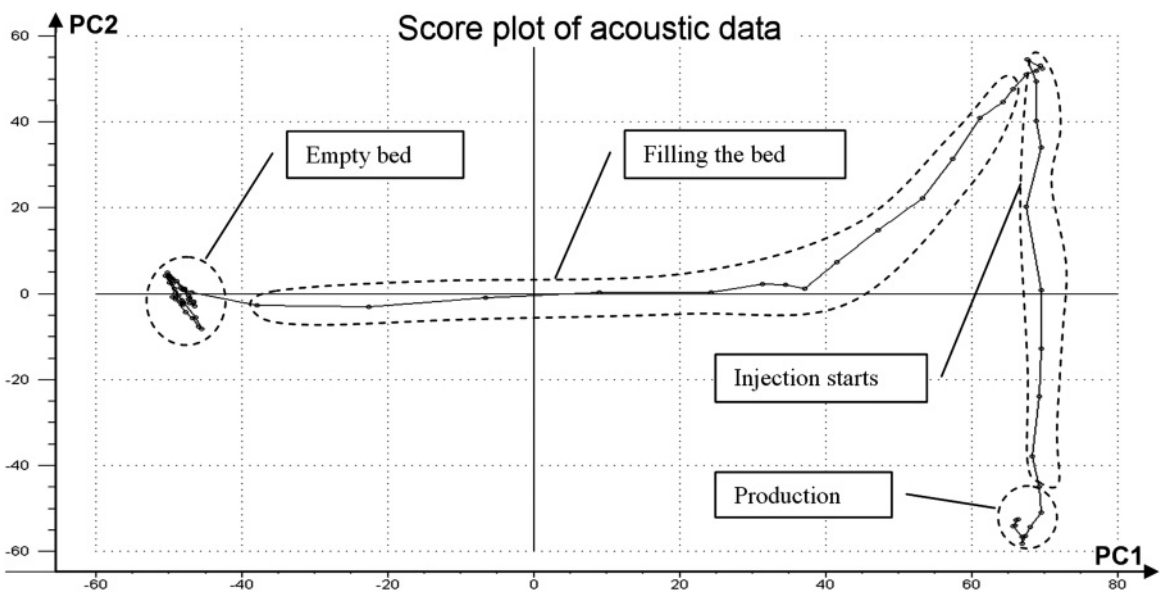


Figure 9.12 *Acoustic based score plot of pricnipal component (PC) 1 and 2 measurements recorded during start-up of the granulator 20 February 2001. Compare with Figure 9.13.*

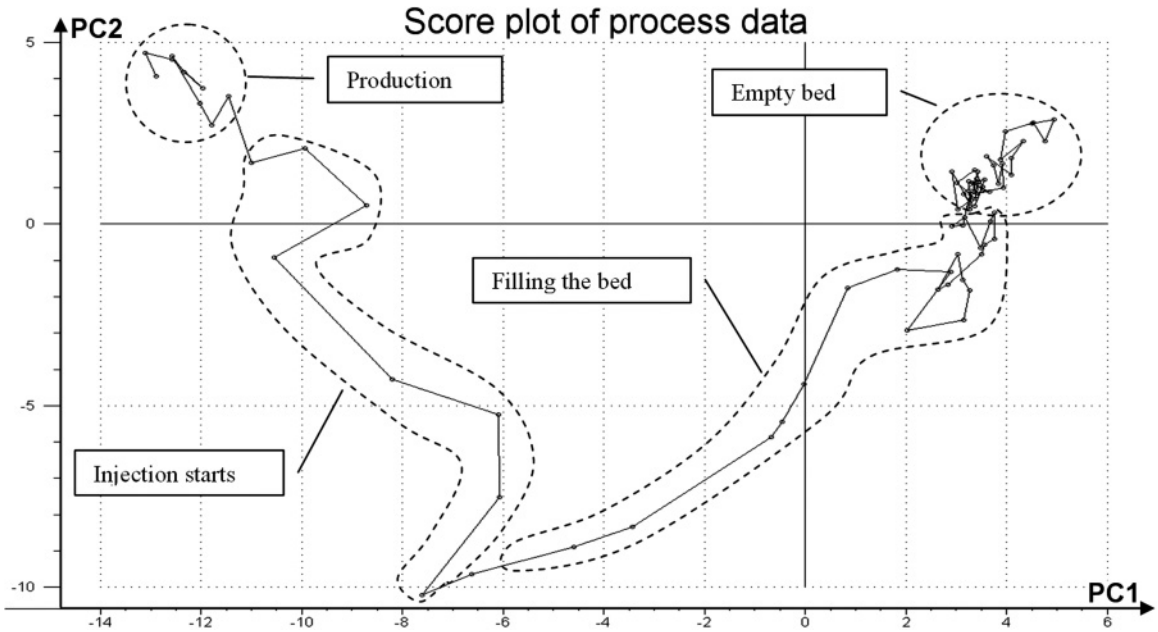


Figure 9.13 *Conventional process data-based score plot of PC 1 and 2 recorded during start-up of the granulator 20. February 2001. Compare with Figure 9.12.*

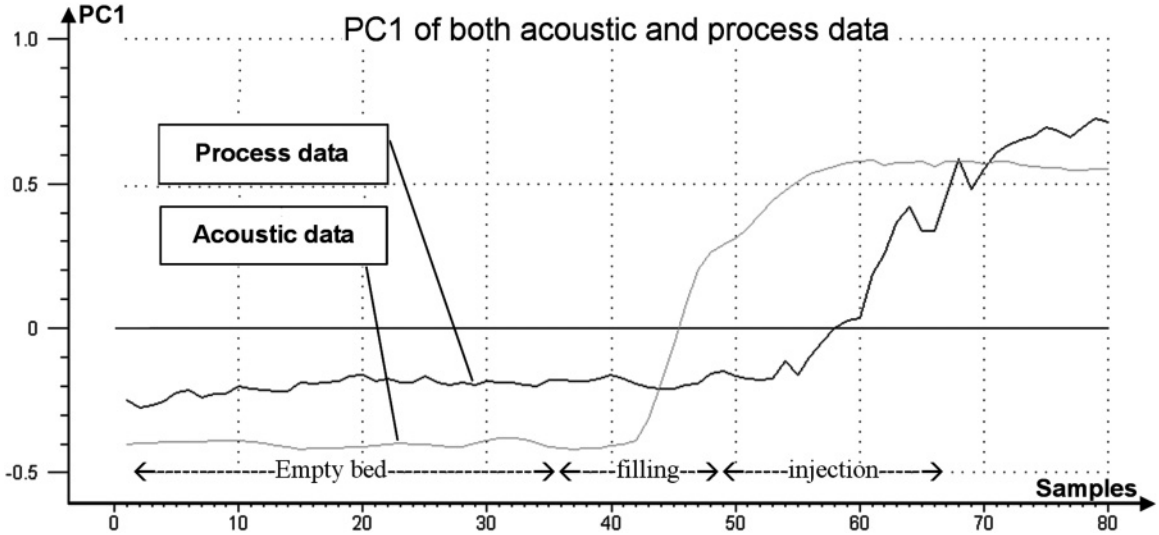


Figure 9.14 *Synoptic comparison of PC1 evolution. T1 for acoustic data compared to T1 for the process data 20 February 2001. The score vectors were normalized for the sake of comparison.*

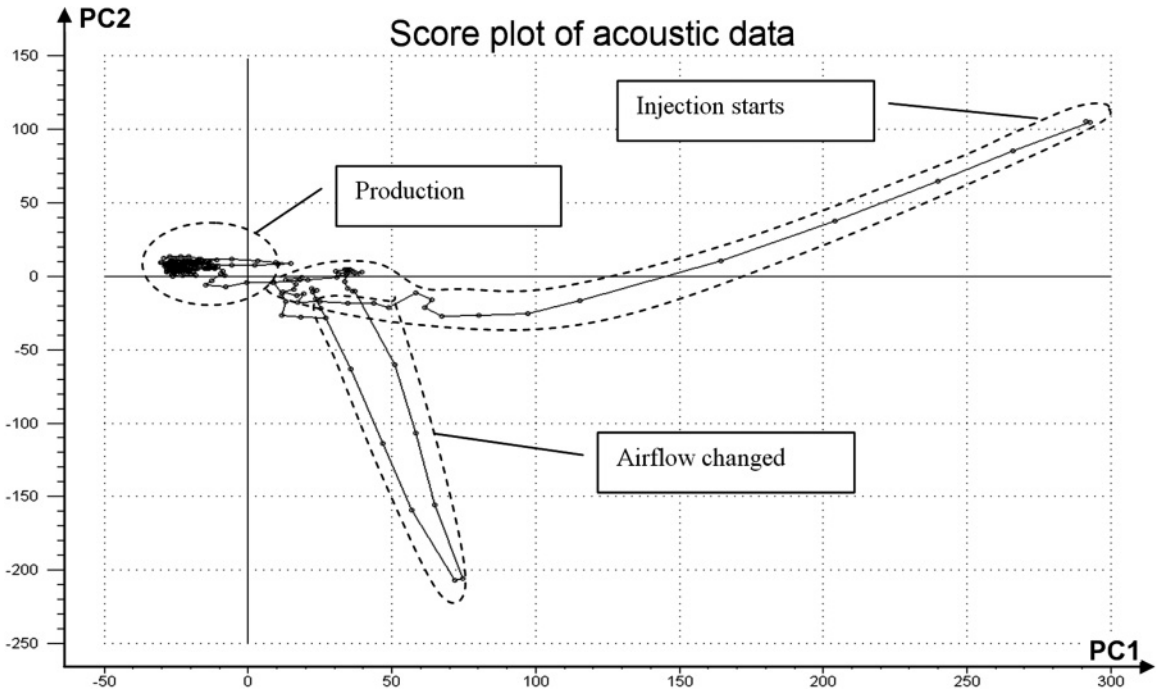


Figure 9.15 *Acoustic-based score plot of PC 1 and 2 of measurements recorded during start-up of the granulator 28 February 2001 (case of malfunctioning).*

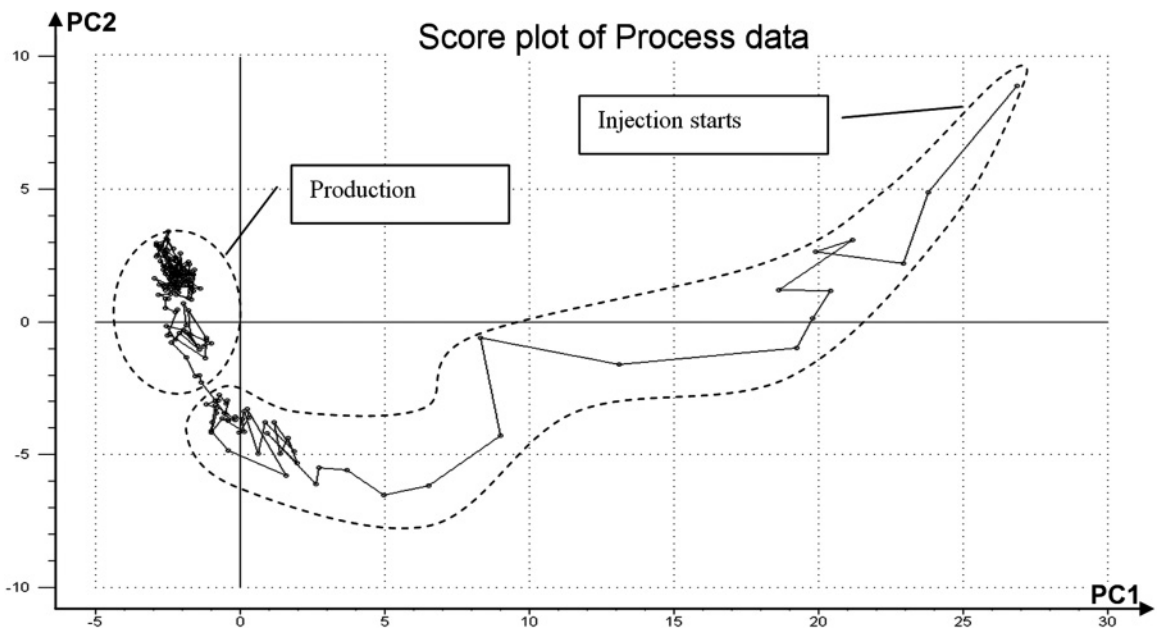


Figure 9.16 *Conventional process data-based score plot of PC 1 and 2 from measurements recorded during start-up of the granulator 28 February 2001. This is illustrative of a process malfunction. This is a new type of operator display based on acoustic chemometrics.*

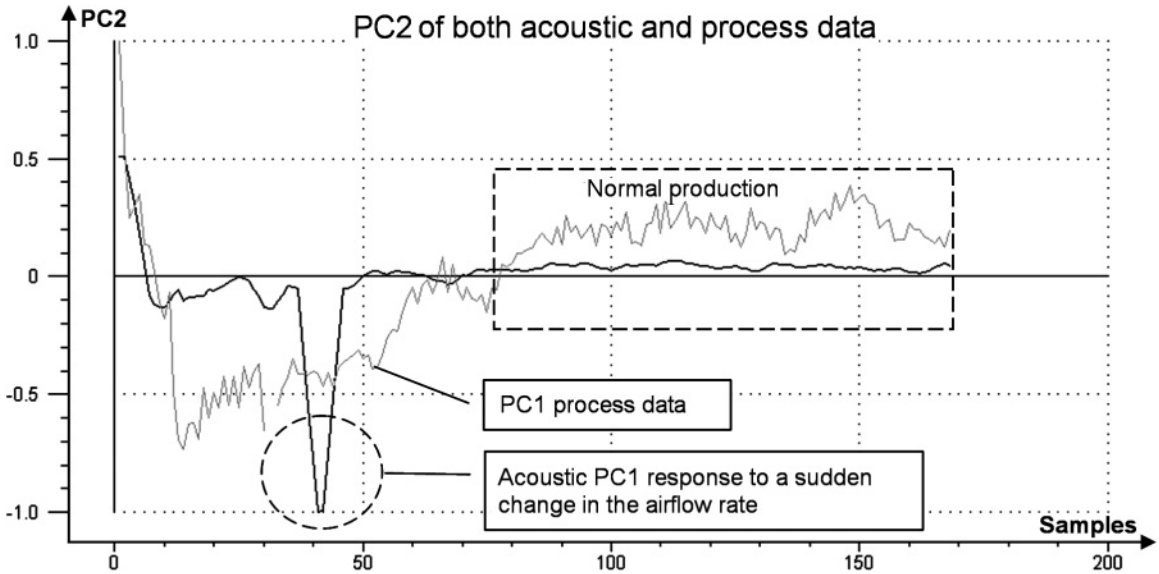


Figure 9.17 *Synoptic comparison of PC2 evolution. T2 for acoustic model and T2 for process data 28 February 2001. The score vectors were normalized for easier interpretation.*

9.3.4 Process monitoring as an early warning of critical shutdown situations

One major objective of this feasibility was to assess the potential of acoustic chemometrics to monitor the general process state of the granulation reactor in order to give reliable 'early warning' if a critical situation occurs in the bed. Critical situations in the fluidized bed are often a result of lump formation and/or layering on the perforated bottom plate of the reactor (see Figure 9.5).

During a run on 15 February 2001 a critical situation did occur in the pilot plant granulator, which was not according to the experimental plan, and thus very 'welcome' for the purpose of showing the value of acoustic chemometrics. A lump formation had been building over several hours and suddenly resulted in an uncontrolled shutdown around 16:30 the same day. Analysis of the acoustic data recorded in the period immediately leading up to this critical period shows an interesting trendline (Figure 9.18). The results presented in Figure 9.18 are based on data from the last 7 hours before the critical shutdown occurred.

From the process trend shown in Figure 9.18 it can be seen that an operator (in an on-line situation) would have had early warning at least 30 min before the shutdown if using the acoustic data (significant deviation from previous trend trajectory is observed). According to the local process-specific operator experience, this time interval would be enough time to take the necessary preventive action to bring the granulator back into normal conditions and avoid an uncontrolled shutdown. Process data from the same time period were also analyzed with PCA. The scores from PC 1 are plotted in Figure 9.20. The first warning shows up at 16:20 as a change in the process data trends, which is probably too late to avoid a critical shutdown of the reactor based on the same process operator consensus. The results from the principal component analysis of the acoustic data can be seen in Figure 9.18 for comparison.

A comparison of Figures 9.18 and 9.19 shows that the acoustic chemometric approach is much more sensitive to changes in the process state(s) of the fluidized bed than the traditional process data alone. Of course an industrial implementation of this process monitoring facility would include both acoustic data and process data, together with relevant chemometric data analysis (PCA, PLS) and the resulting appropriate plots.

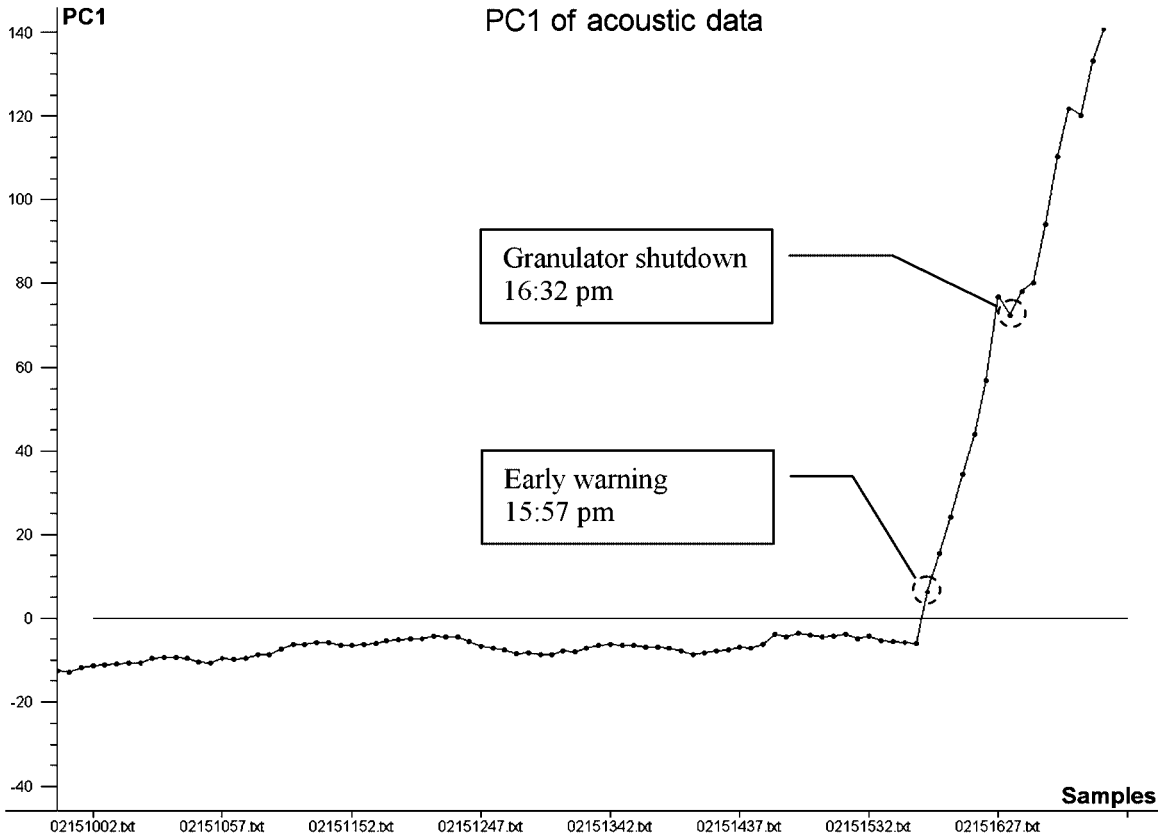


Figure 9.18 *Plot of PC 1. Sensor B was used as the lump formation started in chamber 1. Acoustic warning appears as early as 15:57, approximately 30 min before reactor shutdown. Compare with Figure 9.19, which shows the same situation based on traditional process measurements only.*

Figure 9.20 shows loadings from a PCA of data from all four acoustic sensors A, B, C and D, which clearly shows that the progress in PC 1 in Figure 9.20 is caused by lump formation in chamber 1 measured by sensor B as indicated by the relatively high loading values for this sensor.

9.3.5 Acoustic chemometrics for fluid flow quantification

Acoustic emission from fluid flow through an orifice plate inserted in a pipeline contains a wealth of information, which can be used to predict, for example composition, flow or density [5]. Acoustic signatures from fluid flow are affected by several physical factors: such as flow rate; differential pressure over the orifice plate; static pressure; as well as chemical–physical factors – density, composition, viscosity. It is the objective of PLS modeling to extract the relevant features from the acoustic spectra and make use of these embedded signals in *indirect* multivariate calibration [1,2]. Several successful examples, including prediction of trace concentrations of oil in water, have been reported [5].

Figure 9.21 shows liquid flow through an orifice plate. The relatively high differential pressure loss leads to powerful turbulent flow and cavitation downstream of the orifice plate. Cavitation occurs when the local

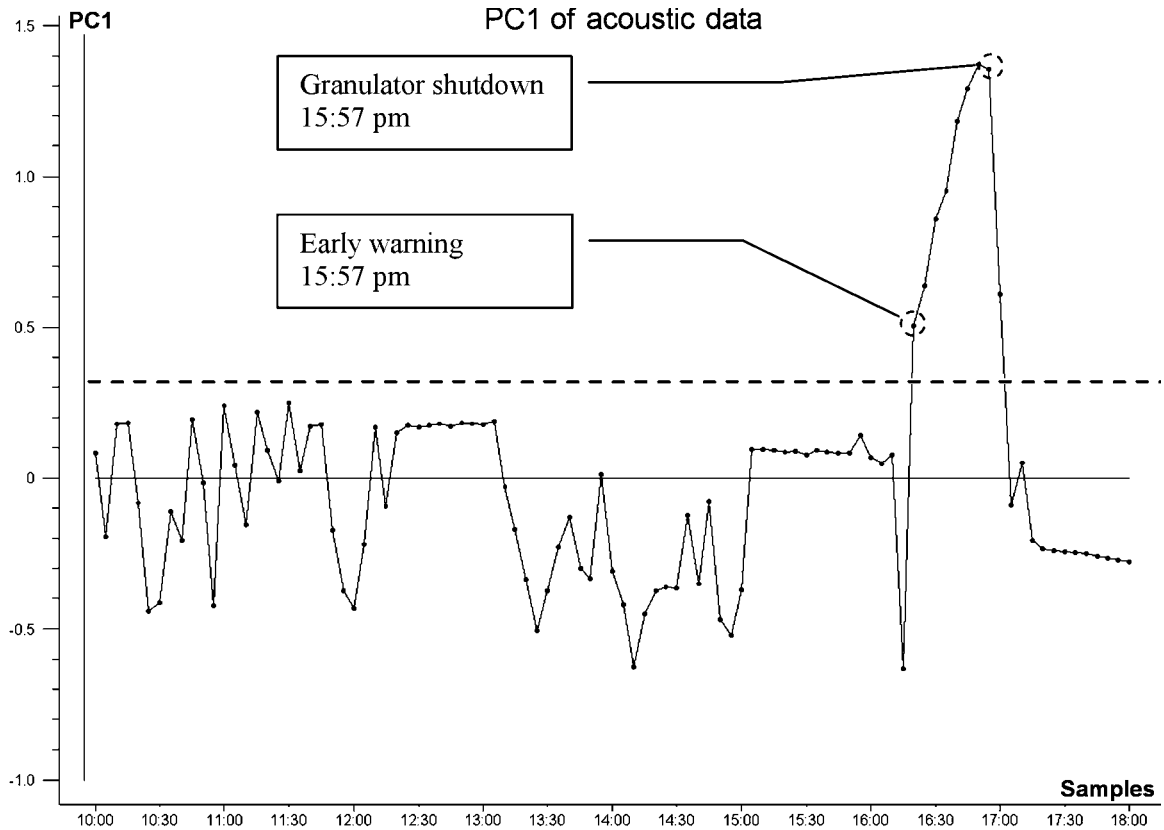


Figure 9.19 *Plot of PC 1 based on traditional process measurements alone. The warning first appears 16:20, only 10 min before shutdown of the reactor, which is insufficient for process recovery.*

hydrostatic pressure is lower than the vapor pressure and microbubbles are generated, as can be seen in Figure 9.21. The vibrations (often audible sound, hence the name 'acoustic' chemometrics) produced by microbubbles are among other factors affected by surface tension, viscosity and by both static and differential pressure. Oil in water reduces surface tension of the water and generates a dramatic change in the vibration signals emitted from the orifice. The vibrations can be recorded using sensors easily mounted on the surface of the pipeline in a convenient *nonintrusive mode*, which means it is not necessary to modify the process equipment.

9.3.5.1 Ammonia Concentration – Industrial Application Example

For industrial fertilizer production reliable ammonia concentration data are essential. An experimental setup for acoustic chemometric prediction of ammonia concentration has been tested in a full-scale industrial plant. Figure 9.22 shows a bypass loop with the orifice plate. The acoustic sensor was again mounted onto the orifice plate [5]. To ensure constant differential pressure and temperature of the ammonia flow, two pressure transmitters and one temperature sensor were used. Reference samples were taken at the sample valve shown in Figure 9.22.

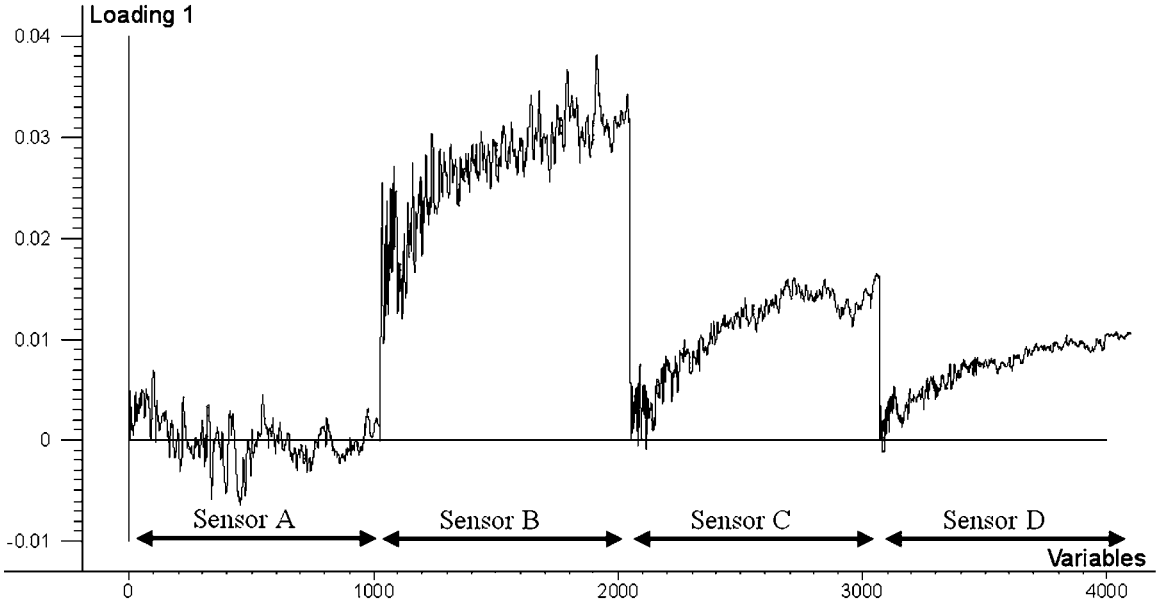


Figure 9.20 *Loading plot for PC 1 shows high values for sensor B which shows that the lump formation started in chamber 1 (sensor B).*

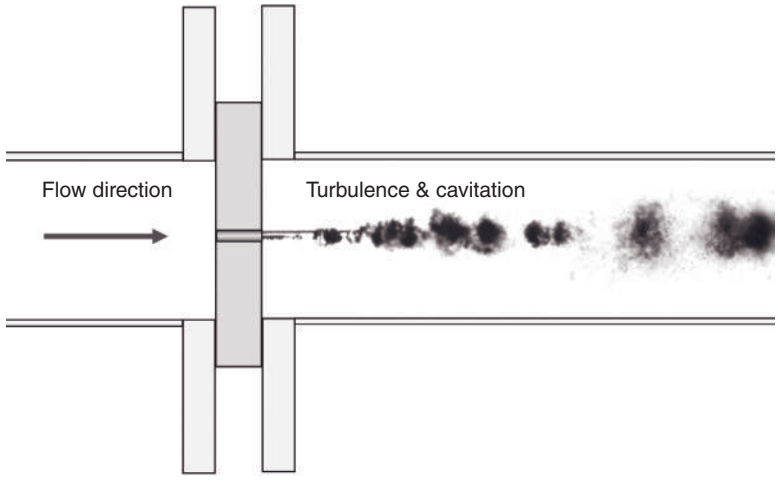


Figure 9.21 *Liquid flow through an orifice plate. Note turbulent flow conditions downstream generated by the high pressure loss – also causing the local pressure to drop below the vapor pressure producing cavitation (microbubbles).*

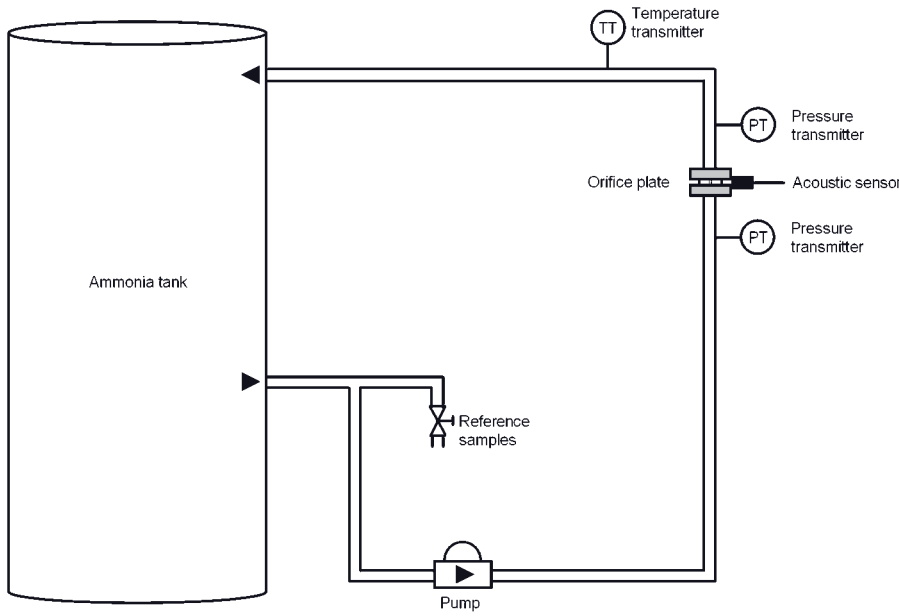


Figure 9.22 *Experimental setup for assessment of acoustic chemometrics potential to predict concentration of ammonia directly from acoustic emission generated by cavitation/turbulent flow.*

9.3.5.2 Ammonia Concentration Calibration and Results

Acoustic spectra were calibrated using PLS regression with six ammonia concentration levels. The reference concentration levels were 0, 0.5, 1, 2, 5 and 8% ammonia, with five replicate measurements at each level. Figure 9.23 shows the PLS-R prediction results validated with two-segment cross-validation (commonly referred to as a test set switch).

NB This pilot study included only six concentration levels; with a five-component PLS model there is a virtual certainty of over fitting the model. Even a two-segment cross validation is no absolute guarantee [2], but does supply the only possible validation basis with any credibility. However, the results in Figure 9.23 indicate and substantiate satisfactory possibilities for continuing to the type of extended calibration work needed in a full-fledged industrial calibration setting.

9.4 Available On-line Acoustic Chemometric Equipment

The equipment needed for acoustic chemometric monitoring can be divided into three categories:

- hardware (sensors, cables, data acquisition unit, amplifier)
- data acquisition software (acoustic recording and digital signal processing)
- chemometric software (analysis and calibration of multivariate models).

The hardware will often be bought as a selected suite of parts from different OEMs or suppliers which together will make up a complete system. Another option – which is better for industrial use – is to buy a commercially available complete solution (e.g. as illustrated in Figure 9.24).

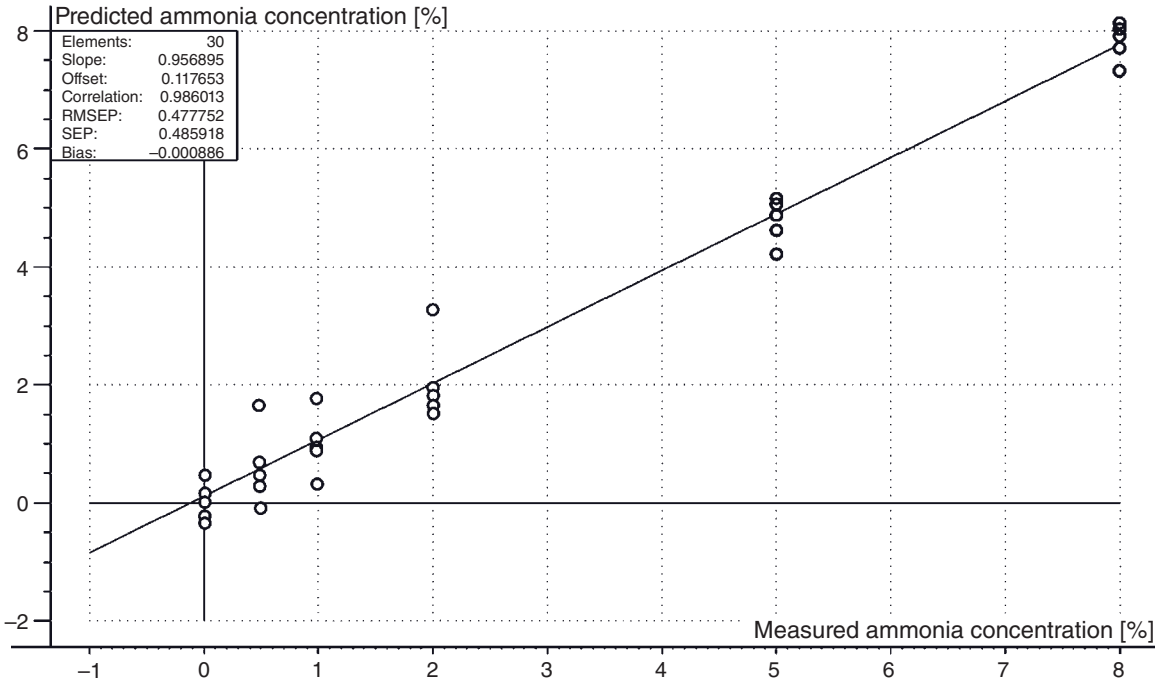


Figure 9.23 *Prediction results for ammonia, validated with two-segment cross validation (test set switch). Slope = 0.96. RMSEP = 0.48% ammonia.*

Figure 9.24 *Process Signature Analyser PSA-100i, Left: laboratory version Right: industrial version.*

The advantages of a turnkey solution are obvious: minimal systems integration and local programming, proven and robust equipment with explosion-proof certification, etc.

The turnkey system illustrated in Figure 9.24 can be mounted close to the sensor which will reduce potential electrical noise picked up by the cable connecting the sensor to the rest of the equipment. The equipment shown can also be used without a computer as a 'stand alone measurement transmitter'. The

multivariate prediction model can be uploaded to the analyser and the predicted parameter of interest transmitted as an industrial standard 4–20 mA signal. There are several such complete systems available on the market.

9.5 Discussion

9.5.1 Granulator monitoring

Sensor positions A (mounting on the orifice plate) and B (mounting on the wall in chamber 1) showed the overall best results for the industrial granulator pilot plant. Acoustic chemometrics could be used to predict crystallization temperature of liquid urea, fluidization airflow, granule moisture content, ammonia concentration and general process states. During the total experimental trial several experimental parameters were changed so that the conditions were only comparable within a timeframe of approximately 1 week. The prediction results for all parameters investigated in this study gave satisfactory results in this context.

We conclude that acoustic chemometrics provide the process operators with useful information which can be used to run the process with fewer critical and work intensive shutdowns.

All experiments were set up so as to be realistic scenarios with respect to multivariate statistical process control (MSPC). The pilot plant granulator is operated exactly as the industrial scale counterpart, but most of the experiments included much more severe variation than will usually be found in an otherwise stable industrial production situation of similar duration (e.g. it is *normally* not necessary to change nozzle types, or to change product types within such short intervals).

An independent issue is that all the most important parameter validations are *both* based on identical 10-segment cross-validations as well as proper test sets. The prediction results for granule moisture content have here been displayed in a time-dependent fashion, which is most relevant for the industrial process operators.

9.5.2 Process state monitoring

Different types of data visualization techniques have been presented to a selection of process operators, normal operating conditions (NOC), clogging of nozzles, temperature excursions (too high, too low), moisture deviations, unplanned level changes, pipeline blocking. Most experience gained from developing these techniques together with industrial personnel has shown that score cross-plots (e.g. score 1 versus score 2 plots) used to monitor a granulation process presented in one of our earlier papers [7] were felt to be too complex for on-line interpretation by process operators. In this chapter we therefore present the results as one vector score plots (Figures 9.18 and 9.20) which generally are more relevant and often provide a better understanding within this critical target group. With a minimum of chemometric training, this result need not at all be fixed and universal, however.

9.5.3 Ammonia concentration monitoring

The full-scale industrial experiment demonstrated the feasibility of a convenient, nonintrusive acoustic chemometric facility for reliable ammonia concentration prediction. The training experimental design spanned the industrial concentration range of interest (0–8%). Two-segment cross-validation (test set switch) showed good accuracy (slope 0.96) combined with a satisfactory RMSEP. It is fully possible to further develop this pilot study calibration basis until a full industrial model has been achieved. There would appear to be several types of analogous chemical analytes in other process technological contexts, which may be similarly approached by acoustic chemometrics.

9.6 Conclusions

The feasibility survey presented here, together with papers [3–7], demonstrates the conclusion of an important first step transforming on-line MSPC acoustic chemometrics from a new experimental concept towards a proven PAT technology. Acoustic chemometrics appears with a broad and versatile application horizon in process technology for a range of physical and chemical parameters: composition (earlier studies looked at oil, fat, ammonia, glycol, ethanol); mixing fractions; mixing progress; fiber length; flow; density; temperature; and system state. There are no doubt many other candidate analytes as well; a certain amount of experimental calibration curiosity is required.

All multivariate calibrations must be based on empirical training and validation data sets obtained in fully realistic situations; acoustic chemometrics is no exception. Many models are in addition based on indirect multivariate calibration. All industrial applications must always be evaluated only based on test set validation. Reference [2] deals extensively with the merits of the various validation methods, notably when it is admissible, and when not, to use cross-validation. See also Chapters 3 and 12, which give further background for the test set imperative in light of typical material heterogeneity and the 'Theory of Sampling'.

References

1. H. Martens and T. Næs, *Multivariate Calibration*, John Wiley & Sons, Ltd, Chichester, 1989 (reprint 1994).
2. K.H. Esbensen, *Multivariate Analysis – in practice*, 5th edition, Camo AS, Oslo, 2001.
3. K.H. Esbensen, M. Halstensen, T.T. Lied, A. Saudland, J. Svalestuen, S. de Silva, B. Hope, Acoustic chemometrics – from noise to information, *Chemom. Intell. Lab. Syst.*, **44**, 61–76 (1998).
4. M. Halstensen, S. de Silva, K.H. Esbensen, Acoustic monitoring of pneumatic transport lines – from noise to information, *KONA Powder and Particle*, **16**, 170–178 (1998).
5. K.H. Esbensen, B. Hope, T.T. Lied, M. Halstensen, T. Gravermoen, K. Sundberg, Acoustic chemometrics for fluid flow quantifications -II: A small constriction will go a long way, *J. Chemom.*, **13**, 209–236 (1999).
6. M. Halstensen, K.H. Esbensen, New developments in acoustic chemometric prediction of particle size distribution – the problem is the solution, *J. Chemom.*, **14**, 463–481 (2000).
7. M. Halstensen, P. de Bakker, K.H. Esbensen, Acoustic chemometric monitoring of an industrial granulation production process – a PAT feasibility study, *Chemom. Intell. Lab. Syst.*, **84**, 88–97 (2006).
8. M. Whitaker, G.R. Baker, J. Westrup, P.A. Goulding, D.R. Rudd, R.M. Belchamber, M.P. Collins, Application of acoustic emission to the monitoring and end point determination of a high shear granulation process, *Int. J. Pharm.*, **205**, 79–92 (2000).
9. G.R. Flåten, R.M. Belchamber, M. Collins, A.D. Walmsley, Caterpillar – an adaptive algorithm for detecting process changes from acoustic emission signals, *Anal. Chim. Acta*, **544**, 280–291 (2005).
10. A. Kupyna, E.O. Rukke, R.B. Schüller, H. Helland, T. Isaksson, Partial least square regression on frequency shift applied to passive acoustic emission spectra, *J. Chemom.*, **21**, 108–116 (2007).
11. D. Daniher, L. Briens, A. Tallevi, End-point detection in high-shear granulation using sound and vibration signal analysis, *Powder Technol.*, **181**, Issue 2, 130–136 6 (2008).
12. I.C. Ifeachor and B.W. Jervis, *Digital Signal Processing, a Practical Approach*, pp. 47–102, Addison-Wesley Publishing Company, Wokingham, 1993.
13. D. Kunii, O. Levenspiel, *Fluidisation Engineering*, 2nd edition, Butterworth Heinemann Publishing, Stoneham, 1991.
14. P. Gy, *Sampling for Analytical Purposes*, John Wiley & Sons, Ltd, Chichester, 1998.
15. L. Petersen, P. Minkkinen, K.H. Esbensen, Representative sampling for reliable data analysis: Theory of sampling, *Chemom. Intell. Lab. Syst.*, **77**, 261–277 (2005).

10

Process NMR Spectroscopy: Technology and On-line Applications

John C. Edwards and Paul J. Giammatteo

Process NMR Associates, LLC, Danbury, CT, USA

10.1 Introduction

Nuclear magnetic resonance (NMR) spectroscopy is perhaps the last organic spectroscopy to actually make the transition from laboratory instrument to on-line process control sensor. For decades NMR has been used as a lab method for at-line or quality control (QC) testing. Most of this routine QC testing has taken place on time-domain (TD) NMR analyzers manufactured by Bruker Optics,[1] Oxford Instruments,[2] Resonance Systems,[3] etc. These TD-NMR spectrometers are fairly simplistic in their approach to NMR experimentation in that the magnets are relatively inhomogeneous, leading to NMR signals that yield only total proton intensity, relaxation times, and their distributions. No chemical shift information is obtained that can then be used to derive relative chemistry distributions of the protons in the sample. Rather, the absolute signal intensity is used to determine the hydrogen content, or the total concentration of a phase in the sample relative to a calibration standard. The raw NMR signal, called the free induction decay (FID), can be used to determine the relative amounts of liquid and solid-state components, or different phases in a given sample. It can also be utilized to derive the relaxation times and their distributions for sample components, for sample components that can then be related to molecular mobility and other related phenomena such as diffusion and pore size distributions. In the era from the first commercial NMRs marketed in the 1950s to the late 1980s, several ASTM test methodologies were developed for the TD-NMR instrumentation. Applications for on-line control were demonstrated in principle but no true applications of the technology were placed on-line for control. These spectrometers operate typically in the 0.05–0.5 Tesla range which equates to a 2–20 MHz resonance frequency for ^{1}H nuclei. To date the vast majority of the analyses have involved protons, but there have been demonstrations of the techniques applied to ^{19}F, ^{27}Al, and ^{23}Na.

What is striking about NMR in the process analytical field is how few peer-reviewed articles actually appear in the literature. One reason for this is the perceived complexity of NMR as an analytical technique. A second reason is the small number of companies actually engaged in the production of appropriately

Process Analytical Technology 2e Edited by Katherine Bakeev
© 2010 John Wiley & Sons, Ltd.

designed NMR analyzers. Though NMR is well established as the ultimate molecular structural analysis tool for analytical chemists, market forces have driven NMR manufacturers such as Varian,[4] Bruker,[1] and JEOL[5] to produce instruments with ever higher magnetic fields to increase spectral resolution and sensitivity. Their market has become the pharmaceutical research establishments, and NMR systems that operate at frequencies below 100 MHz (permanent magnets) are simply not perceived as useful or marketable. In fact, it is very difficult today to buy a spectrometer from these traditional manufacturers at a frequency lower than 400 MHz. It has been left to smaller companies to develop the lower field NMR devices, and there are simply not many companies that have the expertise to develop such a product. This has led to a situation where research establishments, both academic and industrial, simply do not possess the required NMR hardware to perform publishable process NMR articles. The lack of peer-reviewed articles also is due to the practicalities of NMR as a process analytical technique. Most on-line applications have essentially been implemented in a proprietary manner so that the users can derive the competitive benefits of improved control by use of on-line NMR. This is not to say that one cannot find articles on what is called 'on-line NMR' by the superconducting NMR community. These articles typically cover liquid chromatography (LC)-NMR where the NMR probe is coupled directly to an high performance liquid chromatography (HPLC) chromatograph,[6] and some of them cover the direct coupling of a reaction vessel in a fume hood via an HPLC pump to an LC-NMR probe on a superconducting NMR system in a laboratory environment.[7–13] None of these, however, can be considered to be truly on-line, real-time applications in a manufacturing plant environment. An excellent review of these types of applications is given in a 'process NMR' review article authored by faculty at the University of Strathclyde.[14]

In the 1990s the first true on-line NMR analyzers were placed into petrochemical plants and utilized for control. The first NMR technology to be placed on-line in a manufacturing facility was a TD-NMR system developed by Auburn International for measuring the melt index of polyethylene and polypropylene.[15] This technology was widely accepted by the industry and over 100 implementations of it are still in manufacturing operations around the world. This product line is still marketed and supported by Progression, Inc. In the 1990s another small NMR company, Process Control Technology,[16] operating out of Fort Collins, CO, developed a similar TD-NMR spectrometer and implemented on-line applications, though none of these applications are documented. In the past few years with the development of digital processing technologies in the NMR arena a number of companies have emerged selling moderately priced TD-NMR spectrometers ($20 k–60 k). Some of these companies do not actually manufacture accompanying magnets and probes that would allow them to market entire functional NMR systems. Some of these companies are: Spincore,[17] Resonance Systems,[3] Process NMR Associates,[18] NMR Process Systems,[19] Magritek,[20] MR Resources,[21] SpinLock,[22] Tecmag,[23] ABQMR,[24] Stelar,[25] Bruker Minispec,[1] Oxford Instruments,[2] and Progression, Inc.[26]

The TD-NMR technique has also been modified to produce unilateral NMR devices that involve side by side permanent magnets that project a magnetic field into the object being analyzed to produce an external homogeneous volume several millimeters below the object surface. A surface coil is utilized to pulse radio-frequency (RF) energy into the object and the same coil then picks up the weak FID from the protons in the sample. This technique is being investigated in the food and packaging industries to monitor the quality of packaged foods and other materials through the packaging materials. There is also potential for this to be utilized in rubber manufacturing with the NMR monitoring physical properties of tires as they are produced. There are a number of challenges that may limit the true applicability of this NMR technology for on-line analysis and control. These unilateral NMR devices are sometimes called the NMR MOUSE[27–29] and they are marketed by AixNMR,[30] Bruker Optics,[1] Resonance Systems,[3] and Magritek.[20]

A true on-line application of NMR that is widely used in the field and based on this unilateral NMR technology is downhole NMR sensing performed by the gas and oil exploration and production companies such as Schlumberger, Halliburton, Baker Hughes, and Chevron. These tools are typically provided as a

service to the exploration company's customers and are heavily protected by patents and held as proprietary techniques. The basic technology is a large downhole tool that contains a ruggedized spectrometer and unilateral magnet/probe system. The technique utilizes NMR spectroscopy to observe FID and Carr–Purcell–Meiboom–Gill (CPMG) NMR data obtained on the rock formation which the instrument's unilateral magnet is contacting. Analysis of water, oil, and gas content, as well as diffusion coefficients and porosity, are derived from the bulk NMR signal and relaxation data.[31–32] The technique has become particularly useful when relaxation data and diffusivity data can be obtained and displayed simultaneously via new multipulse techniques analyzed by two-dimensional Laplace transformations.[33–34]

On the higher end of the NMR technologies one finds high-resolution (HR) NMR which is the spectroscopy that most chemists are familiar with. This technology involves the use of highly homogeneous magnets that yield NMR spectra that contain the chemical type distribution of the protons in the sample. In research grade instruments the magnetic fields typically run from 7 Tesla up to 21 Tesla, and instrumentation costs anywhere from $200 k–$3000 k. These research instruments are based on superconducting magnet technology that involves expensive alloy magnet coils and liquid helium and nitrogen cryogenic cooling. Many applications of 'on-line NMR' have been presented in the literature based on these research grade instruments but these only actually represent a hyphenation of LC and NMR. There have been several papers published on reaction monitoring by NMR but in the context of on-line NMR they have simply been small scale demonstrations of reactions performed in a fume hood being monitored by LC-NMR. These techniques are not considered to be true on-line process NMR applications. They are highlighted here because they can be considered at-line measurements implemented in a batch reaction environment at a pharmaceutical or specialty chemical facility.[7–14]

In the mid 1990s Elbit-ATI of Israel developed permanent magnet technology that could produce homogeneous fields at 1.4 Tesla. This technology enabled the development of a field-deployable spectrometer capable of obtaining HR-NMR spectra that yield detailed chemical shift information from the protons in the sample. Several applications were demonstrated with Texaco R&D implementing two of these at their Los Angeles refinery – for fuel gas BTU[35] and acid alkylation control.[36] In 1997 the Elbit-ATI NMR company was purchased by the control company Foxboro (now part of the Invensys conglomerate) and began to operate under the name FoxboroNMR Ltd. From 1997–2005 Invensys operated under an exclusive marketing arrangement and installed more than 30 NMR spectrometers in various refinery applications around the world. From 1997–2003 Process NMR Associates was the application development and technical marketing support company that was under contract to FoxboroNMR as they implemented the technology into the field. In 2005 FoxboroNMR Ltd. was spun out of Invensys and became Qualion, Inc.[37] They are currently marketing the process NMR product around the world through various distributors of which Process NMR Associates is one. Qualion is currently the only company that produces an on-line HR -NMR. The applications for this product have been focused almost entirely in the petroleum and petrochemical industries.

10.2 NMR Spectroscopy Overview

10.2.1 The NMR phenomenon

The NMR phenomenon is based on the magnetic properties of nuclei and their interactions with applied magnetic fields either from static fields or alternating RF fields. Quantum mechanically subatomic particles (protons and neutrons) have spin. In some nuclei these spins are paired and cancel each other out so that the nucleus of the atom has no overall spin: that is, when the number of protons and neutrons is equal. However, in many cases the sum of the number of protons and neutrons is an odd number, giving rise to

what are called half-integer spins. In the case of process NMR this is the only NMR case that we must consider. In the case of ^{1}H, ^{31}P, and ^{19}F nuclei the nuclear spin quantum number, I, is 1/2, while for ^{27}Al and ^{23}Na, I is 3/2. As the nuclei are spinning they possess angular momentum, V, as well as a nuclear charge. This gives rise to a magnetic moment, μ, such that:

$$\mu = \gamma V \tag{10.1}$$

where γ is the magnetogyric ratio, which is a constant for particular nuclei and may be viewed as the 'strength' of the magnetic moment. When placed in an external static magnetic field, B_o, the magnetic moments will align themselves relative to the magnetic field into $2I + 1$ energy states. For protons, there are two possible energy states that equate to aligning with or against the applied magnetic field (Figure 10.1).

The magnetic moments do not, however, align directly along the axis of the applied magnetic field. Rather the magnetic moments have a torque applied to them by the applied field. This causes the magnetic moments to rotate around the applied field direction in a motion referred to as Larmor precession. The rate of the precession, ω_o, is proportional to the magnetogyric ratio of the nucleus, which leads to different nuclei precessing around the applied magnetic field at different angular velocities or Larmor frequencies. The angular velocity (ω_o), and frequency (υ) of the precession are expressed as:

$$\omega_o = -\gamma B_o \tag{10.2}$$

$$\upsilon = \omega_o / 2\pi \tag{10.3}$$

Thus, the precession frequency is proportional to the 'magnetic strength' of the nuclear magnetic moment as well as applied magnetic field. As magnetic field strength increases the precession frequency increases. Table 10.1 shows the relationship between applied field B_o and the precession frequency of a single proton.

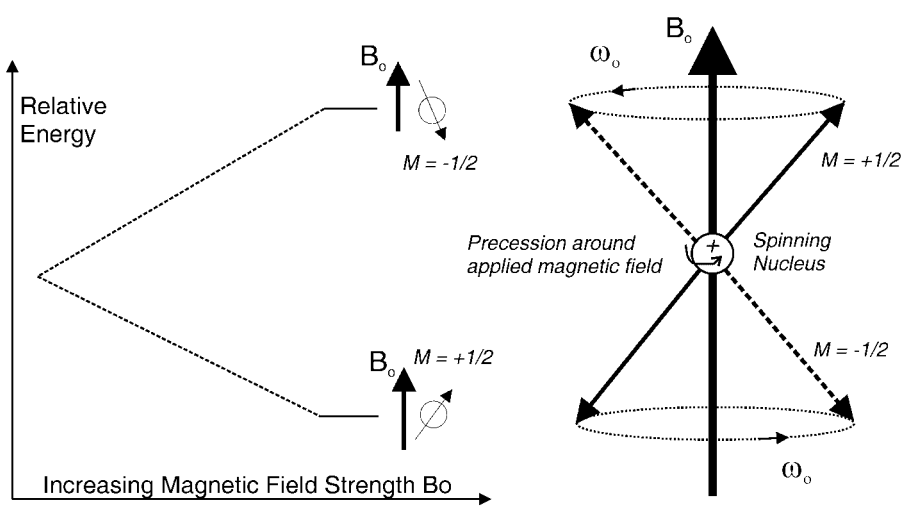


Figure 10.1 *Orientation of proton magnetic moments in an external static magnetic field and the precession of proton magnetic moments around the applied magnetic field.*

Table 10.1 *Precession frequencies of protons in different applied magnetic fields*

B_o (Tesla)	^{1}H resonance frequency (MHz)
0.47	20
1.41	60
7.05	300
11.75	500

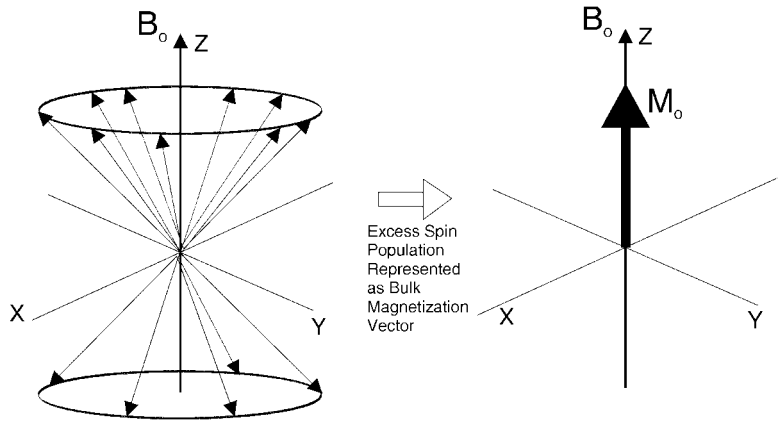


Figure 10.2 *Representation of excess spin population expressed as a bulk magnetization vector* M_o.

Now let us consider a collection of spins in the applied magnetic field. As can be seen from the energy diagram of the two spin-1/2 spin states, the alignment of spins parallel to the applied magnetic field (α) has a slightly lower energy which means that its spin population will be higher than in the antiparallel case (β). This difference in spin population between the two energy levels is known as the Boltzmann distribution and is affected by the difference in energy between the two levels (ΔE) and the temperature (T). The population ratio can be calculated as:

$$N_\alpha / N_\beta = \mathrm{e}^{-\Delta E / k_\mathrm{B} T} \tag{10.4}$$

where N is the number of spins in a given state, and k_B is the Boltzmann constant. The population differences are quite small and correspond only to a difference of 1 in 10^4 even at the highest magnetic fields. It is this small excess of nuclear spins that gives rise to the signal in an NMR experiment. The bulk of the spins cancel each other out and the excess population is summed together to form a bulk magnetization vector, $\mathbf{M}_o$ which is aligned directly along the applied magnetic field axis (Figure 10.2).

Now that we have described how the magnetic moments of the protons in a sample align themselves with the applied magnetic field and how the spin population difference can be represented as a bulk magnetization vector, we can now describe how this vector can be acted upon by applied RF pulses. When the spins are irradiated by an orthogonal RF field, B_1, oscillating at the Larmor frequency (in the case of a 1.4 Tesla NMR system this frequency for protons is 60 MHz), the spins actually begin to precess around the orthogonal B_1 field. The application of the RF field is achieved by creating a powerful square wave of RF radiation, a pulse, at the Larmor frequency of the nucleus to be observed. Over a time frame on the order of

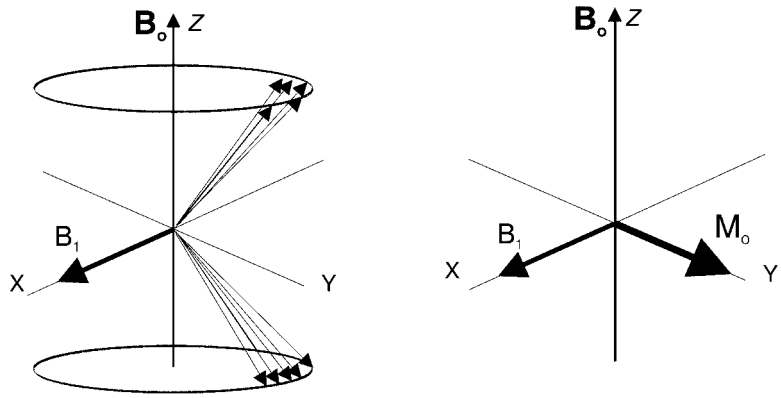


Figure 10.3 *The magnetic field generated by an RF pulse applies a torque to the bulk magnetization vector and drives it towards the transverse plane.*

microseconds the net magnetization vector will rotate into the transverse plane. This RF pulse is applied through a coil whose geometry dictates that the B_1 field be orthogonal to the static magnetic field B_o (Figure 10.3).

After a 90° pulse the bulk magnetization vector will be found on the *y* axis and there will be no magnetization vector observable along the *z* axis. This means that effectively the spin populations of the two energy levels are equal and the individual spins are bunched along the *y* axis in a phase coherent manner. The pulse length, *pw*, required to move the magnetization vector through 90° is typically in the 10–20 μs range. If the RF pulse is left on for twice as long, then the magnetization vector will move to an alignment along the $-z$ axis. This is a peculiar situation in which the spins effectively feel a –273 K temperature. In other words, the spin populations are completely inverted with the excess magnetization being found in the antiparallel (β) energy level. This spin inversion is called an 180° (π) pulse and it is typically twice the length of the 90° pulse. The NMR signal is derived from what happens to the spin system from the moment at which the RF pulse is switched off.

As soon as the RF pulse is turned off the spins that make up the bulk magnetization vector find themselves out of equilibrium with the static applied field B_o, and they begin to equilibrate back to the 'ground state' condition aligned with the magnetic field. The spin populations begin to return to the original condition with spins moving from the antiparallel to the parallel alignment. Thus, the bulk magnetization vector begins to precess around B_o and as the vector reduces in intensity along the *y* axis it simultaneously increases in intensity along the *z* axis. The rate at which the magnetization reappears along the *z* axis is measured as a relaxation time called T_1 (spin-lattice or longitudinal relaxation). At the same time that the spin populations are equilibrating the phase coherence in the transverse plane that was created by the 90° pulse disappears as the spins randomize their distribution in that transverse plane. The rate at which the coherence is lost is measured as a relaxation time called T_2 (spin–spin, or transverse relaxation). These relaxation times will be discussed later in the TD-NMR section of this chapter (Section 10.2.2), as they can be utilized to provide relative phase quantities, pore size distributions, mobility and crystallinity indicators, etc.

The precession that occurs in this time frame can be viewed as a movement of coherent spins inside the coil that the sample is sitting inside. The movement of these magnetic moments generates a weak oscillating voltage. The oscillation in the voltage is at frequencies at or around the Larmor frequency of the nucleus. This voltage represents the NMR signal and contains the chemical shift and relaxation information of all the spins in the sample. As such it contains as many oscillating frequencies as there are different chemistries

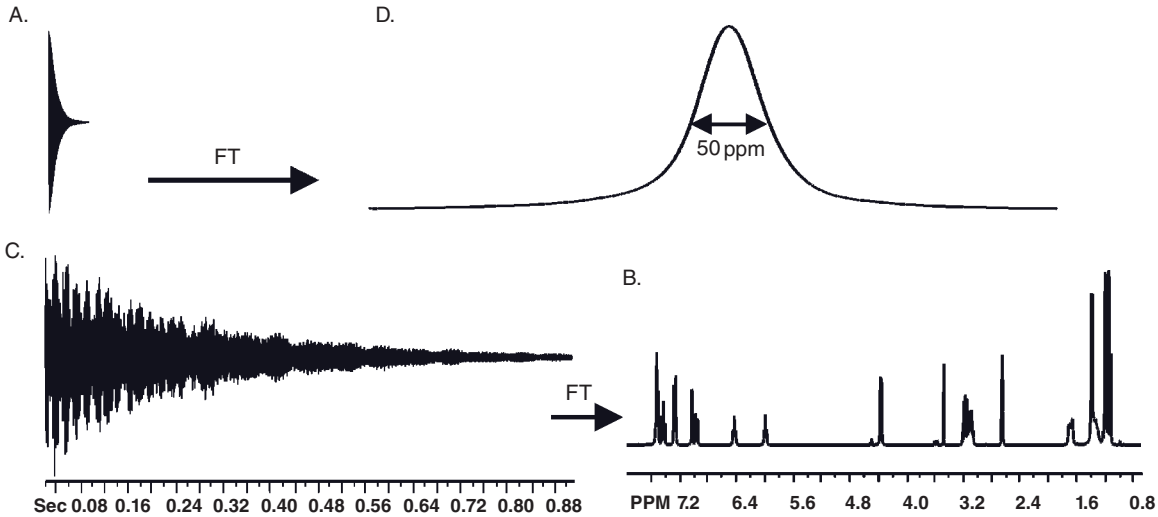


Figure 10.4 *Difference between TD-NMR and HR-NMR data. (A) TD-NMR, FID (×100), 2 ms acquisition, (B) TD-NMR, spectrum, (C) HR-NMR, FID, 900 ms acquisition, (D) HR-NMR, spectrum.*

in the sample, with the amplitude of those voltages being proportional to the absolute concentration of the particular phases or chemistries. The signal starts at a maximum signal intensity and slowly decays as the spin coherence is gradually lost in the transverse plane. This oscillating, decaying signal is called a free induction decay (FID). The FID is the basis for all NMR data and is treated differently depending on the type of NMR system from which it was generated. The FID is analogous to the interferogram obtained from infrared interferometers. Fourier transform (FT) of the FID yields a frequency domain spectrum (Figure 10.4).

TD-NMR is performed on NMR spectrometers that are equipped with lower magnetic field strength magnets with relatively low field homogeneity. Thus, relatively short FIDs on the order of a few milliseconds are obtained and FT of this signal yields broad lines from which no chemical detail can be obtained. However, the data is rich in information regarding the relative amounts of different phases that are present in a sample, such as water and oil, liquid and solid, crystalline and amorphous. The data can be approached in two ways – analysis of the FID or analysis of relaxation times and their relative distributions.

10.2.2 Time-domain-NMR: utilization of the FID and spin relaxation

Figure 10.5 shows the FID obtained on a soap sample. The initial points of the FID yield information on the absolute proton signal intensity which is related to the total proton content in the sample. After the initial few points of the FID however, one must be careful with how the data are viewed quantitatively. Protons in different physical environments will have very different relaxation times. In the FID the differences in transverse (T_2) relaxation are observed readily. In the case of the soap sample the FID contains a rapidly decaying signal at the beginning of the FID along with a signal that decays at a much slower rate.

Multiexponential fitting of the FID data allows the initial (time zero) intensities of the sample components giving rise to the different exponential decays to be determined. All NMR experiments have an initial dead time applied to the receiver before acquiring the NMR signal. This dead time delay protects the receiver from being exposed to high power levels as well as allowing the vibration of the coil (ringing) to dampen

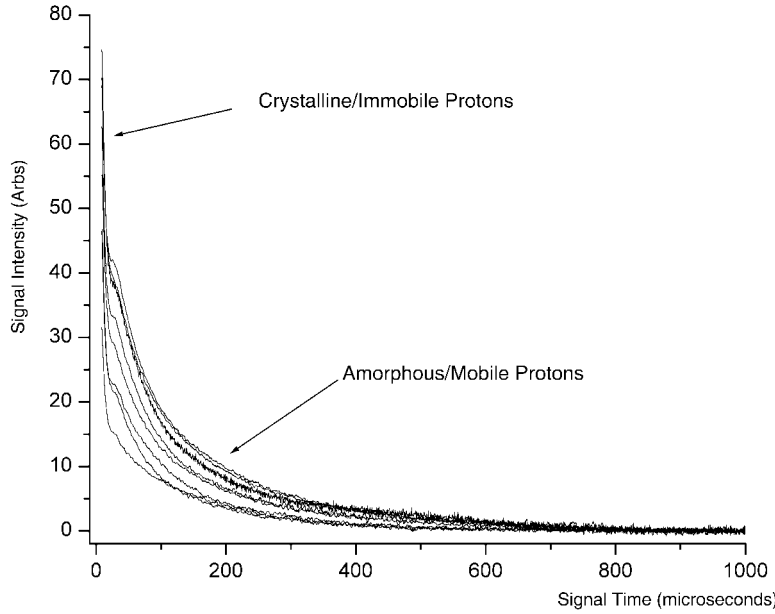


Figure 10.5 *Typical TD-NMR FIDs for heat treated soaps. Crystalline and amorphous with different relaxation properties are observed.*

after the high power RF pulse is switched off. The RF power induces harmonics in the coil which lead to strange distortions in the FID which cause large problems with quantitation and spectral baselines. It is a standard practice to linearly predict back to time zero in order to obtain truly quantitative data.

Transverse relaxation data (T_2) is also obtained on a routine basis and reveals much about the relative amount of various components that might be in a sample. T_2 relaxation is obtained by performing a multi-pulse experiment called the CPMG sequence,[38] named after its inventors. This sequence involves a $90°_x$ pulse and the collection of the first few points of the resulting FID. These first few points are averaged and stored as the first data point of the CPMG data set. The sequence then has a variable delay, τ, built into it during which the spins begin to lose their coherence in the transverse (xy) plane. An $180°_y$ pulse is applied after delay τ, which has the effect of inverting the spins in the transverse plane. However, the direction of coherence change remains the same for each spin, and this now leads to a refocusing of the spins along the y axis (Figures 10.6 and 10.7A). This refocusing creates an echo in the acquired signal and the points acquired at the top of the echo are averaged and stored as the second point of the CPMG data set. After another τ delay a second $180°_y$ pulse is applied and another echo is generated that provides the third point of the CPMG data set. This 2τ delay followed by an $180°_y$ pulse is repeated anywhere from 500–2000 times and leads to the development of a 500–2000 point CPMG data set (Figure 10.7B). Each echo will be slightly less intense than the previous one due to the complete relaxation of a subset of the spins in the time frame of the echo pulses. The CPMG data is an exponentially decaying signal that can contain a number of relaxation rates depending on the sample that is being observed. Sometimes they appear as a single exponential decay and at other times, when there are multiple components, phases, or diffusion rates they must be processed with multiexponential fitting algorithms.[39]

Often the data cannot be satisfactorily fit to multiple specific exponentials because the sample contains a natural distribution of exponentials centered on the specific exponential time decays utilized in the fitting procedure. When specific exponentials cannot adequately describe the data, Laplace inversion routines are

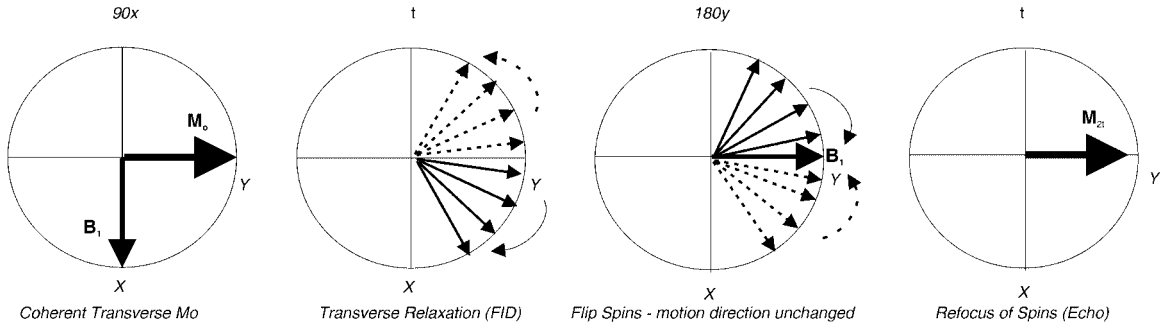


Figure 10.6 *Effect of CPMG pulse sequence on spin system – applying 90× pulse followed by series of 180° pulses. Note that 180° pulse inverts the defocused spins but their direction of motion remains the same, leading to a refocusing of the spins.*

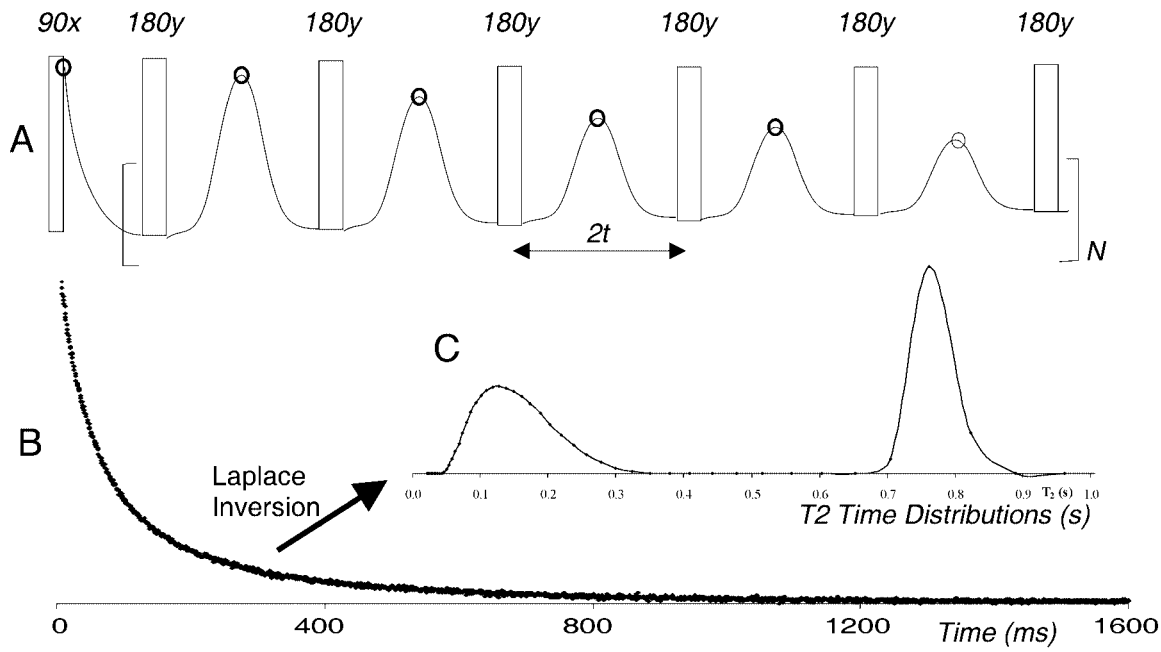


Figure 10.7 *Carr–Purcell–Meiboom–Gill (CPMG) sequence and NMR signal. (A) CPMG pulse sequence, (B) CPMG decay signal, and (C) Laplace inversion of the CPMG decay signal.*

used to extract the distribution of exponential relaxation rates from the CPMG data (Figure 10.7C).[40–42] One then obtains a plot that describes the distribution of T_2 relaxation rates in the sample. The presence of multiple distributions typically indicates the presence of two phases or components while the shape and range of the distributions yields information on physical properties of the sample. For example, in a rock sample containing only water in the pores, a broad range of T_2 values is indicative of the interaction of the water with the pore walls. A range of pore sizes leads to differences in relative diffusion distances as well as the number of collisions of the water molecules with the pore wall. Collisions with the wall are a relaxation

mechanism that reduces T_2. Thus, the Laplace transform 'spectrum' yields quantitative information on the size distribution of rock pores. This same idea can yield porosity information on any porous material containing a proton-containing liquid. This technique can be utilized to derive porosity, droplet size in emulsions, and diffusion information on many porous materials.[43–46]

10.2.3 High-resolution NMR: obtaining a spectrum with resolved chemical shift information

Another physical aspect of the magnetic environment of a nucleus that we have not discussed yet is the effect of electrons. The chemistry of hydrocarbons revolves around molecular structures held together by electronic bonds that vary greatly in electronic density due to the effects of multiple bonding, aromaticity, electronegative functionality, electron donating functionality, free pairs of electrons, etc. Any given nucleus in a molecule has an electronic environment that is varied depending on where that nucleus is in the molecule and what is attached to it. When the electrons of a sample are placed in a strong static magnetic field the magnet induces a circulation of the electrons around the applied field. This circulation of electrons in turn generates its own magnetic field that acts in opposition to the applied field. This means that every nucleus feels the effect of the main field perturbed by the induced magnetic field produced by the circulating electrons in its environment. The induced magnetic fields from the electrons are said to 'shield' the nucleus from the static applied field, reducing its overall effect. As this shielding is variable depending on the electron density around the nucleus, then the absolute magnetic field that the individual nuclei experience varies depending on the chemistry surrounding the nucleus under observation. The differences in magnetic field at the nucleus lead to differences in Larmor precession frequencies for each proton in a unique chemical environment, which leads to distributions of frequencies in the observed FID. Upon Fourier transformation the distribution of frequencies gives rise to a distribution of NMR peaks with each peak corresponding to a unique chemical environment present at an intensity proportional to its concentration. The excellent resolution that can be obtained between different chemical functionalities and structurally different environments has made NMR the premier structural chemistry tool used by chemists today.

The magnets in HR-NMR are designed to yield a highly homogeneous field at the point of analysis. As an example (keeping to process NMR frequency ranges), at a magnetic field strength of 1.47 Tesla the protons in the sample will resonate at 60 MHz ± 300 Hz. The 600 Hz range of frequencies that are observed in a complex sample is due to the electronic shielding effect of the electrons in the sample causing the Larmor frequencies to be dependent on the sample chemistry. These differences in Larmor frequency are termed chemical shifts. The chemical shift scales directly with magnetic field strength, which means that if the shifts are described in terms of frequency (Hz), the shifts vary from one spectrometer to another as the magnetic field strength changes. In order to allow consistent comparison of NMR data, the chemical shift frequencies are divided by the resonance frequency of the spectrometer (i.e. 60 MHz spectrometer frequencies are divided by 60 and 300 MHz spectrometer frequencies are divided by 300). Thus the units of this normalization are Hz/MHz, or ppm. This is why all NMR data are plotted with a scale showing the NMR peak shifts in terms of ppm. The chemical shift range observed for ^{1}H is 600 Hz or 10 ppm. Figure 10.8 shows the resolution difference between 300 MHz data (obtained on a superconducting magnet spectrometer) and that acquired at 60 MHz (on a permanent magnet process spectrometer). Though the resolution of the 60 MHz data is lower all peaks are readily observed and identified. This chemical specificity is what makes NMR such a powerful tool for many application of process analytical spectroscopy. The ability to observe well-resolved resonances for different molecules in a complex mixture allows it to be particularly useful for reaction monitoring.

The NMR experiment performed is the simplest one imaginable. The ^{1}H nuclei in the sample are irradiated with a single square pulse of RF energy (~1 W) lasting about 10–12 μs. After waiting a dead time of 10 μs, to allow the coil to physically stop oscillating from the power input it experienced, the receiver is

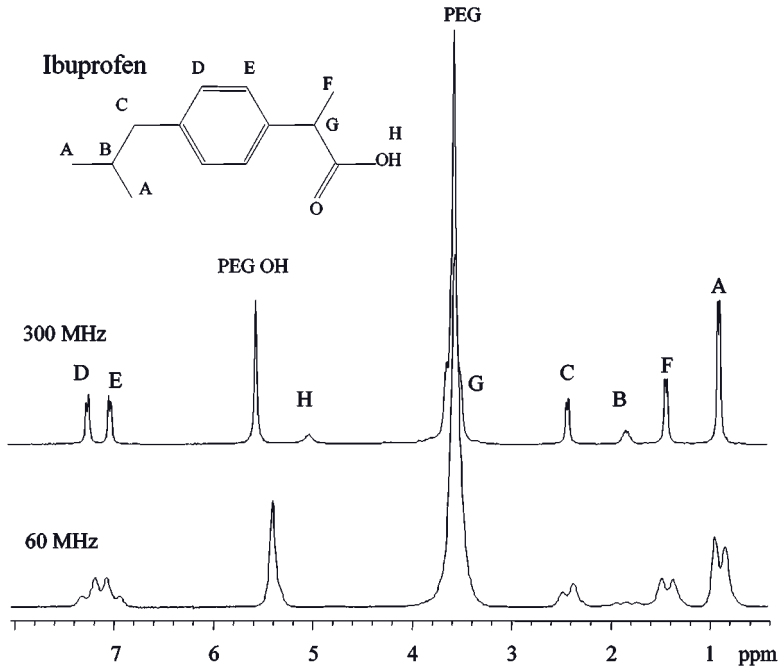


Figure 10.8 *Comparison of high-resolution FT-NMR spectra obtained on ibuprofen sample at resonance frequencies of 300 MHz (B_o = 7.06 Tesla) and 60 MHz (B_o = 1.47 Tesla).*

turned on and an FID lasting a fraction of a second is acquired and digitized. The FID is then Fourier transformed to yield a frequency space spectrum.

10.3 Process NMR Instrumentation

10.3.1 Spectrometer and magnet design

TD-NMR and HR-NMR spectrometer systems have a majority of components in common. All spectrometers consist of a magnet, magnet temperature sensors, magnet heater power supply, RF frequency synthesizer, pulse programmer, transmitter/amplifier, sample probe, duplexor, preamplifier, receiver, and ADC, all controlled by a computer. In addition to these items a HR-NMR has several other requirements which include an electromagnetic shim set, a shim power supply, and a second RF locking channel tuned to the resonance frequency of ^{7}Li. The second RF channel is identical to that of the observed ^{1}H channel. Figures 10.9 and 10.10 show the basic setup of TD-NMR and HR-NMR spectrometers, respectively.

10.3.1.1 TD-NMR Spectrometers: schematic shown in Figure 10.9

The magnets utilized in these systems are typically neodymium–iron or samarium–cobalt disks with appropriately designed pole shoes that are placed with a variable air gap between them to accommodate an NMR probe designed to allow sample tubes (or flow probes) of 5–40 mm diameter. The magnet disks are typically wrapped along their sides with double wound wire to heat magnets to maintain a stable 35 °C operating

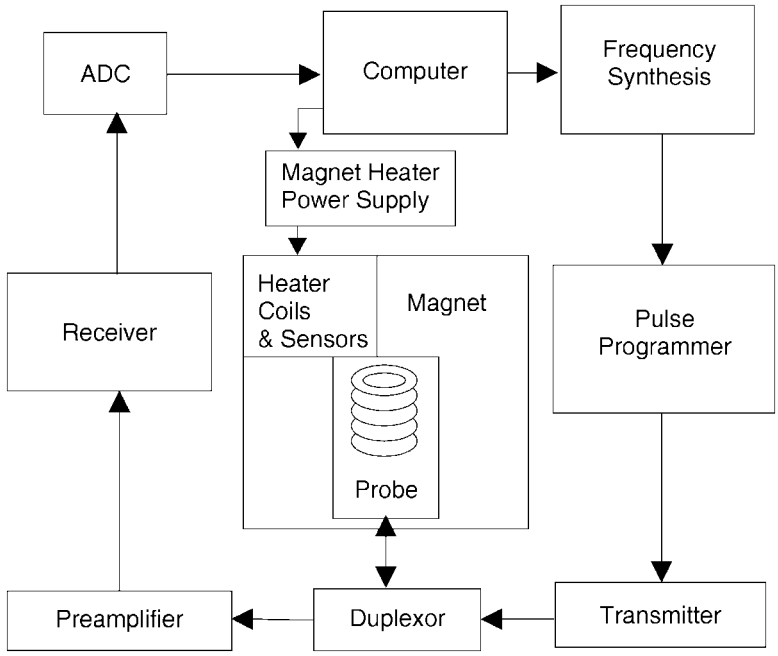


Figure 10.9 *TD-NMR spectrometer design.*

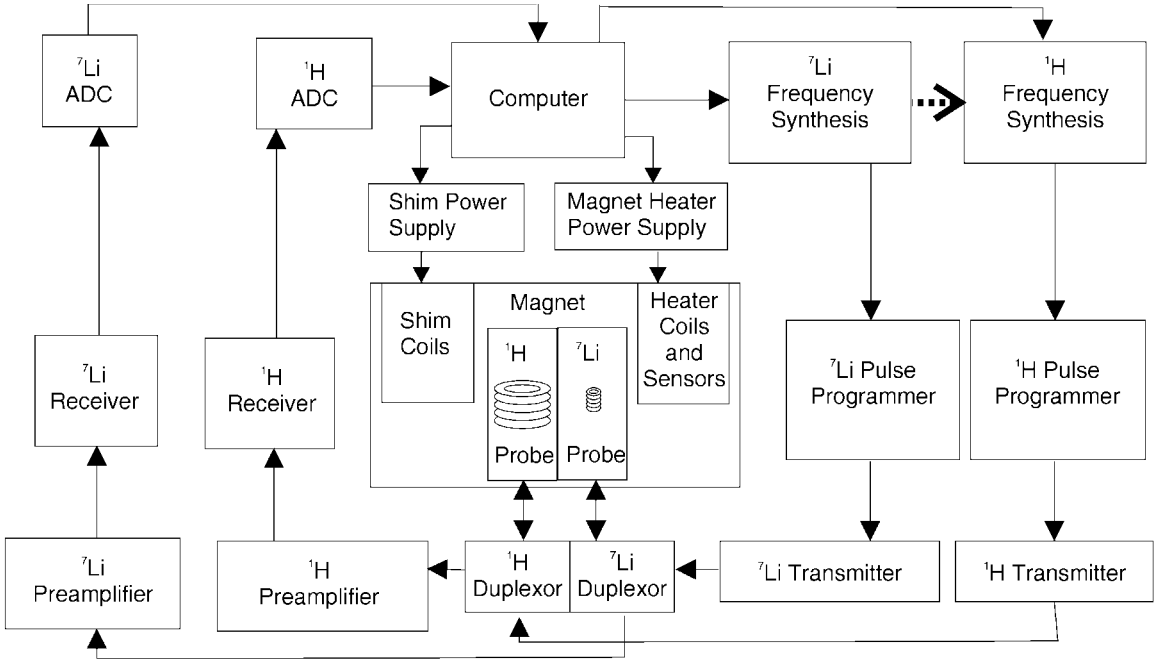


Figure 10.10 *HR-NMR spectrometer design.*

condition, or they are thermostatted to be heated a few degrees above room temperature. This heating stabilizes the flux variation in the magnets, providing a more stable magnetic field. The magnets are mechanically aligned with geometrically oriented bolts that displace the magnet pole faces with respect to each other. A trained operator can watch the length of the FID ring-down while adjusting these bolts and optimize the magnetic field to produce the longest ring-down FID, which is indicative of a more homogeneous magnetic field at the sample which will in turn give rise to narrower resonance peaks.

10.3.1.2 HR-NMR Spectrometers: schematic shown in Figure 10.10

In these discussions we will concentrate on the details of the HR-NMR spectrometer and magnet system. They represent the most complex configuration of process NMR technology. Similar process TD-NMR instrumentation is typically less complex.

Enclosure The main cabinet is divided into two compartments: one for the magnet, and one for the electronics. The magnet compartment is maintained at a temperature of 24 ± 1 °C for ultimate permanent magnet stability. The electronics compartment is on a separate air conditioner and is less rigorously temperature controlled. Figure 10.11 shows several pictures of a Qualion Process NMR spectrometer in a NEMA enclosure configured for implementation on-line.

Magnet and shims The magnet itself comprises multiple neodymium–boron–iron segments in a complex bonded construction. Integral to the construction are the application of a heater array and associated temperature sensors, as well as an array of 40 specifically oriented shim coils that are placed on the wall of the magnet at the pole gap around the bore of the magnet where the sample probe will be inserted. These shims are utilized to improve the homogeneity of the field by passing current through them to produce electromagnetic adjustments to the basic flux of the magnet array. The magnet materials themselves are maintained at a temperature of 41 °C while the enclosure housing the magnet (white box visible in the left side of the enclosure) is maintained at 37 °C. The heater control has two proportional-integral derivative (PID) loops that accept measurement input signals from thermistors mounted on the magnet itself and in the envelope. The outputs of these two PID loops control the currents to electric heater strips to a sensitivity of ±0.5 mK. The magnet and probe are pictured in Figure 10.11B. The basic resolution of the magnet is around 120 Hz (2 ppm) after mechanical shimming with eight bolts on either end of the magnet. The electrical shim brings that resolution to about 2 Hz, on a nonspinning water sample. The digital-to-analog converter boards, that are used to convert the computer's 40 digital shim current values to DC currents in the shim coils, are mounted in the enclosure of the magnet. The magnet weight is ~100 kg and the shape is approximately a cube with sides of 42 cm. Optimum field strength for this design is 1.35 Tesla which equates to a ^{1}H resonance frequency of 58 MHz.

Probe The probe is constructed of stainless steel and has Swagelok fittings top and bottom for attachment to the process sample line. The construction of the probe includes a Dewar that reduces the sample heat transfer through the walls of the probe to the magnet components. At the sampling zone the inner wall of the probe is constructed of alumina ceramic which is specialty welded to the stainless. The probe is pressure tested to 1.035×10^{4} Pa (103.4 bar, 1500 psi). The duplexor and preamplifier are built into the base of the probe.

RF generation, transmission and spectrometer locking The NMR probe is in fact a two-channel probe that is capable of obtaining an ^{1}H NMR signal while locking the spectrometer through a separate channel operating at the resonance frequency of ^{7}Li. At the sampling point of the probe there are two coils – the ^{1}H

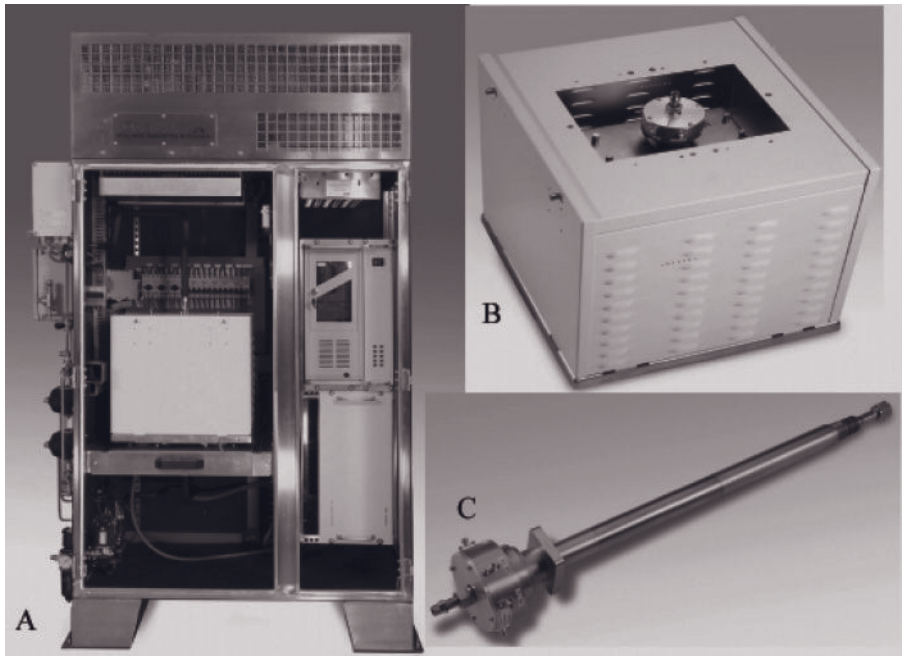


Figure 10.11 *Qualion Process NMR Analyzer operating at 60 MHz. (A) Spectrometer housed in area certified enclosure, (B) NMR magnet system and probe, (C) NMR probe.*

tuned coil is wrapped around the process sample line (around the outside of the alumina ceramic section) which has an interior dimension of 7 mm. A second smaller coil is mounted in the Dewared vacuum area of the probe directly adjacent to the ^{1}H coil. The smaller coil is wrapped around a sealed capillary of LiCl containing a paramagnetic relaxation agent. ^{7}Li is a high receptivity, quadrupolar nucleus that yields a strong signal with a very short relaxation delay requirement between NMR pulses. The spectrometer can work in several modes to 'lock' the NMR spectrometer in order to produce high-resolution spectra that are immune to the effects of magnetic field drift during the NMR acquisition.

The requirement of electronic locking capability on a permanent magnet based NMR spectrometer system arises from the fact that magnetic fields are always drifting due to the effects of temperature cycles in the magnetic material. These drifts when converted to NMR resonance frequencies are on the order of Hz/min. Thus, if one obtains a signal-averaged NMR spectrum over the course of several minutes the resolution of the spectrum can be greatly decreased as the drift causes the peaks to broaden. To avoid this situation the Qualion spectrometer constantly observes the resonance frequency of the ^{7}Li signal through the lock channel. The drift of the magnet can be observed and quantified and when it is time to pulse on the sample through the ^{1}H channel, the transmitter frequency is adjusted to keep the next FID at the same resonance offset as the previous one. Other approaches are also available in the processing steps to remove the effects of magnet drift. We will discuss them further in the data processing discussion.

10.3.2 Sampling and experimental design

TD-NMR sampling is decidedly more difficult than HR-NMR because on-line applications of low-resolution NMR have tended to focus on solid materials such as tablets, polymers, toothpaste, cellulose, foodstuffs,

and packaged goods. In many ways the sampling of these types of materials is extremely difficult when considering that the sample must be introduced into a restricted probe for analysis in a consistent and repeatable manner. For many on-line projects the sampling system is oftentimes more expensive than the NMR. For this reason, the TD-NMR analyses that have moved on-line have focused predominantly on polymer samples that can be melted and introduced to the NMR probe. Most applications of TD-NMR must be considered as at-line applications because the sample requirements are too problematic and have to be addressed either by human or robot sample preparation.

In the case of HR-NMR the main stipulation is that all samples introduced to the analyzer must be in an entirely liquid form in order for the protons in the entire molecular distribution to be observed. Solids will not provide observable ^{1}H signals and their presence will therefore reduce the efficacy of the on-line predictions or the model development. A sampling tap from the process lines must deliver separate streams (up to five) to the sample system at a flow rate of 260–340 liters per hour. The sample system must be designed to provide each sample stream to the NMR probe with a temperature variation of less than 3 °C. Each sample system is custom designed to meet the above criteria by working in close collaboration with process engineers on the temperature, flows, pressures, viscosities, and solids content found in the various process lines.

Excessive water in samples should be removed if possible, but there are processing methodologies to remove the effects of water contamination on the chemometric models used for predictions. Water appears in a specific area of the spectrum and can be digitally removed from the spectrum during the processing stages.

10.4 Postprocessing Methodologies for NMR Data

As described in Section 10.2, the final output from the NMR spectrometer to the computer is an FID. Typically 2048–4096 digital points are accumulated in the FID and the next step is to improve the potential resolution of the FID by 'zero-filling' the FID to 16384 digital points by adding zeroes to the end of the FID. Upon Fourier transformation (FT) the resultant spectrum contains 16 384 points describing a spectral width of 2000–4000 Hz depending on the settings in the ADC.

Following the FT the NMR data are phased so that the baseline is as flat as possible and all resonance peaks are positive. In an NMR experiment the first data point of the FID can be at any combination of phases depending on the timing of the various RF pulses and delays in the NMR experiment table. This means that upon applying the FT, the spectrum will typically be out of phase with the peaks dispersed about the baseline or even completely negative. Automated methods of phasing the spectrum have been developed, and for different applications parameters can be set to allow for extremely reproducible phasing results. This is of vital importance in the chemometric methodologies as it is important that the same data can be automatically processed to produce identical processing results. Otherwise small phase changes can be interpreted as being small chemical differences between samples and these can then translate into erroneously assigned chemical and physical property variations that are not truly there.

The process NMR data are not stored in the traditional manner utilized for laboratory NMR data. On laboratory instrumentation a single FID is stored as the final result of the experiment. The FID represents a signal averaging of perhaps 16–64 FIDs produced by 16–64 pulses of RF. On a process NMR instrument the final data that are saved to disk are of the fully processed spectrum. There are a number of steps that must be taken in the processing of the FID to the final spectrum. In order to remove small magnetic field drift from the NMR data that cannot be fully addressed by the ^{7}Li lock circuitry, each FID is Fourier transformed, autophased, and then autoreferenced. The autoreferencing is a requirement since the process samples being analyzed do not contain the usual NMR deuterated solvents or the TMS reference to establish the

0 ppm reference of the spectrum (a procedure that is standard in lab-based superconducting NMR). The operator involved with developing process NMR applications must establish the chemical shift of several intense peaks in the samples that are being analyzed. The NMR software can then be instructed to peak-pick the relevant resonances and assign them to the particular chemical shift defined by the operator. Then each individual one pulse spectrum is coadded after having been aligned by the referencing procedure. This means that the NMR spectra are averaged in the frequency domain (spectral data) rather than the time domain (FID data). The fact that pure samples are being analyzed allows for adequate signal-to-noise ratio to be obtained on a single pulse. This technique does not work as well when one is using a process NMR instrument in a laboratory environment where the samples are always diluted significantly in deuterated solvents. The basic steps of NMR data processing are shown schematically in Figure 10.12.

The databases of NMR data that are the basis of chemometric model development for on-line process control can be developed via several different processing approaches.

- The actual spectral data can be directly modeled without digital resolution reduction or spectral normalization.
- The data can be 'binned' which involves integrating the NMR spectral region from 12 to −2 ppm in increments of 0.02, 0.05, or 0.1 ppm. This effectively reduces the digital resolution of the spectral data but maintains the quantitative nature of the NMR data. These arrays of integral values become the NMR spectrum that is imported into the chemometric modeling software. If required, these 'integral spectra' can be normalized such that the sum of the integrals is equal to a fixed value. Figure 10.13 shows a diesel fuel data set binned with a 0.05-ppm integration resolution.
- The NMR spectra can be integrated to reflect the chemistry functionality observed in the ^{1}H NMR spectrum. This leads to the development of a 'carbon chemistry spectrum' that can be utilized as the *X* matrix data of a chemometric model. NMR calculated chemistry values can also be appended to the chemistry integral data so that various ratios and calculated chemistry parameters can be brought into the regression modeling. This approach allows specific chemistry types and parameters to be directly observed as correlators to physical properties being investigated. Figure 10.14 shows the final data set

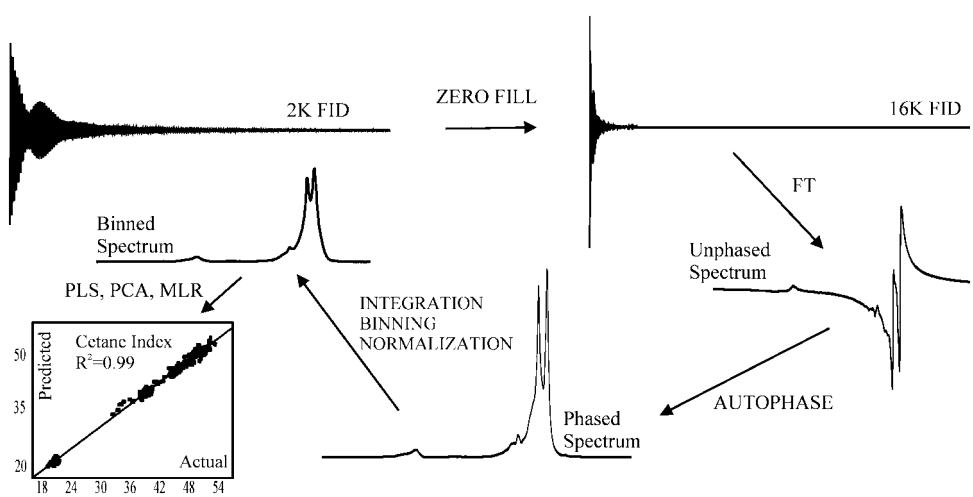


Figure 10.12 *Processing scheme utilized in HR-NMR.*

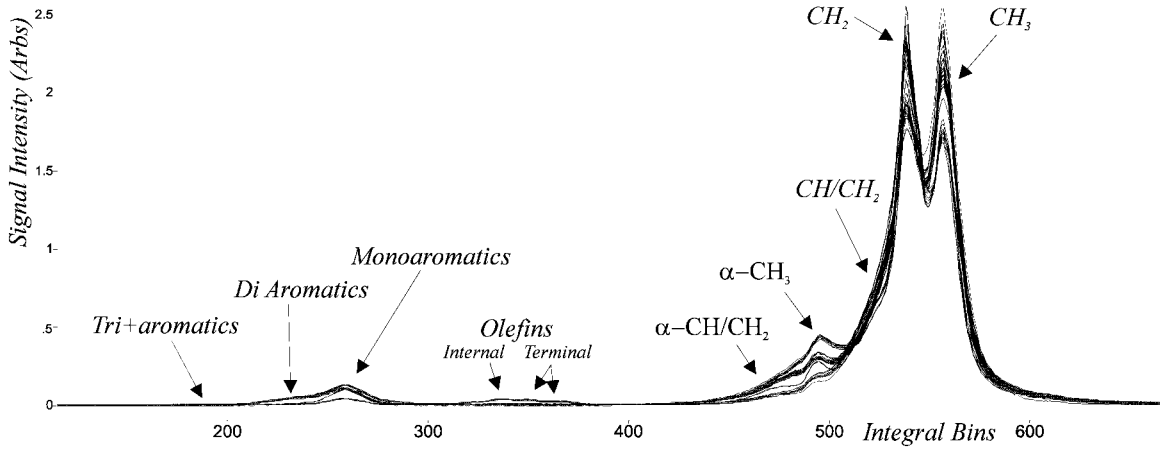


Figure 10.13 *Diesel NMR spectra processed with 700 integral points obtained with an increment of 0.02 ppm across a chemical shift range of 12 to −2 ppm. Basic chemistry functionality is also indicated.*

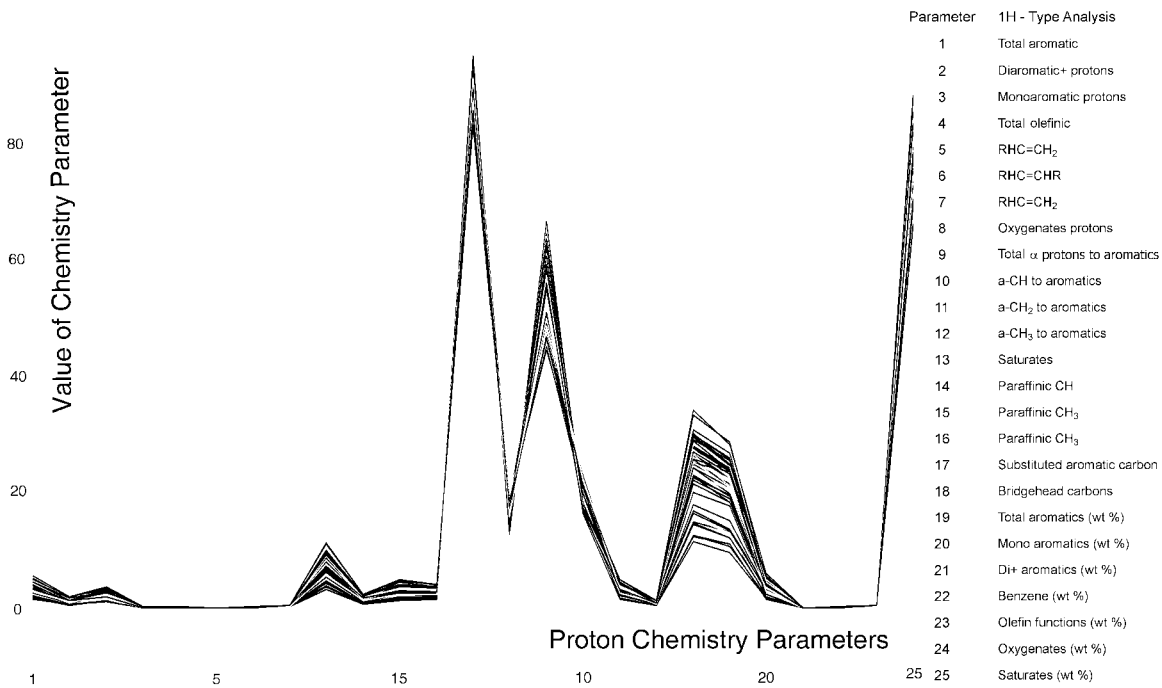


Figure 10.14 *Process NMR data from a series of vacuum gas oils represented in terms of proton chemistry.*

on a series of diesels where spectra were created from the various integrals describing the proton chemistry in the sample: i.e. mono-, di-, and tri-aromatics, total aromatic-H, olefinic-H, alpha-H, alpha-CH_2, alpha-CH_3, iso-paraffinic CH/CH_2, linear CH_2, CH_3, etc.

10.5 Advantages and Limitations of NMR as a Process Analytical Technology

10.5.1 Advantages

NMR spectroscopy is a truly quantitative technique when the experiments are performed with appropriate experimental delays (to enable full relaxation to thermal equilibrium after each RF pulse) and processed with robust procedures. The spectral response is identical for all proton chemistries in the sample and the chemical functionality observed is orthogonal in nature and falls into distinct regions of the NMR spectrum. With appropriate sampling and experimental conditions this means that the NMR response of liquids and gases containing protons is also linear and quantitative. In Figure 10.15 we show the spectrum of a gasoline obtained on a process NMR instrument. The NMR spectrum can be broadly described by the following breakdown of chemical shift regions with sample chemistry when describing petroleum hydrocarbons:

- 0–2.0 ppm: the aliphatic region of the spectrum clearly shows distinct peaks for CH_3, CH_2, and CH protons in this region.
- 2.0–3.4 ppm: the alpha-proton region shows distinct peaks for CH_3 and CH_2/CH directly adjacent to an aromatic ring. The CH_2 bridging two phenyl groups is also observed in this region.
- 3.0–4.2 ppm: alcohol and ether protons associated in gasoline with ethanol or ether additives are observed in this region.
- 4.5–6.5 ppm: olefin (alkene) proton region – specific olefin bond types can readily be quantified. Conjugated diolefin signal intensity can also be observed in the 6–6.5 ppm region.
- 6.5–8.5 ppm: aromatic proton region, can also be subdivided into mono-, di-, and tri- plus aromatic regions.

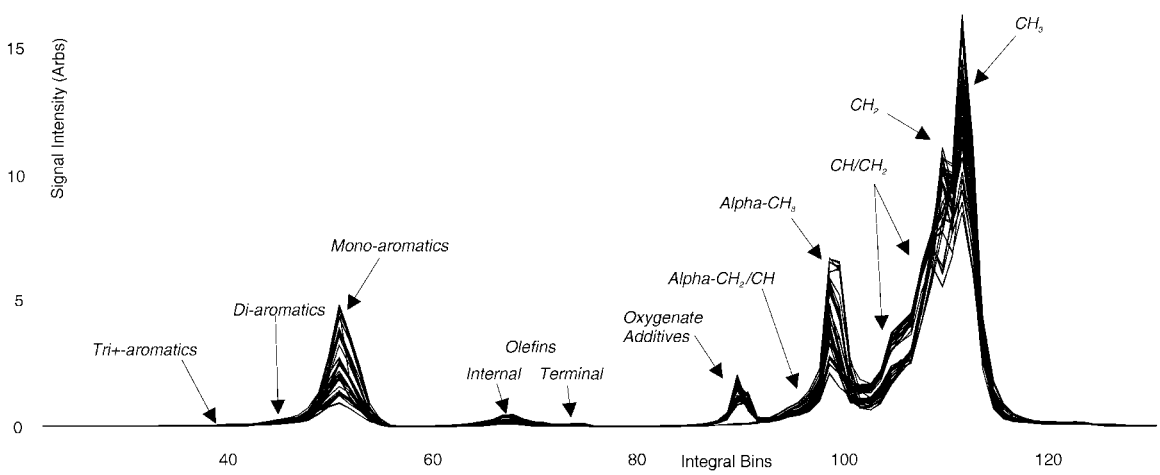


Figure 10.15 *Gasoline NMR spectra processed with 140 integral points obtained with an increment of 0.1 ppm across a chemical shift range of 12 to −2 ppm. Basic chemistry functionality is also indicated.*

One of the advantages of NMR is that water, or solvents, fall in specific regions of the spectrum and can be digitally removed from the spectrum by simple processing procedures, and hence do not interfere with analysis.

The entire sample volume (10×7 mm (length $\times$ diameter)) is observed during the experiment and optics' fouling is not an issue. The NMR result is not affected by sample color and receives an equal response from black, brown, yellow, or colorless samples.

The linear response of sample chemistry across all hydrocarbons allows spectrum-property regressions to be performed on a wide range of samples in a given chemometric model. The modeling approach can literally include individual parameter models that span a single product type (e.g. all diesels) or span several product types (e.g. naphtha, kerosene, diesel, gas oils). Thus, the modeling is far more robust and does not have to be continually updated and thoroughly localized as is often the case for NIR or mid-IR spectroscopy.

10.5.2 Limitations

Samples must be physically brought to the NMR magnet and probe via a sample line and sampling system. This limits the number of streams that a single NMR unit can effectively monitor with a useful duty cycle. Heating of samples entering the NMR analyzer is required for several reasons:

- Large temperature variation between streams will lead to magnetic field fluctuations that will reduce the repeatability of the NMR experiment.
- Solid hydrocarbon materials are not observable in process NMR instruments. In order to be observed the molecules under analysis must be entirely in the liquid state and must be above a nominal viscosity threshold. In heavy or waxy petroleum streams the samples must be heated to approximately 80 °C to lower viscosity and ensure melting and solubility of waxy components.

NMR spectra can be greatly affected by the presence of paramagnetics such as iron in the sample stream. Care must be taken to appropriately filter the sample streams to prevent particulates entering the magnet zone.

NMR observes the chemistry of only the proton nucleus (though it can observe many other nuclei independently). This means that hetero and metallic chemistry cannot be observed directly. Thus, sulfur, nitrogen, oxygen, and metals cannot be directly analyzed by NMR, though secondary correlations can be obtained from the proton chemistry of the sample. In combination with electron spin resonance (ESR) analyzers that can operate in the fringe fields of the NMR magnet the presence of paramagnetic metals and free radicals can be quantified.

NMR spectroscopy is relatively insensitive compared to optical spectroscopies such as Fourier transform infrared, requiring acquisitions of several minutes to obtain high signal-to-noise ratio spectra adequate for process analytical applications. The linear response and low sensitivity of NMR also prevents its use for observing very low level contaminants (<1000 ppm) in complex mixtures.

10.6 On-line and At-line Applications

As stated previously, the focus of this review is the truly on-line application of NMR in an industrial manufacturing facility. There are many papers in the literature describing chemometric applications of laboratory NMR data to produce chemometric relationships between NMR data and physical/chemical properties.[47–67] These applications will not be discussed in detail here but are noteworthy as the approaches described can

be implemented on-line on HR-NMR or TD-NMR instrumentation. The discussion here will also include the work of several groups who have described superconducting NMR system reaction monitoring in the laboratory by directly coupling reaction vessels to laboratory NMR systems.[7–13]

10.6.1 Time–domain NMR

Progression, Inc. began deploying TD-NMR instrumentation, under the trade name Magneflow IMR, into petrochemical plants in the late 1980s when they operated under the name Auburn International. They currently have over 100 systems in the field monitoring polyethylene (melt index, density, and crystallinity) and polypropylene pellets (for xylene solubles (XS), ethylene content (C_2), Charpy modulus, flexular modulus).[68–70] The NMR analysis involves grabbing ~50 ml of sample and pneumatically delivering it to the NMR analyzer. There it is heat treated and the FID is analyzed with a three-component fitting routine to extract a fast Gaussian (isotactic crystalline), slow Gaussian (intercrystalline), and an exponential (amorphous) component from the signal. The relative amounts of these three components for each sample are regressed against known laboratory test results and models derived for on-line prediction. The duty cycle on the analyzer is around 5–10 min.

A series of TD-NMR applications to determine total styrene content in styrene-butadiene polymers was described by researchers at Phillips Petroleum.[71–73] PLS analysis of the NMR FID signals from a series of styrene-butadiene rubbers yielded robust, transferable models that allowed styrene content to be calculated based on relaxation time differences observed in the FID. The FID was comprised of an initial rapid decay (the rigid styrene blocks have relatively short T_2 relaxation times), and a much longer decaying signal from the butadiene blocks (much higher in molecular mobility leading to longer T_2 relaxation times).

By far the most widespread use of NMR in an on-line production environment is the utilization of downhole exploration tools by petroleum service companies such as Schlumberger, Halliburton, and Baker Hughes. Articles on these unilateral NMR systems are found in the patent databases,[74–76] academic literature,[77–81] and on-line resources provided by the exploration companies. The references provided here are just a few examples in a very prolific field. The technique is applied in high-temperature and pressure situations and currently is used down to a depth of about 10 km (6 miles) to produce a picture of water/oil content in the adjacent rock formations as well as to derive permeability, diffusivity, and hydrocarbon chemistry information. Mobile unilateral NMR systems such as the NMR-MOUSE are also being developed in order to take benchtop NMR systems into the field to perform analysis of geological core samples at the drill site.[82] NMR analyzers are also being developed to determine the bitumen content and properties in tar sand production and processing.[83–84]

The same techniques applied in this field are also directly applicable in food manufacturing process control, though there do not appear to be any NMR systems currently on-line. In previous studies it was determined that TD-NMR would be an excellent tool for monitoring fat, water, and protein content in the manufacturing of various dairy products such as butter, cream cheese, and cheese.[85–91] Companies such as Unilever appear to be developing applications in this area but to date no reporting of truly on-line analysis has appeared in print. The techniques appear to be relegated to at-line benchtop situations. The process NMR of liquid-like materials such as dairy products, fats, vegetable oils, and beverages is fairly straightforward from a sampling standpoint. Sample can be readily transported to the NMR instrument and can be observed under slow flow conditions or under static conditions by incorporation of valve closing and opening sequences into the NMR experimental pulse sequence.

BOC and Bruker have been working together on a 'no contact check weighing' (NCCW) system based on NMR for measurements on the packaging line of a pharmaceutical vial loader. This analyzer utilizes the absolute FID intensity to determine the proton density of liquid and powder samples in vials passing through the magnet. The NMR system operates at 8–10 MHz and is sizeable with vial dimensions.[92]

One of the areas where traditional NMR methodology has problems is in the on-line inspection of packaged foods or beverages on a conveyor line. The fact that industrial conveyors have an operational speed of several meters per second means that the NMR analysis time is reduced to milliseconds, and the only experiment possible is literally a single pulse opportunity. The NMR experiment would typically occur over a distance of about 2–10 m, which is problematic with regard to sensor development. An alternate method for on-line analysis utilizing NMR technology in an untraditional way is to implement NMR signals that are generated by the motional relativity of the samples passing through a magnetic zone that is irradiated continuously with RF fields.[93] The RF excitation zones are irradiated at different power levels that would effectively cause 90° and 180° spin flips to be applied to the nuclear spins traveling through the RF zone. With a combination of pulsed field gradients and evolution times it is possible to acquire an FID, an echo, and T_1/T_2 information. Also with the application of a constant field gradient along the direction of motion of the sample it is possible to obtain NMR microimages of the sample. Successful application of an NMR analyzer of this type would mean that real-time analysis of food and horticultural samples could be obtained. This would allow the rapid detection of foreign bodies in the product (observing objects such as sticks, plastics, insects, etc., that are not detectable in current X-ray assays), as well as sample damage (freeze damage of fruits and vegetables, bruising of fruits, storage spoilage, malformation, phase separations). Other applications might include monitoring sodium content in soups and canned goods (T_1 and T_2 measurements of brine can be related to sodium content), moisture and fat content, as well as undesired bone fragments present in meats,[94] and moisture content of baked goods during the baking process.[95] This type of NMR analyzer is currently in the development stages and the potential market for such an analyzer is a strong incentive to drive development to completion.

10.6.2 High-resolution NMR: chemometric applications

The first application of high-resolution (HR) NMR in a process environment was in 1995 when an Elbit-ATI 53 MHz NMR analyzer, developed by Texaco R&D, was placed on a cogeneration unit's fuel gas stream to measure BTU content of fuel gas at their Los Angeles refinery.[35] The analyzer measures BTU, specific gravity, methane content, hydrogen content, olefinic H, and hydrocarbon content; the NMR compares favorably with the gas chromatograph (GC). Unlike the GC, process NMR does not require component speciation to determine BTU and specific gravity. Rather the BTU is calculated directly from the C-H bond types that are observed in the ^{1}H NMR spectrum. A typical ^{1}H NMR spectrum of fuel gas is shown in Figure 10.16. Typical stream conditions are 345–380 Pa (3.45–3.79 bar, 50–55 psi) and 20–25 °C resulting in a total concentration of less than 6 μmol of gas in the NMR probe. To improve sensitivity, the NMR averages 32 scans per analysis. However, before each scan, a new sample of gas is purged into the probe. Each analysis is an average of 32 different samples and results in a fuel gas analysis being performed every 30 min. Further, as required by the refinery's environmental permit, a pure methane standard spectrum is obtained daily, and the measured methane value is compared to the predicted methane value from the calibration curve at the pressures and temperatures obtained during the daily calibration run. Predicted versus measured methane values must be within ±2% for the NMR to be within daily calibration.

In 1997, based on the operational success of the fuel gas analyzer, a second system was developed by Texaco R&D to measure the acid strength of sulfuric acid in an acid alkylation unit at the same Los Angeles refinery.[36,96,97] In the alkylation process butane (C_4) gases are taken from the refining process and brought into contact with concentrated sulfuric acid, and are acid catalyzed to octane (C_8). A side reaction of the process leads to the buildup of C4-based polymers known as acid soluble oils (ASO). The justification for installation of the NMR was based on high acid costs at that particular plant. At the time the acid strength set point for the alkylation process was 90.5 wt% H_2SO_4. The ability to control acid strength in the third of three Stratco contactors (in series) to 88.0–89.0 wt% H_2SO_4 translated into a $1 million dollar per year

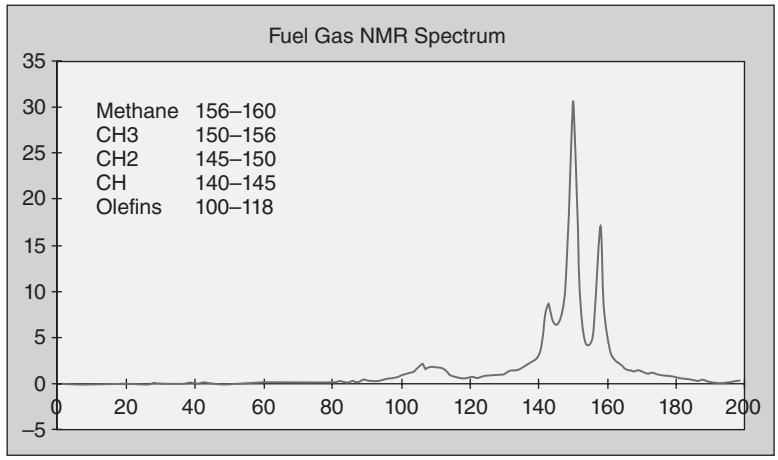


Figure 10.16 *^{1}H NMR spectrum of fuel gas at Texaco Los Angeles refinery.*

savings in acid costs. In addition to the reduction of acid costs it was hoped that close control of the unit by NMR would lead to improved operating efficiencies and improved product quality. These operating efficiencies included: (i) the ability to expand the utilization of hydrocarbon feed types and distributions (C_3–C_5); (ii) the ability to regulate acid feed rates with changing feed compositions; and (iii) the ability to operate the unit at or near design limits. Safety improvements were also realized by reducing the need for manually grabbing acid samples for titration. Two approaches were taken – analysis of the emulsion reaction zone of the Stratco unit, and analysis of the settled spent acid before circulation back into the unit. In the emulsion zone the NMR can monitor the C_3–C_5 hydrocarbon/acid ratio, the ASO content, and the acid strength. In the settler application, the ASO content is determined along with the acid strength and shown to be inversely proportional to the acid strength. Figure 10.17 shows the data from 14 spent acid samples. The chemical shift position of the sulfuric acid peak is related to the acid strength of the sulfuric acid. By monitoring the position of the acid resonance the sulfuric acid strength can be determined from a simple calibration curve (Figure 10.18).

In 1997, the Foxboro Company (later known as Invensys), purchased Elbit-ATI NMR and began to market the analyzer as the Foxboro NMR as well as the Invensys MRA. In 1998 BP placed an NMR analyzer on a reformer at their Coryton refinery in the United Kingdom.[98] The analyzer was placed on-line to control the reformer based on research octane number (RON) and motor octane number (MON) as well as to determine the total aromatics and benzene content by ASTM GC methods. Figure 10.19 shows a series of NMR data compared to the RON of each sample; it can plainly be seen that RON increases as aromatics increase. Partial least-squares (PLS) regression modeling of reformate NMR data sets yields extremely robust octane models that require minimal updating. Figure 10.20 shows the performance of a RON model over the course of a 3-week period. The interesting thing in this comparison is that an on-line NIR was being utilized side by side with the NMR. Both were operating on models that encompassed a range of RON values from 98–104. For a considerable period the reformer operated in a high severity condition and octane numbers by NMR were found to be 106, whereas the NIR was predicting a full octane lower. CFR engine tests confirmed that the NMR prediction was correct and proved that the NMR model predictions were linear beyond the range of the model. A typical RON model for reformate samples is shown in Figure 10.21.

Performance of PLS regression models to predict benzene and total aromatics by GC methods were equally robust. Model maintenance was minimal and the spectrometer operated within validation limits for periods of over a year.

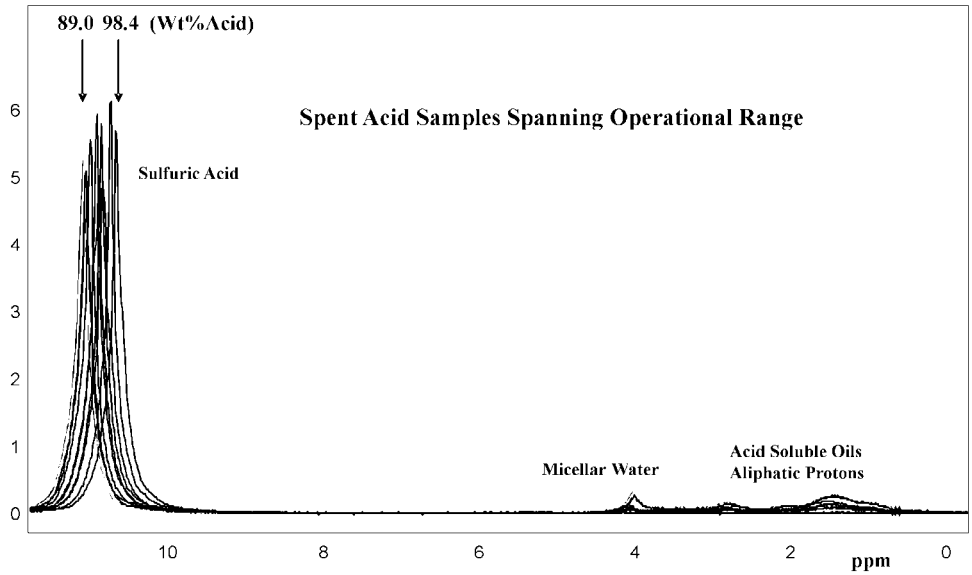


Figure 10.17 *^{1}H NMR spectra of spent acid samples spanning operational range of sulfuric acid concentrations. Shift of sulfuric acid peak with change in acid strength is indicated. Acid soluble oils and micellar water are readily quantified.*

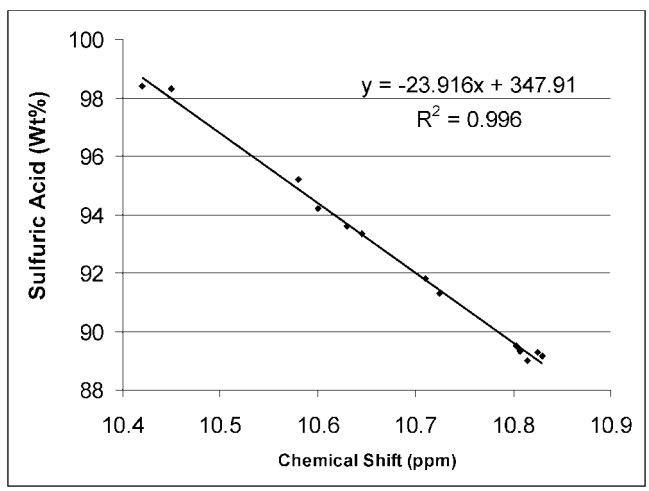


Figure 10.18 *Calibration describing the relationship between chemical shift and sulfuric acid concentration in spent acid from a refinery alkylation process.*

In 2000 two major petrochemical companies installed process NMR systems on the feed streams to steam crackers in their production complexes where they provided feed forward stream characterization to the Spyro reactor models used to optimize the production processes.[99–101] The analysis was comprised of PLS prediction of n-paraffins, *iso*-paraffins, naphthenes, and aromatics calibrated to GC analysis (PINA) with speciation of C_4-C_{10} for each of the hydrocarbon groups. Figure 10.22 shows typical NMR spectral variability for naphtha streams. Table 10.2 shows the PLS calibration performance obtained with cross validation for

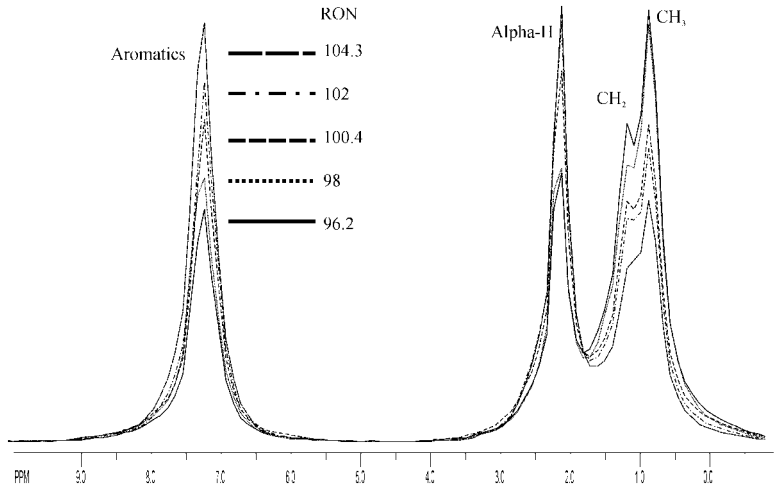


Figure 10.19 *Overlaid ^1H process NMR spectra of reformate samples of varying research octane number. Aromatic proton signals increase with increasing octane number.*

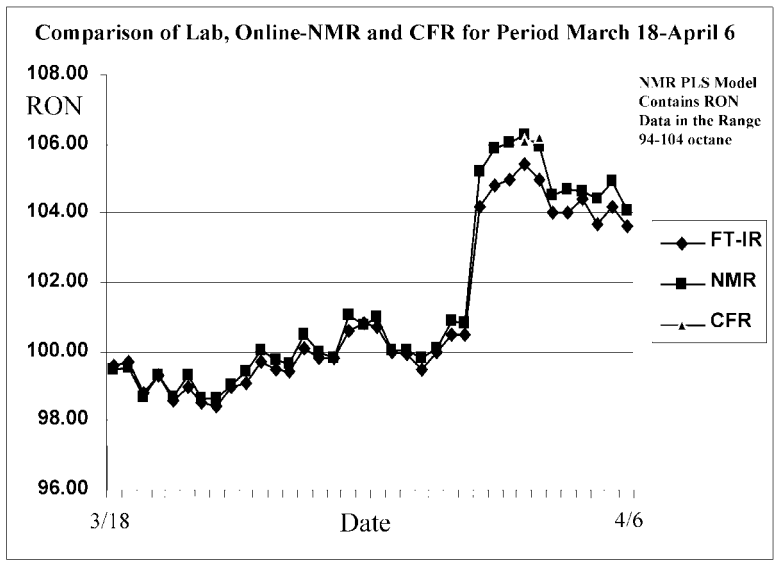


Figure 10.20 *Comparison of online NMR and FT-NIR RON predictions on reformate product. CFR engines confirmed correct NMR prediction 2 octane numbers above the upper range limit of the model.*

the PINA analyses predicted for the optimization of the steam cracker. An example of 4 days of on-line predictions for the PINA total hydrocarbon type groups is shown in Figure 10.23. A typical PLS model for an aromatic component, given in Figure 10.24 shows the performance for toluene prediction from NMR data.

During the period of 2000–2008 about 40 high-resolution process NMR systems were placed on-line in refineries around the world. In Australia Caltex Brisbane refinery installed a gasoline blending NMR system

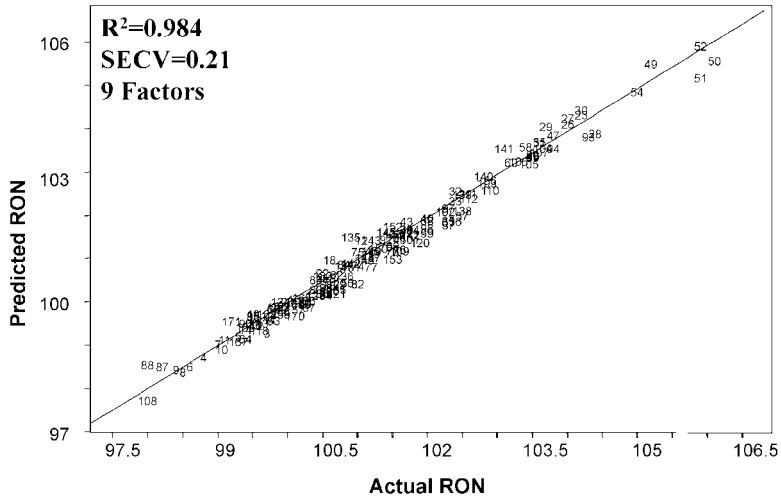


Figure 10.21 *RON PLS model for reformate streams.*

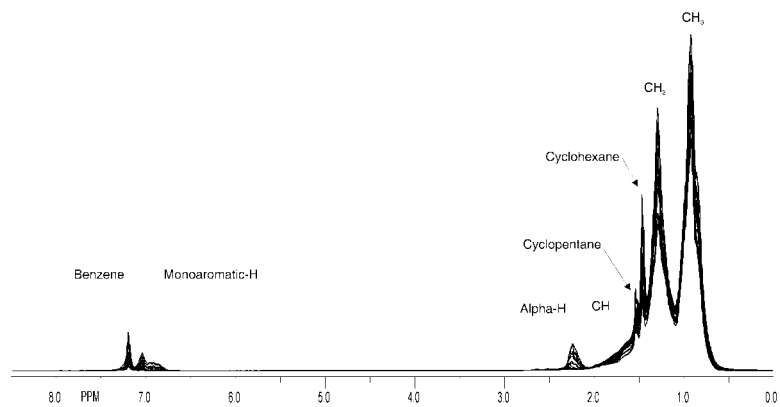


Figure 10.22 *^{1}H process NMR spectra and related chemistry of naphtha streams feeding a steam cracker.*

and now certifies and releases gasoline based on NMR analysis.[102] The NMR gasoline analysis comprises of chemometric prediction of RON, MON, benzene, ASTM-D86 distillation points (T10, T50, T90, FBP), aromatics, and olefins. PDVSA-Isla refinery applied NMR to the optimization of a crude unit observing chemical and physical properties of the incoming crude oil as well as heavy naphtha and kerosene products.[103] The crude oil true boiling point (TBP) was predicted by NMR and revealed a large difference in the calculated TBP (from paper assay information on the blended crudes); the actual TBP observed by NMR contributed to the reconciliation process of the optimization project.

Economic analysis performed for refineries in certain markets have calculated that the benefit of being able to increase kerosene and jet fuel production yield was an improvement of 3–6 cents per barrel over previous operational conditions. On an 180 000 barrel per day crude unit this equates to a benefit of \$2 000 000–4 000 000 per year. Several other refiners are utilizing NMR analyzers on the feed and products of crude units for control and optimization,[104,105] AGIP has an NMR analyzer for monitoring the feed,

Table 10.2 *Performance of PLS cross-validation modeling of NMR correlated to GC-PINA*

	R^2	SECV[a] (wt%)	Minimum (wt%)	Maximum (wt%)
Normal C4	0.842	0.71	0	7
Normal C5	0.941	1.58	4	31
Normal C6	0.925	0.81	5	18
Normal C7	0.953	0.51	0	8
Normal C8	0.968	0.32	0	6
Normal C9	0.952	0.23	0	5
Normal C10	0.893	0.16	0	2
Normal C11	0.864	0.03	0	0.3
Total normal paraffins	0.962	0.98	25	60
Iso-C4	0.831	0.32	0	3
Iso-C5	0.957	1.25	1	22
Iso-C6	0.922	1.32	5	32
Iso-C7	0.898	0.89	1	10
Iso-C8	0.962	0.39	0	7
Iso-C9	0.937	0.42	0	7
Iso-C10	0.962	0.22	0	5
Iso-C11	0.876	0.05	0	0.5
Total iso-paraffins	0.956	1.01	20	45
Cyclopentane	0.991	0.08	0	4
Methyl-cyclopentane	0.955	0.37	0.5	10
Cyclohexane	0.994	0.25	0.2	15
Methyl-cyclohexane	0.953	0.49	0	10
Naphthene – C7	0.954	0.35	0	6
Naphthene – C8	0.959	0.51	0	9
Naphthene – C9	0.921	0.44	0	7
Naphthene – C10	0.916	0.05	0	1
Total naphthenes	0.993	0.53	5	37
Benzene	0.998	0.10	0.3	12
Toluene	0.973	0.15	0	5
Ethyl-benzene	0.912	0.06	0	0.6
Xylenes	0.948	0.18	0	3.5
Aromatic-C9	0.949	0.20	0	3.5
Aromatic-C10	0.856	0.20	0	1.6
Total aromatics	0.987	0.35	2.5	17

[a] Standard error of cross validation.

intermediate, and final products of a lubricant oil manufacturing plant in Livorno, Italy.[106] The NMR in this application is predicting ^{13}C NMR-derived chemistry parameters (fraction aromatic carbon (*F*a), fraction naphthenic carbon (*F*n), fraction paraffinic carbon (*F*p)), density, refractive index, simulated distillation points, and viscosity index). Finally a number of refiners have placed NMR analyzers on fluid catalytic cracking units (FCCU) and are utilizing the NMR to obtain real-time analysis of feed and cracked products such as light cycle oil (LCO). Typical parameters analyzed are simulated distillation, ^{13}C NMR derived carbon type parameters (*F*a, *F*p and *F*n), micro-Conradson residue (MCRT), refractive index, density, and sulfur content.

Clean fuels regulations in both the European and American markets have had a substantial impact on a refiner's ability to maximize product draws at the refinery front end. Extremely low sulfur requirements for

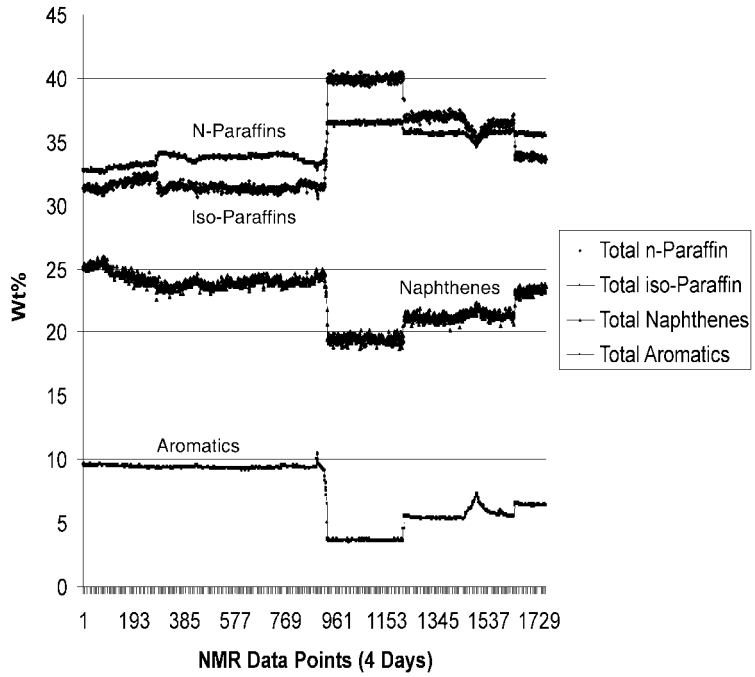


Figure 10.23 *Four day performance of ^{1}H NMR PLS model predictions for GC-PINA hydrocarbon group totals.*

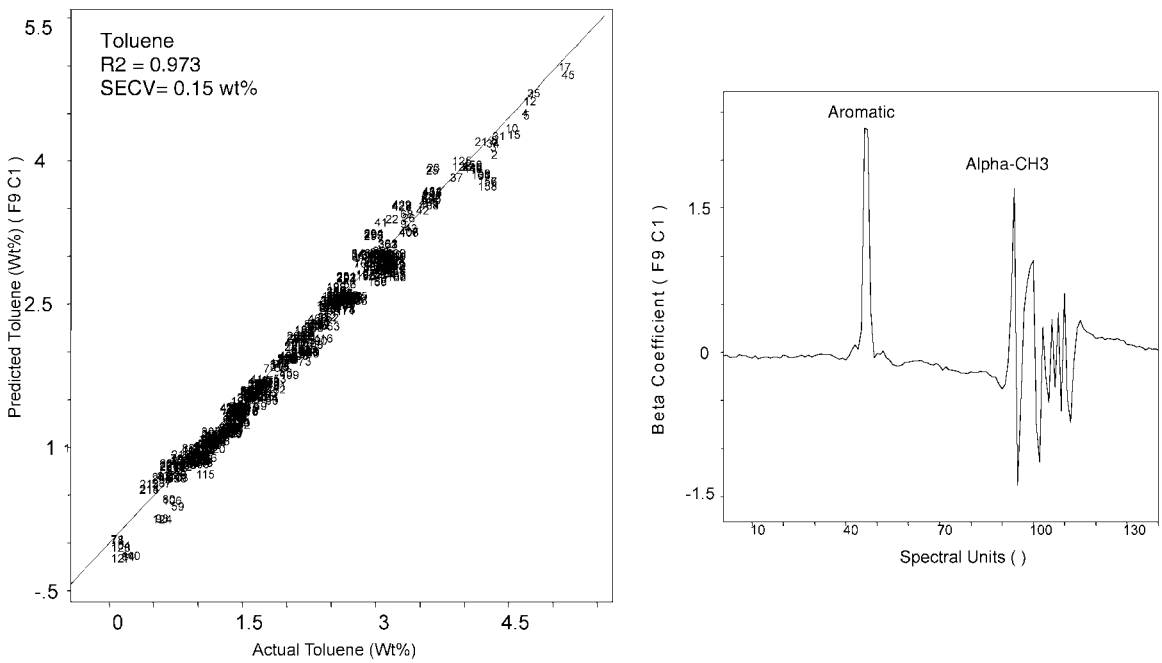


Figure 10.24 *Typical PLS model performance in GC-PINA prediction – ^{1}H NMR spectrum correlated to toluene content (wt%).*

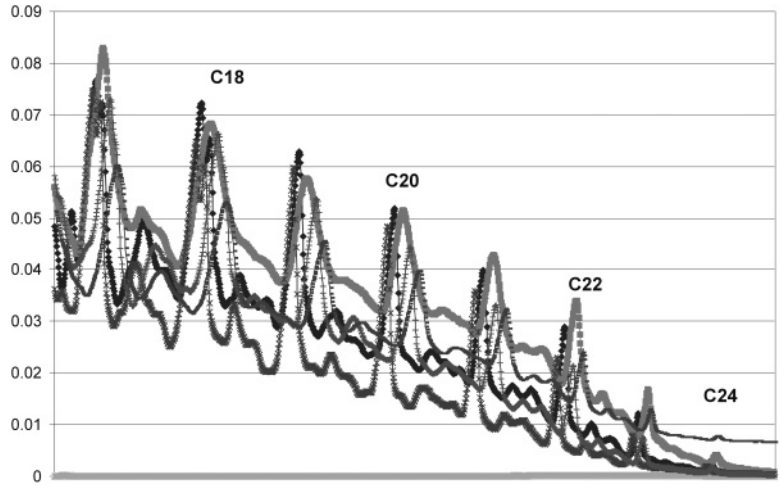


Figure 10.25 *C_{17}–C_{24} expansion of the simulated distillation chromatograms for a variety of diesel samples.*

gasoline and diesel have resulted in many refiners undercutting initial diesel distillation in order to ensure minimum sulfur breakthrough in the final product blends. Dibenzothiophenes in particular represent a class of sulfur compounds that are in the diesel distillation range, are difficult to detect on-line, are difficult to remove, and, contribute to sulfur concentrations near the final product limits. Depending on a refinery's crude supply and crude distillation capacity, a conservative estimate of 300–500 barrels per day of lost diesel production is typical as refiners minimize dibenzothiophene breakthrough. In an effort to improve the on-line control of dibenzothiophene breakthrough the combination of NMR and Simulated Distillation (SimDis) GC is being utilized to monitor the carbon number distributions in diesel distillation cuts. Multidetector SimDis analyses of diesel fuel enable the simultaneous measurement of carbon number, sulfur, and nitrogen species. These techniques have confirmed the dibenzothiophene components retention times correspond to the C_{19}/C_{20} region of the SimDis chromatogram. Correlations can be obtained relating the individual carbon number content from SimDis (Figure 10.25) with ^{1}H NMR spectra. ^{1}H NMR will be utilized to predict carbon number distributions in real time (Figure 10.26). This unique NMR measurement strategy enables crude unit operations to cut and control crude unit diesel production as closely to the dibenzothiophene distillation limit as possible without the direct measurement of these compounds.

10.7 Current Development and Applications

Expansion of process NMR into reaction monitoring is an area that seems to be ready to emerge, particularly in the pharmaceutical area. The chemical detail present in ^{1}H NMR spectra is unrivalled for monitoring chemical conversions. Microreactor technology combined with NMR probe design will allow small scale reactions to actually take place in the NMR magnet itself. Incorporation of HR-NMR and TD-NMR in the same NMR analyzer is another development that would allow simultaneous observation of sample chemistry and NMR relaxation. For example, in heavier petroleum streams viscosity is nonlinear in its response to sample chemistry which makes linear PLS modeling a challenge. However, T_2 relaxation times are directly related to viscosity, so a combination NMR analyzer would work well and configurations can be imagined where the same magnet and probe could be used for the two analyses.

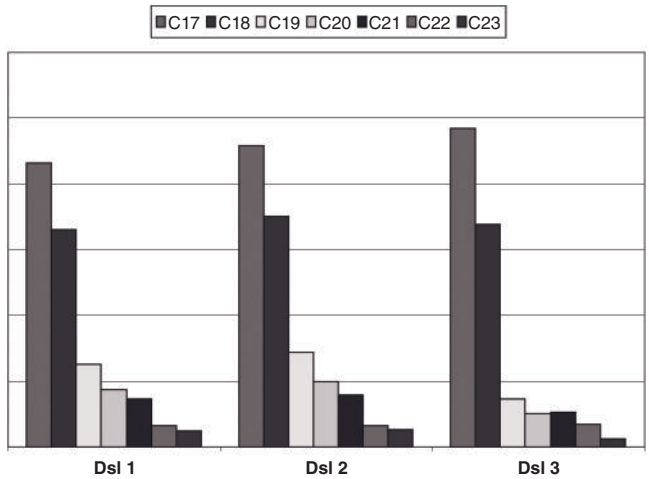


Figure 10.26 *Examples of ^{1}H NMR predicted carbon number distributions for three light gas oil samples.*

The combination of process NMR with electron spin resonance (ESR) will also expand process NMR's capabilities in areas where observation of free radicals or paramagnetic metals is required. Petroleum-related projects can benefit greatly from NMR analysis for hydrocarbon chemical and physical property determination in combination with ESR yielding quantitation on vanadium content of samples, as well as asphaltene content of heavier materials. Asphaltenes contain stable free radicals that are readily observed by ESR. Again a single magnet ESR-NMR hybrid instrument can be envisioned thus expanding the capabilities of a process NMR instrument. Hybridization of multiple analyzers into the NMR housing so that a single sample system can be used to deliver samples to a combination of spectroscopies (NMR, Mid-IR, NIR, Raman, ESR, UV-Vis) will lead to redundancy of analyses and a wider scope of possible parameters.

Finally, a step away from ^{1}H NMR in some applications may open up other areas of process analysis. In the current configuration the only nuclei that are really feasible are ^{19}F, ^{31}P, ^{27}Al, and ^{23}Na. Sodium content of liquid foodstuffs is an obvious area for ^{23}Na NMR. ^{31}P NMR has been targeted as a way to identify biological and chemical warfare agents in the field. A portable version of the 60 MHz NMR analyzer could be mounted in a van and driven to scenes of inquiry.

10.8 Conclusions

Our focus here has been predominantly on petroleum-related applications. However, many applications exist in biofuels, dairy, food manufacturing, beverages, pharmaceuticals, petrochemicals, etc. Feasibilities for many applications in these other sectors of industry have already been carried out. However, the main barrier to NMR entering these arenas seems to be the lack of realization, on the part of engineers, physicists and chemists, that NMR technology can be found in a compact form that is field-mountable, and produces stable, quantitative, reproducible data that can be automatically analyzed via algorithms and chemometrics to produce real-time process data. Our hope is that this short chapter will lead more researchers and process analysis personnel to think of NMR as a viable technique with certain advantages over other spectroscopies currently in use.

References

1. Bruker Optics, Billerica, MA, USA, http://www.brukeroptics.com/minispec.html
2. Oxford Instruments, Abingdon, UK, www.oxford-instruments.com
3. Resonance Systems, Ltd., Yoskar-Ola, Russia, www.mobilenmr.com
4. Varian, Inc., Palo Alto, CA, USA, www.varianinc.com/nmr
5. JEOL, Ltd., Tokyo, Japan, www.jeol.com
6. K. Albert, *On-Line LC-NMR and Related Techniques*, John Wiley & Sons, Ltd, Hoboken, 2002
7. M.A. Bernstein, M. Stefinovic, and C. J. Sleigh, Optimising reaction performance in the pharmaceutical industry by monitoring with NMR, *Magn. Reson. Chem.*, **45**, 564–571 (2007).
8. M. Maiwald, H.H. Fischer,Y-K. Kim and H. Hasse, Quantitative on-line high-resolution NMR spectroscopy in process engineering applications, *Anal. Bioanal. Chem.*, **375**, 1111–1115 (2003).
9. M. Maiwald, O. Steinhof, C. Sleigh, M. Bernstein, and H. Hasse, Quantitative high-resolution online NMR spectroscopy in pharmaceutical reaction and process monitoring, in *NMR Spectroscopy in Pharmaceutical Analysis*, U. Holzgrabe, I. Wawer, B. Diehl (Eds), Elsevier, Oxford, 471–490, 2008.
10. S. Clark, N.W. Barnett, M. Adams, I.B. Cook, G.A. Dyson, and G. Johnston, Monitoring a commercial fermentation with proton nuclear magnetic resonance spectroscopy with the aid of chemometrics, *Anal. Chim. Acta*, **563**, 338–345 (2006).
11. M. Otta, H.H. Fischer, M. Maiwald, K. Albert, and H. Hasse, Kinetics of oligomerization reactions in formaldehyde solutions: NMR experiments up to 373K and thermodynamically consistent model, *Chem. Eng. Process.*, **44**, 653–660 (2005).
12. H. Weber, L. Brecker, D. De Souza, H. Griengl, D.W. Ribbons, and H.K. Weber, Online NMR for monitoring biocatalysed reactions – the use of lipases in organic solvents, *J. Mol. Catal. B: Enzym.*, **19–20**, 149–157 (2002).
13. M. Maiwald, H.H. Fischer, Y.-K. Kim, K. Albert and H. Hasse, Quantitative high-resolution on-line NMR spectroscopy in reaction and process monitoring, *J. Magn. Reson.*, **166**(2), 135–146 (2004).
14. A. Nordon, C.A. McGill, and D. Littlejohn, Process NMR spectrometry, *Analyst*, **126**, 260–272 (2001).
15. A.K. Roy and S.A. Marino, NMR in process control, *Am. Lab.*, 32–33 (1999).
16. Process Control Technology Corp., Fort Collins, CO, USA, www.pctnmr.com
17. SpinCore Technologies, Inc., Gainesville, FL, USA, www.spincore.com
18. Process NMR Associates, LLC, Danbury, CT USA, www.process-nmr.com
19. NMR Process Systems, LLC, Font du Lac, WI, USA, www.nmr-automation.com
20. Magritek Ltd., Wellington, New Zealand, www.magritek.com
21. MR Resources, Fitchburg, MA, USA, www.mrr.com
22. Spinlock SRL, Cordoba, Argentina, www.nmr-spectrometers.com
23. Tecmag, Houston, TX, www.tecmag.com
24. ABQMR, Inc, Albuquerque, NM, USA, www.abqmr.com
25. Stelar, s.r.l, Mede (PV), Italy, www.stelar.it
26. Progression, Inc., Haverhill, MA, USA, www.progression-systems.com
27. S. Anferova, V. Anferov, M. Adams, *et al.*, Construction of a NMR-MOUSE with short dead time, *Concepts Magn. Reson.*, **15B**(1), 15–25 (2002).
28. S. Anferova, V. Anferov, D.G. Rata, *et al.*, A Mobile NMR device for measurements of porosity and pore size distributions of drilled core samples, *Concepts Magn. Reson.*, **23B**(1), 26–32 (2004).
29. B. Blumich, J. Perlo, and F. Casanova, Mobile single-sided NMR, *Prog. Nucl. Reson. Spectrosc.* **52**, 197–269 (2008).
30. aixNMR, Aachen, Germany, www.aixnmr.com
31. G.R. Coates, L. Xiao, and M.G. Prammer, *NMR Logging – Principles and Applications*, Gulf Professional Publishing, Houston, 2001.
32. D. Allen, C. Flaum, T.S. Ramakrishnan, *et al.*, Trends in NMR logging, *Schlumberger Oilfield Review*, Autumn 2000, 2–19.
33. B. Sun, In situ fluid typing and quantification with 1D and 2D NMR logging, *Magn. Reson. Imaging*, **25**(4), 521–4 (2007).
34. M.D. Hürlimann, L. Burcaw, and Y.-Q. Song, Quantitative characterization of food products by two-dimensional D-T2 and T1–T2 distribution functions in a static gradient, *J. Colloid Interface Sci.*, **297**, 303–311 (2006).
35. P.J. Giammatteo, J.C. Edwards, T. Cohen, and M.W. Blakley, Process applications of NMR on flowing gaseous streams. Part III. The installation and operation of an on-line NMR in a refinery, presented at 37th Experimental NMR Conference, Pacific Grove, CA, March 17–22, 1996.

36. J.C. Edwards and P.J. Giammatteo, On-line acid strength measurement and sulfuric acid alkylation process control using process NMR, *ISA Tech/Expo Technol. Update*, **2**, 63–67 (1998).
37. Qualion, Inc., Haifa, Israel, www.qualion-nmr.com.
38. S. Meiboom and D. Gill, Modified spin-echo method for measuring nuclear relaxation times, *Rev. Sci. Instrum.*, **29**(8), 688–691 (1958).
39. H.T. Pedersen, L. Munck, and S.B. Engelsen, Low-field ^{1}H nuclear magnetic resonance and chemometrics combined for simultaneous determination of water, oil, and protein contents in oilseeds, *J. Am. Oil Chem. Soc.*, **77**(10), 1069–1077 (2000).
40. R. Lamanna, On the inversion of multicomponent NMR relaxation and diffusion decays in heterogeneous systems, *Concepts Magn. Reson.*, **26A**(2), 78–90 (2005).
41. D. Le Botlan, L. Ouguerram, L. Smart, and L. Pugh, Characterization of a semisolid state in milk fat through T2* resolved T1 distributions by time domain nuclear magnetic resonance, *J. Am. Oil Chem. Soc.*, **76**(2), 255–261 (1999).
42. N. Marigheto, L. Venturi, D. Hibberd, K.M. Wright, G. Ferrante, and B.P. Hills, Methods for peak assignment in low-resolution multidimensional NMR cross-correlation relaxometry, *J. Magn. Reson.*, **187**, 327–342 (2007).
43. M.L. Johns and K.G. Hollingsworth, Characterisation of emulsion systems using NMR and MRI, *Prog. Nucl. Magn. Reson. Spectrosc.*, **50**(2–3), 51–70 (2007).
44. J. Arnold, C. Clauser, R. Pechnig, S. Anferova, V. Anferov, and B. Blümich, Porosity and permeability from mobile NMR core-scanning, *Petrophysics*, **47**(4), 306–314 (2006).
45. P.T. Callaghan, S. Godefroy, and B.N. Ryland, Diffusion–relaxation correlation in simple pore structures, *J. Magn. Reson.*, **162**, 320–327 (2003).
46. M. Braia, C. Casieri, F. De Lucab, P. Fantazzinie, M. Gombiae, and C. Terenzib, Validity of NMR pore-size analysis of cultural heritage ancient building materials containing magnetic impurities, *Solid State Nucl. Magn. Reson.*, **32**(4), 129–135 (2007).
47. T.M. Alam and M.K. Alam, Chemometric analysis of NMR spectroscopy data: A review. *Annual Reports on NMR Spectroscopy*, G.A. Web (Ed.), **54**, 41–80, Elsevier, 2005.
48. R. Meusinger, Qualitative and quantitative determination of oxygenates in gasolines using ^{1}H nuclear magnetic resonance spectroscopy, *Anal. Chim. Acta*, **391**, 277–288 (1999).
49. A. Nordon, C. Meunier, R.H. Carr, P.J. Gemperline, and D. Littlejohn, Determination of the ethylene oxide content of polyether polyols by low-field ^{1}H nuclear magnetic resonance spectrometry, *Anal. Chim. Acta*, **472**, 133–140 (2002).
50. J. Burri, R. Crockett, R. Hany, and D. Rentsch, Gasoline composition determined by ^{1}H NMR spectroscopy, *Fuel*, **83**, 187–193 (2004).
51. A.S. Sarpal, G.S. Kapur, S. Mukherjee, and A.K. Tiwari, PONA analyses of cracked gasoline by ^{1}H NMR spectroscopy. Part II, *Fuel*, **80**, 521–528 (2001).
52. G.S. Kapur, A.P. Singh, and A.S. Sarpal, Determination of aromatics and naphthenes in straight run gasoline by ^{1}H NMR spectroscopy. Part I, *Fuel*, **79**, 1023–1029 (2000).
53. T.W. Skloss, A.J. Kim, and J.F. Haw, High-resolution NMR process analyzer for oxygenates in gasoline, *Anal. Chem.*, **66**, 536–542 (1994).
54. R. Meusinger and R. Moros, Determination of octane numbers of gasoline compounds from their chemical structure by ^{13}C NMR spectroscopy and neural networks, *Fuel*, **80**, 613–621 (2001).
55. V. Bansal, G.J. Krishna, A.P. Singh, A.K. Gupta, and A.S. Sarpal, Determination of hydrocarbons types and oxygenates in motor gasoline: A comparative study by different analytical techniques, *Energy Fuels*, **22**, 410–415 (2008).
56. J.C. Edwards, P.J. Giammatteo, G.P. Firmstone, and P.D. Cusatis, Development of a process NMR gasoline analysis system, Presented at 37th Experimental NMR Conference, Pacific Grove, 1996.
57. D.A. Le Febre and M.S. Zetter, Process for controlling blending using nuclear magnetic resonance spectroscopy, United States Patent 5,796,251, 1998.
58. R. Meusinger, Gasoline analysis by ^{1}H nuclear magnetic resonance spectroscopy, *Fuel*, **75**(10), 1235–1243 (1996).
59. B. Diehl and G. Randel, Analysis of biodiesel, diesel and gasoline by NMR spectroscopy – a quick and robust alternative to NIR and GC, *Lipid Technology*, **19**(11), 258–260 (2007).
60. M.R. Monteiro, A.R.P. Ambrozin, L.M. Liao, E.F. Boffo, L.A. Tavares, M.M.C. Ferreira, and A.G. Ferreira, Study of Brazilian gasoline quality using hydrogen nuclear magnetic resonance (^{1}H NMR) spectroscopy and chemometrics, *Energy Fuels*, **23**(1), 272–279 (2009).
61. F.H. Larsen, F. van den Berg, and S.B. Engelsen, An exploratory chemometric study of ^{1}H NMR spectra of table wines, *J. Chemom.*, **20**, 198–208 (2006).

62. B. Basu, G.S. Kapur, A.S. Sarpal, and R. Meusinger, A neural network approach to the prediction of cetane number of diesel fuels using nuclear magnetic resonance (NMR) spectroscopy, *Energy Fuels*, **17**, 1570–1575 (2003).
63. V. Bansal, G.S. Kapur, A.S. Sarpal, V. Kagdiyal, S.K. Jain, S.P. Srivastava, and A.K. Bhatnagar, Estimation of total aromatics and their distribution as mono and global di-plus aromatics in diesel-range products by NMR spectroscopy, *Energy Fuels*, **12**, 1223–1227 (1998).
64. G. Del Campo, J.I. Santos, N. Iturriza, I. Berregi, and A. Munduate, Use of the ^{1}H nuclear magnetic resonance spectra signals from polyphenols and acids for chemometric characterization of cider apple juices, *J. Agric. Food Chem.*, **54**, 3095–3100 (2006).
65. G.E. Pereira, J-P. Gaudillere, C. Van Leeuwen, *et al.*, ^{1}H NMR and chemometrics to characterize mature grape berries in four wine-growing areas in Bordeaux, France, *J. Agric. Food Chem.*, **53**, 6382–6389 (2005).
66. I. Duarte, A. Nio Barros, P. S. Belton, R. Righelato, M. Spraul, E. Humpfer, and A. M. Gil, High-resolution nuclear magnetic resonance spectroscopy and multivariate analysis for the characterization of beer, *J. Agric. Food Chem.*, **50**, 2475–2481 (2002).
67. R. Consonni and L.R. Cagliani, Geographical characterization of polyfloral and acacia honeys by nuclear magnetic resonance and chemometrics, *J. Agric. Food Chem.*, **56**(16), 6873–6880 (2008).
68. R. Garner and P. Mayo, Online prediction of mechanical properties in PP using a low resolution NMR, available www.progression-systems.com/pdfs/PP0905.pdf accessed 19 November 2009.
69. S. Marino, Product consistency and manufacturing flexibility using process NMR, available www.progression-systems.com/pdfs/PP paper 2003.pdf accessed 19 November 2009.
70. S. Marino, Advances in polyethylene analysis by NMR, available www.progression-systems.com/pdfs/PE%20 paper%202005.pdf accessed 16 December 2009.
71. M. Sardashti, J.J. Gislason, X. Lai, C.A. Stewart, and D.J. O'Donnell, Determination of total styrene in styrene/butadiene block copolymers by process NMR and chemometrics modeling, *Appl. Spectrosc.*, **55**(4), 467–471 (2001).
72. J. Gislason, H. Chan, and M. Sardashti, Calibration Transfer of chemometric models based on process nuclear magnetic resonance spectroscopy, *Appl. Spectrosc.*, **55**(11), 1553–1560 (2001).
73. X. Lai, M. Sardashti, B.J. Lane, J.J. Gislason, and D.J. O'Donnell, Simulation studies on modeling process NMR data by using chemometrics approach, *Appl. Spectrosc.*, **54**(1), 54–61 (2000).
74. S. Chan and D.T. Georgi, In-situ heavy-oil reservoir evaluation with artificial temperature elevation, United States Patent 6,755,246, 2004.
75. G. Taicher and A. Reiderman, NMR well logging apparatus and method, United States Patent 6,018,243, 1997.
76. C.M. Edwards and D.T. Georgi, High-resolution high-speed NMR well logging device, United States Patent, 7,012,426, 2006.
77. B. Sun and K.J. Dunn, A global inversion method for multi-dimensional NMR logging, *J. Magn. Reson.*, **172**(1), 152–160 (2005).
78. G.J. Hirasaki, S.W. Lo, and Y. Zhang, NMR properties of petroleum reservoir fluids, *Magn. Reson. Imaging*, **21**(3–4), 269–277 (2003).
79. R.L. Kleinberg and J.A. Jackson, An introduction to the history of NMR well logging, *Concepts Magn. Reson.*, **13A**(6), 340–342 (2001).
80. J. C. Goswami, A. Sezginer, and B. Luong, On the design of NMR sensor for well-logging applications, *IEEE Antennas Propag.*, **48**(9), 1393–1402 (2000).
81. S. Godefroy, M. Fleury, F. Deflandre, and J.-P. Korb, Temperature effect on NMR surface relaxation in rocks for well logging applications, *J. Phys. Chem. B*, **106**(43), 11183–11190 (2002).
82. B. Blumich, S. Anferova, R. Pechnig, H. Pape, J. Arnold, and C. Clauser, Mobile NMR for porosity analysis of drill core sections, *J. Geophys. Eng.* **1**, 177–180 (2004).
83. K. Mirotchnik, K. Allsopp, A. Kantzas, and D. Marentette, Quantification of bitumen using NMR, *United States Patent* 6,630,357 (2003).
84. Z. Yang and G. J. Hirasaki, NMR measurement of bitumen at different temperatures, *J. Magn. Reson.*, **192**(2), 280–293 (2008).
85. D. Gabriele, M. Migliori, R. Di Sanzo, C.O. Rossi, S.A. Ruffolo and B. de Cindio, Characterisation of dairy emulsions by NMR and rheological techniques, *Food Hydrocolloids*, **23**, 619–628, (2009).
86. M.D. Hürlimann, L. Burcaw, and Y. Song, Quantitative characterization of food products by two-dimensional D–$T2$ and $T1$–$T2$ distribution functions in a static gradient, *J. Colloid Interface Sci.*, **297**, 303–311 (2006).
87. G. Guthausen, H. Todt, W. Burk, D. Schmalbein, and A. Kamlowski, Time-domain NMR in quality control: more advanced methods, in *Modern Magnetic Resonance*, G.A. Webb (ed.), Springer, Netherlands, 2006.

88. H. Todt, G. Guthausen, W. Burk, D. Schmalbein, and A. Kamlowski, Time domain NMR in quality control: standard applications in food, *Modern Magnetic Resonance*, G.A. Webb (ed.) Springer, Netherlands, 2006.
89. E. Trezza, A.M. Haiduc, G.J.W. Goudappel and J.P.M. van Duynhoven, Rapid phase-compositional assessment of lipid-based food products by time domain NMR, *Magn. Reson. Chem.*, **44**, 1023–1030 (2006).
90. H.T. Pedersen, S. Ablett, D.R. Martin, M.J.D. Mallett, and S.B. Engelsen, Application of the NMR-MOUSE to food emulsions, *J. Magn. Reson.*,**165**, 49–58 (2003).
91. A.M. Haiduc, E.E. Trezza, D. van Dusschoten, A.A. Reszka, and J.P.M. van Duynhoven, Non-invasive 'through-package' assessment of the microstructural quality of a model food emulsion by the NMR MOUSE, *LWT- Food Sci. Technol.*, **40**, 737–743 (2007).
92. J.M. McKendry, R. Selway, and J.A.W.M. Corver, Method for magnetic field tracking in a NMR check weighing system, United States Patent 7061239 (2006).
93. B.P. Hills and K.M. Wright, Motional relativity and industrial NMR sensors, *J. Magn. Reson.*, **178**, 47–59 (2006).
94. G.H. Sørlanda, P.M. Larsena, F. Lundby, A.-P. Rudic, andT. Guiheneuf, Determination of total fat and moisture content in meat using low field NMR, *Meat Sci.*, **66**, 543–550 (2004).
95. S.B. Engelsen, M.K. Jensen, H.T. Pedersen, L. Nørgaard and L. Munck, NMR-baking and multivariate prediction of instrumental texture parameters in bread, *J. Cereal Sci.*, **33**, 59–69 (2001).
96. P. Giammatteo and J. Edwards, Reliable refinery control using process NMR, *Control*, **6**, 71–74 (1999).
97. Process NMR Associates, www.process-nmr.com/fox-app/alkltn.htm.
98. Process NMR Associates, www.process-nmr.com/reformer_and_platformer.htm
99. J. C. Edwards and P. J. Giammatteo, Detailed hydrocarbon analysis of naphtha by on-line NMR: integration of real-time NMR Feed analysis with advanced process control and optimization, Presented at Eastern Analytical Symposium, Somerset, NJ, November 18–21, 2002.
100. P.A. Barnard, C. Gerlovich, and R. Moore, The validation of an on-line nuclear magnetic resonance spectrometer for analysis of naphthas and diesels, presented at the ISA Analytical Division Meeting, Calgary, Alberta, Canada, 2003.
101. J.C. Edwards, Steam cracker performance optimized by process MRA, *Control*, **16**(6), 69–71 (2003).
102. L. Corrado, L.J. Roobottom, and L.D. Palmer, Application of magnetic resonance analysis (MRA) for the production of clean fuels, Presented at Chemeca 2005, Brisbane, Australia, 2005.
103. R. Seshadri, A. Offerman, J. Cijntje, R. McFarlane, and V. Lough, Optimization of a refinery crude unit, *Pet. Tech. Q.*, 1–7, Winter 2002.
104. P.J. Giammatteo and J.C. Edwards, On-line analysis of crude feeds and distillation products: utilization of on-line NMR on a refinery's crude distillation unit, Presented at Eastern Analytical Symposium, Somerset, New Jersey, 2002.
105. Process NMR Associates, available www.process-nmr.com/pdfs/Crude Unit Control – Eastern Analytical Symposium.pdf accessed 19 November 2009.
106. S. Guanziroli, R. Giardino, A. Farina, V. Lough, and J.C. Edwards, Process MRA – lube plant application, presented at the European Refining Technology Computing Conference, Milan, Italy, 2003.

11

Fluorescent Sensing and Process Analytical Applications

Jason E. Dickens

Preclinical Development, GlaxoSmithKline, Research Triangle Park, NC, USA

11.1 Introduction

Luminescence, in particular photoluminescence, constitutes a well-established discipline in analytical science where the cited hallmarks include remarkable sensitivity, wide dynamic range and low detection limits (~10^{-12} M under suitable conditions). These collective merits are often unrivaled by other optical techniques, and hence its wide adoption in the life sciences for determining trace constituents in biological and environmental samples. Moreover, its fast response, high spatial resolution and remote sensing capabilities make it attractive for real-time analytical problems such as process manufacturing (process analysis or PAT) and field applications.[1]

While near-infrared (NIR) and other conventional spectroscopic process analytical technologies (PAT) are proving effective for process understanding and control across industries, there are some applications where these techniques are insufficient. These include applications that require detection at low analyte concentrations, higher measurement precision, high sensitivity or a wide dynamic range.[2] Steady state intrinsic photoluminescent sensing is well suited to fill this gap when the target analyte(s) is a native fluorophore and when the blank sample matrix contains negligible emissive interferants with similar spectral properties. Although steady state intrinsic fluorescence provides moderate discrimination power to resolve multiple fluorohores with similar spectral properties, various approaches are available (e.g., chemometrics and chemical optodes) to provide the required resolution. In addition, emission enhancement reagents and fluorescent derivatization tags are available for weak and nonfluorescent analytes, respectively.[3]

The purpose of this chapter is to provide an overview of luminescence techniques for real-time monitoring. This review provides: (i) an introduction to luminescence fundamentals; (ii) an outline of various luminescence techniques, with particular focus on steady state photoluminescent sensing; (iii) an overview of the emission measurement risks; and (iv) a discussion of current and potential PAT applications.

Process Analytical Technology 2e Edited by Katherine Bakeev
© 2010 John Wiley & Sons, Ltd.

11.2 Luminescence Fundamentals

Luminescence is a well-established class of analytical spectroscopic techniques where a species emits light after excitation. Emission is an electronic transition from an excited state as opposed to the ground state as is the case in most other spectroscopies. Photoluminescence, or light-induced fluorescence (LIF), is the most common route to induce emission where sufficient incident photons of a particular energy excite the target species via absorption. Although less common, nonradiative excitation can also occur via a chemical reaction termed chemiluminescence. Unless otherwise stated, the terms luminescence and fluorescence within this review infers excitation by light induction.

Photoluminescence involves three types of information: (i) emission versus wavelength (spectral-domain), (ii) intensity across a specific wavelength bandwidth (intensity-domain); and (iii) emission decay over time (time-domain or lifetime) where each fluorophore has a unique lifetime. Photoluminesence detection modes in general are classed as either steady state or time-domain where the former involves either spectral- or intensity-domain detection modes.

11.2.1 Luminescence nomenclature

In recent years luminescence nomenclature has become confusing within the literature and in practice. Luminescence involves both phosphorescence and fluorescence phenomena. While luminescence is the appropriate term when the specific photochemical mechanism is unknown, fluorescence is far more prevalent in practice. Moreover, the acronym 'LIF' has historically inferred 'laser'-induced fluorescence; however, in recent years it has evolved to the more general term 'light'-induced fluorescence due to the various light sources found within laboratory and real-time instruments. Within this chapter fluorescence and LIF are interchangeable terms.

11.2.2 Luminescence processes

Luminescence is a two-stage electronic transition process involving excitation followed by emission as illustrated by the energy diagram in Figure 11.1 first suggested by Jablonski, where the ground state, singlet and triplet excited states are depicted as S_0, S_1 and T_1, respectively.[3,4] As the LIF term implies, light of a particular energy induces excitation (transition 1) where an electron is promoted to an excited state (e.g., S_1 or S_2). After vibrational relaxation (transition 2) the electron either returns to the ground state with an emission of a photon (transition 4; fluorescence) or the competing nonradiative intersystem crossing (transition 3) process occurs where the electron moves to a lower excited state. Within this S_1 state, the electron further relaxes and then either returns to the ground state by emitting a photon, decays further via a nonradiative external conversion (transition 5) or the less prevalent disallowed singlet to triplet intersystem crossing transition occurs (transition 6). This latter transition leads to either phosphorescence (transition 7) or the predominant external conversion route, in particular at ambient temperatures and above. Additional competing processes are discussed later within this review (see luminescent sensing risks).

11.2.3 Fluorophore classification

A fluorescent species is termed a fluorophore or fluor analogous to a chromophore in ultraviolet-visible (UV-vis) spectroscopy. Fluorophores are native chemical species or moieties that exhibit emission upon excitation, often by incident light of the proper wavelength and intensity. Molecular structural analysis serves

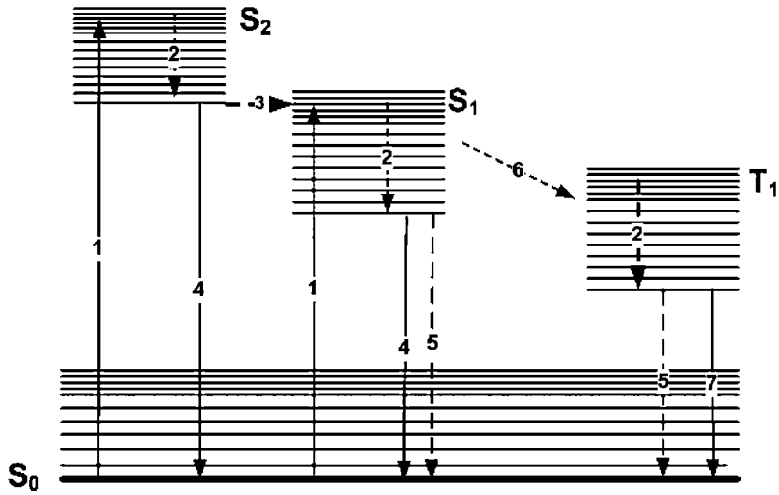


Figure 11.1 *Luminescence Jablonski diagram.*

as a good first estimate of emission properties where it is most often associated with extensive conjugated π-electron bonding systems as found with aromatic, cyclic or closed ring structures such as polycyclic aromatic structures or highly unsaturated molecules. Thousands of native fluorophores are known such as anthracene, organic dyes, fluorescein, coumarin and cyanine.[3] Few saturated organic compounds and limited inorganic molecules produce intense emission. While structural analysis serves as a good first estimate of emission properties, for determining emission exceptions do occur and thus the target analyte should be screened by empirical spectrofluorometry.[5]

The application potential of intrinsic LIF monitoring is based on the wide number of native fluorophores across the petrochemical, chemical, and pharmaceutical manufacturing industries, and among other sectors such as marine science, and defense, including surveillance and environmental (see Section 11.6). There is particular interest in its use within the pharmaceutical industry as a significant number of active pharmaceutical ingredients (APIs) are probable fluorophores[6–8] while most excipients are weak or negligible emitters. Excipients derived from natural sources such as cellulosics however contain weak fluorophores and may be a problematic interferants in some situations.[9] Nevertheless, when emissive interferants are negligible, LIF is well suited for very low API content where other typical process spectroscopic techniques (e.g., near-infrared spectroscopy, NIR) provide inadequate detection.

Other fluorophores such as biochemical reactants and intermediates (e.g., proteins or amino acids, cellular biofluorophores, etc.) are also of interest.[10] For example, intrinsic LIF sensing has been demonstrated in bioprocesses where protein productivity is optimized by monitoring green fluorescent protein (GFP), a noninvasive quantitative fusion marker of foreign protein production in *Escherichia coli*.[11] Similarly, monitoring intrinsic cellular biofluorophores, reduced nicotinamide adenine dinucleotide phosphate (NADPH), tryptophan, pyridoxine and riboflavin, by steady state LIF sensing during the growth phase has shown good correlation with biomass concentration and other biological parameters for well-defined fermentation media.[12–14]

Extrinsic fluors are produced via a chemical reaction where the added reagent either enhances emission of a weak emitter through association or the analyte is derivatized with a fluor tag. 8-Hydroxyquinoline (HQS) is an example of an extrinsic complexing reagent (Reaction 11.1) where the native ligand is a marginal fluorophore but forms intense emitting metal chelates.[15] This approach affords sensitive detection of

various cations including Ca, Zn, Cd, Al and Mg.[16–18] Unfortunately, most other metal-HQS complexes are negligible emitters, thus inhibiting wider use of this sensitive reagent.

$$Mg^{2+} + HQS^{2-} \leftrightarrow Mg(HQS)_2^{2-} \qquad \text{(Reaction 11.1)}$$

Some weak fluorophores exhibit enhanced emission behavior upon association within restrictive environments such as membranes, micelles and guest-host complexes.[19] Inhibiting environments of organic media such as surfactants above their critical micelle concentration and cyclodextrins are examples of reagents that enhance emission intensity.[19] These types of emission enhancements have been realized with various organic compounds such as aromatic pesticides,[20] mycotoxins (e.g, aflatoxins, zearalenone and ochratoxins)[21] and hallucinogenic drugs, *N*,*N*-dimethyltryptamine, mescaline and ibogaine.[19]

Extending the utility of fluorescence to various nonfluorophores is achieved via chemical derivatization methods, also termed labeling or tagging methods (Reaction 11.2). Numerous commercial fluorescent tags are available with disparate reactive functional groups.[3] For example, derivatives of fluorescein, fluorescein isothiocyanate (FITC) are reactive toward nucleophiles such as amines and sulfhydryl groups.

$$A + B_{tag} \rightarrow AB_{tag} + h\nu_1 \rightarrow AB^* \rightarrow AB + h\nu_2 \qquad \text{(Reaction 11.2)}$$

Fluoroimmunoassays comprise a subclass of extrinsic labeling methods where various selective antigen (Ag)- antibody (Ab) immunoassay fluorescent labeling schemes yield a emission signal. One common scheme involves an enzyme-linked immunosorbent assay (ELISA) depicted in Figure 11.2 where the free Ab is tagged with a fluorophore. Numerous analytes can be detected via these types of selective lock-and-key methods.[3]

Chemiluminescence is another extrinsic approach where emissive intermediates are produced through the course of a chemical reaction (Reaction 11.3) and thus does not involve incident radiation. The reaction of luminol (5-amino-2,3-dihydro-1,4-phthalazinedione) and hydrogen peroxide is a classic example of such a reaction (Reaction 11.4).

$$A + B \rightarrow AB^* \rightarrow [Products] + \text{light}\ (h\upsilon) \qquad \text{(Reaction 11.3)}$$

$$\text{luminal} + H_2O_2 \rightarrow \text{3-APA}^* \rightarrow \text{3-APA} + \text{light}\ (h\nu) \qquad \text{(Reaction 11.4)}$$

As a process analytical solution, these extrinsic reactive approaches necessitate an extrinsic optode[22] (see later discussion), an on-line sample conditioning system[23] or an at-line solution such as a flow injection analysis (FIA) system[24] or other autonomous solutions.[3] Reaction kinetics, post analysis cleanup such as rejuvenating a substrate (optode, immobilized based immunoassays, etc.) among other complexities are additional considerations for these types real-time analysis methods.[3,7]

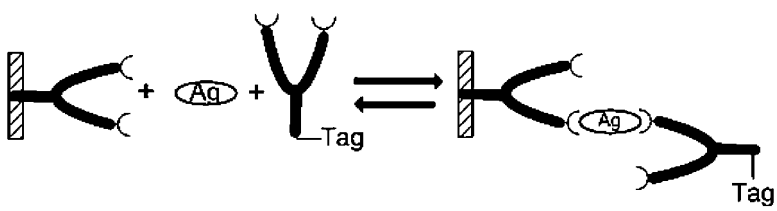


Figure 11.2 *Antigen–antibody fluorescent labeling scheme (Reaction 11.2).*

11.3 LIF Sensing Fundamentals

11.3.1 Light-induced fluoresence (LIF) sensing classification

LIF sensing is classified on the basis emissive properties of the analyte and the detection mode. Intrinsic LIF sensing is the most common in real-time analytical applications where a native fluorophore(s) is monitored. In contrast, extrinsic LIF sensing involves an additive reagent(s) that is either immobilized on a suitable substrate (e.g., fiber optic) or added via autonomous means as described above. Steady state or time-domain detection further categorizes LIF sensing, where the former is more common for real-time problems. Steady state intrinsic LIF sensing most often is achieved with filter-based or photometric instruments due to the underpinning broad and mostly featureless absorption and emission spectra in the liquid and solid states. LIF photometers available on the market and tailored for real-time PAT and field applications contain either single or multiple excitation–emission combination detection schemes where the former are blank-limited devices. That is, these devices cannot distinguish among multiple fluorophores with overlapping or interfering background emissions. When the underpinning emissions spectra of multiple fluors do not overlap to a large extent, a multiple excitation–emission detection scheme provides spectral selectivity (see Section 11.4.1) or in some cases chemometric analysis including data preprocessing can resolve overlapping emissive spectral responses.

Improved LIF sensing discrimination power is required for sample matrices that contain multiple fluorophores with similar spectral emission properties or when background emission is problematic. Distinguishing among airborne bioagent hazards and common emissive interferants (albuminous, epithelium, and cellulous materials as well as aromatic hydrocarbons), is a prime example where higher selective detection is required. This can be achieved via the lifetime properties of each fluorophore, by an optode approach or both.

Emission lifetime provides a powerful selectivity dimension as each fluorophore has a unique decay rate. While laboratory fluorescent lifetime measurement systems are widely available and the concept of lifetime LIF sensing has been demonstrated,[25,26] commercial LIF lifetime sensors are not yet available on the market for PAT and field applications to monitor organic analytes or bioagents. Once available, lifetime LIF sensors could revolutionize fluorescent based monitoring by affording superior detection merits along with sufficient recognition power.

Extrinsic LIF optodes provide another alternative approach that involves chemical agent(s) immobilized at the distal tip of a fiber-optic probe to impart analyte recognition, enhance/quench emission or both. There are various LIF optode detection schemes such as quenching, resonance energy transfer (RET), guest–host complexation and photo-induced electron transfer (PET) analyte recognition. Oxygen sensing is a classic example of an LIF quenching optode. LIF oxygen sensing is based on emission quenching of decacyclene or an equivalent agent (e.g., metal–ligand complexes; $[Ru(Ph_2phen)_3]^{2+}$) in the presence of oxygen and follows the Stern–Volmer relationship (Equation 11.4).[3] Bromine, iodine and chlorine, as well as SO_2 and NO are other examples of LIF-quenching optodes. Examples of RET sensors include pH and glucose ion sensing such as Ca^{2+}, Mg^{2+}, K^+ (in the presence of Na^+). Finally, analyte recognition optodes include metal chelates, glucose complexes and detection of various analytes via protein association immunosensors.[3,27] Various molecular recognition optodes have also been demonstrated that are based on modified cyclodextrins, molecular imprints and other guest-host complexes. Other than O_2, pH and CO_2 optodes, extrinsic LIF sensors are generally not yet available on the market.[28] A complete treatment on LIF optodes and time-domain sensing can be found in dedicated books on the topic and cited literature.[3,22,27,29,30]

Two-dimensional spectrofluorometry along with appropriate chemometric treatment is yet another alternative for real-time multicomponent analysis,[3] which has been utilized in bioprocess applications to monitor various cellular biofluorphores.[12–14] The fundamental principles of this type of spectroscopy can be found elsewhere.[3,31]

11.3.2 Luminescence spectroscopy

A discussion on steady state fluorescent monitoring necessitates a distinction between spectroscopic and photometric measurements. The former involves a grating-based spectrofluorometer where full spectrum excitation and emission multivariate spectra are acquired. In contrast a filter photometer involves optical elements (e.g., optical filters) to isolate excitation and emission bands thereby resulting in a univariate output emission response.

Spectrofluorometry underpins all luminescent techniques for characterizing the target analyte and the associated sample matrix. LIF occurs from the vacuum UV through the NIR range, but is most common between 250 nm and 800 nm. A typical fluorescent profile contains an absorption spectrum and an emission spectrum as depicted in Figure 11.3. Spectrofluorometric excitation profiling at emission maxima and synchronous scanning are common methods to characterize a fluorophore.

Solid and solution phase fluorescent spectra at room temperature exhibit relatively broad, often mostly featureless excitation (absorption) and emission spectra, particularly when compared to mid- and far-infrared spectroscopies. These spectra are often mirror images of each other but there are several exceptions as a result of either disparate molecular geometries between the ground and excited states or when the fluor is an excimer.[32]

Fluorescent spectra are further characterized by the Stokes shift (δ), which is the differential between excitation and emission peak maxima ($\delta = \lambda_{ex} - \lambda_{em}$) and range from 10 nm to 150 nm. The Stokes shift and other spectra characteristics (i.e., shape and number of bands) are dependent upon the fluorophore, the sample matrix and other conditions (e.g., pH, temperature, etc.).

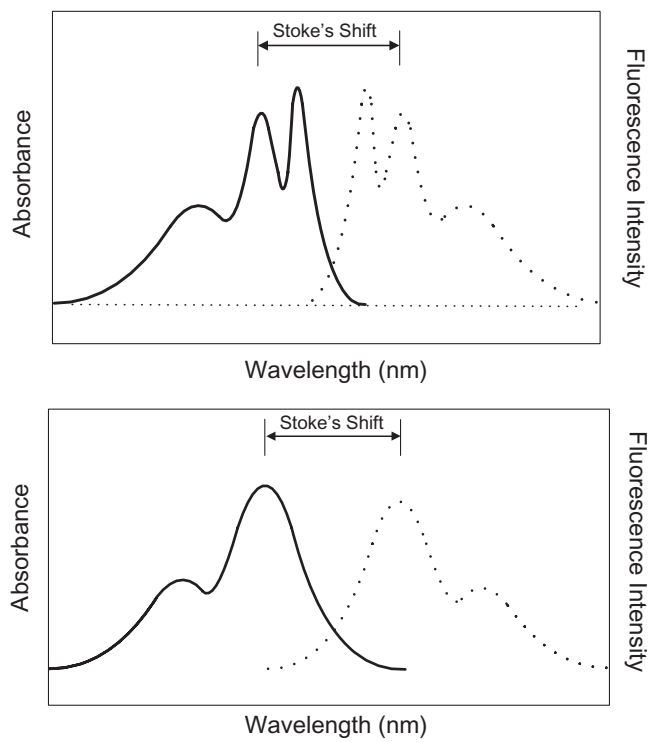


Figure 11.3 *Representative absorption (—) and emission (---) fluorescent spectral profiles.*

11.3.3 LIF signal response function

The fluorescent advantage is comprised of both the benefits of the Stokes shift and the dependence of excitation power on the emission signal, which leads to superior sensitivity and low detection limits with intense excitation. The Stokes shift is advantageous relative to ratiometric or absorption-based spectroscopic techniques (NIR, UV-vis, etc.) as the emission signal is detected against a negligible background (i.e., dark background signal); assuming the blank sample matrix does not contain interfering fluorophores and spectral interferants (e.g., Raman and Rayleigh scattering). The lower background signal and reduced noise results in superior signal-to-background ratios (SBR) and of more fundamental importance higher signal-to-noise (SNR). The latter is described in Equation 11.1, which accounts for the sources of variance within an optical system:

$$\sigma_T^2 = \sigma_S^2 + \sigma_D^2 + \sigma_O^2 + \sigma_B^2 \tag{11.1}$$

where σ_T is the total signal standard deviation and σ_S , σ_D , σ_O and σ_B is the optical source, detector, supporting optics (dispersive elements, etc.) and background contributions to the emission signal variance observed.[32] Consequently, LIF filter photometers relative to dispersive instruments exhibit higher SNR as these devices are constructed with mostly solid-state components that results in lower $\sigma_{O.}$

The direct proportional effect of source optical power on the emission signal is described by Equation 11.2:[33]

$$I_L = P_0 L \gamma \phi \varepsilon c \tag{11.2}$$

where I_L is the luminescent signal intensity, P_0 the incident excitation optical power, L the optical path length, γ the luminescent collection efficiency, ϕ the fluor luminescent quantum efficiency, ε the extinction coefficient of the luminescent species (fluor) and c is the concentration of the fluor. $P_{0,}$ L and γ instrument parameters and the latter three variables (ε, ϕ, c) are dependent on the fluorophore and the sample matrix. The magnitude of the source optical power and its precision (σ_S) influence the response functions as described by Equations 11.1 and 11.2. Relative to a conventional source such as the common xenon arc lamp, intense and stable light sources such as lasers, light-emitting diode (LED) arrays and laser diode (LD) arrays are ideal as they can yield a strong emission response with low signal noise. Real-time dynamic optical power control is a further benefit of these sources. This capability provides tailored analytical performance (sensitivity, detection limits, etc.) without a significant change in precision, which is not afforded by detector-gain limited optical techniques (e.g., NIR, UV-Vis, etc.). Characteristics and performance of the optical source and excitation throughput are pivotal instrument specification criteria and is thus discussed further in the following sections.

11.4 LIF Sensing Instrumentation

This section describes characteristics and selection of commercially available spectral and photometric LIF process instruments. While a comprehensive assessment was the objective, the author acknowledges that some vendors and instruments may have been inadvertently omitted. Nevertheless, the following survey of LIF instruments suitable for process and field environments was based on instruments defined within scientific literature and those actively marketed by common process analytical instrument vendors. Thus, emerging in-line LIF instruments is perhaps a potential gap within the following discussion.

Although spectrofluorometry has been ubiquitously adopted within the life sciences for complex clinical assays and analysis, its application to process analytical problems in routine manufacturing has been more

limited as compared to other spectrocopies (e.g., NIR, IR, UV-vis). This is due to the insufficient market availability of suitable process spectrofluorometers where there is for laboratory-based spectrometers. This is surprising based on the numerous reports demonstrating the utility of full-spectrum fluorescent monitoring in biochemical manufacturing applications.[11,34–38] The rationale for *in situ* spectrofluorometry is based on the need for sufficient control of a biochemical manufacturing process, which are often complex multifaceted problem necessitating the monitoring of multiple biomarkers. There are numerous additional real-time problems necessitating a full-spectrum ruggedized instrument. Delta (DELTA Light & Optics, Denmark) however appears to be the sole provider of a in-line process fluorometer (BioVIEW) for this type of application.[39] The BioVIEW detection scheme is a 16-channel filter wheel photometer rather than grating spectrometer. Nevertheless, appropriate cycling of each filter wheel generates sufficient 2D spectra acquired at 20-nm resolution without loss of information relative to higher resolution grating based instruments.[38] Consequently, the superior optical throughput in filter photometers provides optimal detection performance (SNR, sensitivity, etc.) relative to grating spectrofluorometers.

Carl Zeiss, Inc. also describes a spectrofluorometer system for process monitoring,[40] but it does not currently appear as a standard marketed product on their web site. HORIBA Jobin Yvon also markets a fluorescent process analyzer, but it is a laser-induced time-domain based measurement system tailored for uranium or equivalent analysis.[41] Finally, while numerous miniature spectrofluorometers are also available (Carl Zeiss, StellarNet Inc., Ocean Optics and Avantes), they are not packaged and configured for process applications. Although there is an established need and continued growing interest in real-time process spectrofluorometry, relative to conventional process spectroscopic instruments such as NIR, UV-vis and Raman, commercial process spectrofluorometers are currently available on a very limited basis.

In contrast, various LIF photometric instruments are available for process and field applications. Most of these instruments are solid state single-channel (single excitation and emission combination) filter based instruments and thus are appropriate for detecting a single fluor within a matrix containing negligible background fluorescence (i.e., blank-limited). These devices are preconfigured for the detection of the target fluorophore and most often are deployed for in-line intrinsic monitoring.

It is useful in the context of LIF photometers to differentiate between analyzers and sensors. An analyzer is a system of moderate size and weight that requires external utilities (power, plumbing, etc.), often the routing of fragile fiber optics or an instrument-to-probe conduit and a hardened or permanent installation. In most cases analyzers are not attractive solutions for remote monitoring (e.g., field), 'mobile' applications (e.g., pharmaceutical tumble blending) and distributed deployment across a manufacturing line or factory. These real-time analytical problems necessitate 'sensors', which are compact, self-contained and provide versatile operational attributes (e.g., powering, communications, etc.).

There are numerous providers of commercial LIF photometric analyzers and sensors for process and field analytical problems. Teledyne Analytical Instruments offers the 6000 series photometer analyzer with a flow cell for integration within a flowing liquid stream.[42] Custom Sensors and Technology (CST) provides a similar photometric analyzer with versatile process integration options including flow cells and insertion transmission and reflectance probes for monitoring various matrices including liquids, solids, pastes and slurries.[43] Both of these devices are broadband instruments with a xenon arc lamp optical source and suitable optical filtering. Turner Design[44] also offers various in-line and at-line LIF instruments that are often tailored for liquid phase applications such as water quality and flowing streams. With the market availability of LEDs throughout the UV-vis range[45] range, various LED-based analytical sensors have recently appeared on the market. For example, Turner Design supplies several submersible LIF sensor devices for marine science applications. In addition, CST in collaboration with GlaxoSmithKline has recently developed a novel LED array-based LIF sensor with versatile analytical and operational merits to meet various PAT and field application requirements across sectors.[46]

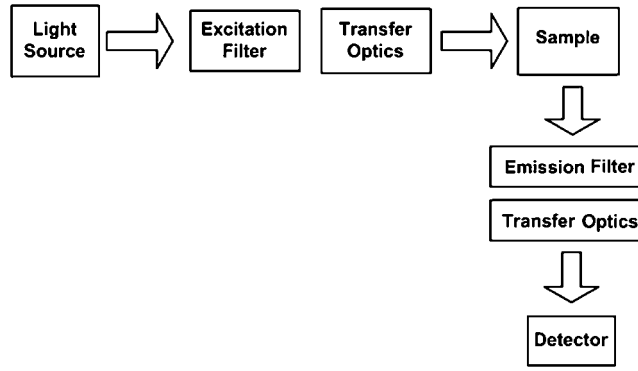


Figure 11.4 *General LIF photometric block diagram.*

11.4.1 LIF photometric instrument specification

Photometric devices require optical specification to detect the target fluorophore. Figure 11.4 depicts the general block for an LIF photometric device. The excitation source is either a broadband white source such as the common xenon arc lamp or a narrow band source such as an LED, LD or a laser. The most common detector is a photomultiplier tube (PMT), but photodiodes can also be used for low sensitivity applications. The transfer optics often include focusing lenses, fiber optics and in some optical configurations a dichroic mirror for splitting the incident radiation and the resultant emission. Photometers are either single-channel with a dedicated excitation-emission combination or multi-channel with various excitation-emission combinations via autonomous mechanical filter wheels. With the advances in microelectro-optical elements such as LEDs and smaller PMT detectors, alternative and perhaps more attractive detection schemes are now possible.

The detection performance of an LIF photometric device is governed by the emission filter(s), excitation filter(s), detector type, the excitation source and the detection scheme.[29,32,33] The selection of optical elements and device configuration as it relates to the detection performance is further described by expanding the collection efficiency term in Equation 11.3:[32,33]

$$I_L = P_0 \mathrm{L} f_{ex}(\theta,\lambda)\, f_{em}(\theta,\lambda)\, f_{\mathrm{det}}(\theta,\lambda)\phi\varepsilon\mathrm{c} \tag{11.3}$$

where f_{ex}, f_{em} and f_{det} describe the optical characteristics (θ solid angle and λ wavelength) of the incident excitation and emission collection pathways and the detector, respectively. Consequently, the λ term accounts for the wavelength bandwidth and the transmission properties of the optical elements across the respective absorption and excitation spectral profiles of the analyte.

The response function and the associated analytical merits for absorption spectroscopic techniques (e.g., NIR, UV-vis and infrared) are determined by the optical path length, detector gain, signal averaging and spectral resolution. The LIF detection performance is also governed by these parameters but is also influenced by critical parameters associated with the excitation source (e.g., optical power, pulse rate, etc.) as previously discussed.[32,46]

There are numerous excitation sources available for LIF instruments. A xenon arc lamp is the most common light source within commercial LIF analyzers. While they offer uniform broad spectral coverage across the UV-vis range and sufficient uniform power output, they are low precision sources and do not offer sufficient real-time or dynamic optical power control. The white output also necessitates an excitation

filter(s), which attenuates the incident excitation source intensity between ~10 to 60%, thereby reducing the emission signal response. They are also less attractive sources for battery operation due to the high current power (25 A) required and are ignition sources which are often considered an explosion hazard without suitable precautions.[47]

In contrast, LEDs and LDs as LIF excitation sources offer many attractive attributes. First, these sources afford superior emission signal precision and accuracy due to their high spectral quality, narrow spectral characteristics (quasi-monochromatic) and reliable output. Due to the narrow bandwidth output these sources also often do not require an excitation filter, which affords maximum possible excitation intensity throughput for optimal emission response. When an excitation filter is required (i.e., small Stokes shift), the resulting attenuation may be offset by increasing the source optical power to provide the desired signal characteristics. Moreover, these sources afford real-time 'dynamic' optical power control[10,47] to exploit the optical power fluorescent advantage. This is advantageous as it provides a powerful approach in optimizing the response functions (i.e., SBR and SNR) and analytical merits (sensitivity, detection limits, etc.) that is not possible with conventional white light sources and in absorption spectroscopies. This can be accommodated by various means such as the diode drive current or selective diode activation within the array.[10,47] LEDs are also compact thereby affording arrays or spectral domain versatility without the need for additional high overhead optics (e.g., grating).[25] Finally, the significantly reduced power requirement of these sources is much more suitable for battery powered devices. In recent years these narrow band light sources have become more common within various instrumental devices as various commercial vendors now provide LEDs throughout the UV-vis range.[45,47]

Lasers have been omitted as a potential light source for process LIF instruments as they are inappropriate in many real-time applications due to several significant limitations. Consequently, before the advent of LEDs suitable for fluorescence analysis, lasers were the only excitation source affording exploitation of the optical power emission advantage. The benefits of lasers include intense and controllable optical power, monochromatic or high narrowband spectral quality and temporal and spatial coherence. Their drawbacks include more limited disparate laser energies for exciting a wide range of target fluorophores, greater power consumption, cost and more limited longevity, which are all problematic for real-time applications such as manufacturing.[32,47]

In a LIF photometer specific optical elements within the detection scheme isolate excitation and emission radiations as depicted in Figure 11.5. These optical elements are often centered about the respective excitation and emission maxima and encompass most of the spectral bandwidth while preserving the Stokes shift advantage. For small Stokes shifts, the optical elements can encompass a portion of the excitation and emission spectral bandwidths and may not necessarily be centered about the maxima. The optical elements are

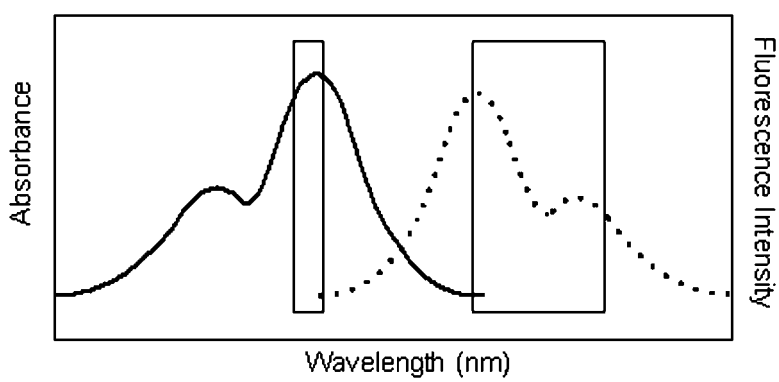


Figure 11.5 *LIF photometric excitation and emission isolation.*

most often optical filters such as band-pass and interference filters. For a broadband device containing a white excitation source (e.g., xenon flash lamp), these elements are optical filters coincident with the respective spectral characteristics of the target fluorophore. Secondary emissions of a white optical sources should also be considered when selecting appropriate filters.[33] Within a narrow band device, if the excitation source (LED or an equivalent source) provides sufficient monochromatic properties an emission optical filter is required. When the excitation LED bandwidth is too large or the fluor Stokes shift is too small an excitation filter is required to ensure a negligible background and optimal SBR and SNR. Compared to available excitation filters, a wider range of disparate LEDs are available for fluorescent analysis in providing target excitation at the desired absorption wavelength.[47]

11.4.2 LIF Instrument selection

As with any process analytical application, instrument selection is based on the required analytical merits (sensitivity, dynamic range, precision and accuracy, etc.), process and environmental conditions, integration complexities (mechanical and controls automation) and operational and maintenance requirements. Because of the wide disparity in analytical performance and functionalities among photometric and spectroscopic LIF process instruments, selection should be carefully weighed on the basis of the technical problem, instrumental cost, implementation complexities, ease of use, commercial and legacy maturity, level of vendor support and cost of ownership.

LIF filter photometers are simple to specify, operate and maintain. Moreover, their univariate output relative to spectral instruments is also a key benefit as it greatly simplifies data acquisition, plant controls automation integration, data management, data processing and calibration. Single-channel instruments with a single excitation–emission detection scheme are appropriate for monitoring a target fluorophore within a matrix containing limited or negligible emissive interference such as in water quality applications and in many pharmaceutical manufacturing applications with suitable formulations as described previously.[9,46,48] They are also robust as the detection scheme is based on solid-state optics. Multichannel filter-wheel based LIF photometers (e.g., the Delta BioVIEW) and spectrofluorometers are better suited for multiple fluorophore applications such as bioreactor fermentation monitoring.[12,38,39]

11.5 Luminescent Detection Risks

A discussion on real-time fluorescent monitoring would not be complete without a treatment on measurement risks, which include competing excited state processes and sample matrix effects. For the former several molecular-level processes compete with emission including dynamic quenching, static quenching, trivial quenching, excited state reactions and energy transfer as represented in Figure 11.6. Both static and dynamic quenching require molecular contact between the fluorophore and the quencher. Dynamic quenching refers to collisional encounters whereas static quenching is the formation of a stable complex. With the former, the quencher must diffuse to the fluorophore during the lifetime of the excited state and the interaction can induce nonradiative deactivation. The effect of quenching (dynamic or static) on fluorescence intensity is described by the Stern–Volmer equation (Equation 11.4):

$$\frac{F_0}{F} = 1 + k_q \tau_0 [Q] = 1 + K_D [Q] \tag{11.4}$$

where F_0 and F are the fluorescence intensities in the absence and presence of the quencher, respectively; k_q is the bimolecular quenching rate constant; τ_0 is the lifetime of the fluorophore in the absence of the quencher; $[Q]$ is the concentration of the quencher; and K_D is the Stern–Volmer quenching constant. In

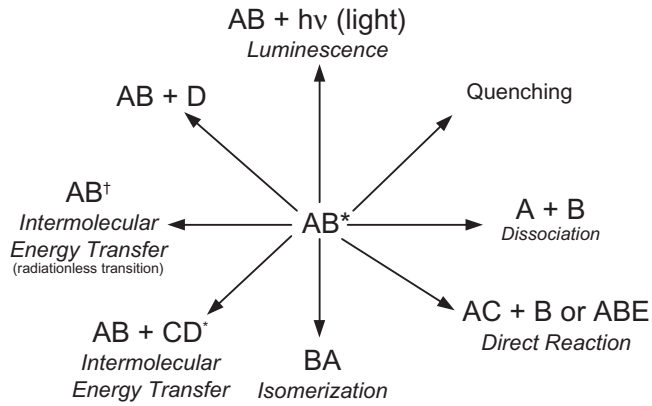


Figure 11.6 *Fates of excited species.*

general, static and dynamic quenching can be distinguished by their disparate dependence on temperature and viscosity. Dissolved oxygen and halogens such as chlorine ion are well known dynamic quenchers. Trivial quenching is also a common competing process. This involves reabsorption of emitted radiation by other sample constituents or the target fluorphore and is most prevalent at higher concentrations. Consequently, this type of quenching is also termed the inner-filter effect. Finally, long-range or Förster quenching occurs between molecules without collision. This type quenching interaction is considered a result of dipole-dipole coupling between a donor (excited fluorophore) and an acceptor (quencher).[32] With both dynamic and long-range quenching the quencher is promoted to an excited state. However, because the latter is a 'softer' interaction, the excited quencher can emit radiation (sensitized fluorescence) rather than relax via a nonradiative transition.

Luminescence is often much more sensitive to molecular dynamics than other optical techniques where temperature, viscosity, pH and solvent effects can have a significant influence on the emission response. Analyte degradation for light sensitive fluors and photobleaching for static measurements also influence the emission signal. Because of the wide variety of potential matrix effects, a thorough investigation should be conducted or the sample matrix well understood in terms of its potential impact on emission response. A complete discussion on the fate of the excited states and other measurement risk considerations can be found elsewhere.[3,4,32]

11.6 Process Analytical Technology Applications

LIF methods can be applied in-line, at-line and on-line for real-time monitoring as discussed throughout this chapter. In-line or *in situ* intrinsic LIF is by far more prevalent in real-time applications such as PAT as it is nondestructive and simple to deploy along with attractive analytical merits. In-line application can be accomplished by direct insertion *in situ* probes or flow cells. This type of monitoring is utilized for real-time analyte quantification monitoring and detection of process endpoints and faults.

At-line LIF methods are either based on intrinsic detection or extrinsic approaches. The former often involves static measurements which are prone to photobleaching and thermal effects. These problems can often be addressed by optimizing the excitation source output (i.e., optical power and pulse rate) or by sample agitation. Flow injection analysis or other autonomous sample prepreparation schemes are possible to facilitate various at-line extrinsic methods such as various selective fluoroimmunoassays.[3,24]

In some manufacturing process analysis applications the analyte requires sample preparation (dilution, derivatization, etc.) to afford a suitable analytical method. Derivatization, emission enhancement, and other extrinsic fluorescent approaches described previously are examples of such methods. On-line methods, in particular those requiring chemical reaction, are often reserved for unique cases where other PAT techniques (e.g., UV-vis, NIR, etc.) are insufficient (e.g., very low concentrations) and real-time process control is imperative. That is, there are several complexities to address with these types of on-line solutions to realize a robust process analysis method such as post reaction cleanup, filtering of reaction byproducts, etc. Nevertheless, real-time sample preparation is achieved via an on-line sample conditioning system.[23] These systems can also address harsh process stream conditions (flow, pressure, temperature, etc.) that are either not appropriate for the desired measurement accuracy or precision or the mechanical limitations of the in-line insertion probe or flow cell. This section summarizes some of the common LIF monitoring applications across various sectors.

11.6.1 Petrochemical, chemical and nuclear field applications

Biofouling, biomediation and water quality are the most widely known intrinsic LIF applications. For example, LIF sensing has been applied for the detection of leaks from subsea pipelines and risers.[49] This is significant as on a global basis there are thousands of miles of pipelines containing crude oil, hydraulic fluids and production chemicals. This application most often involves monitoring native fluorophore(s) such as polycyclic aromatic hydrocarbons (PAHs) or a fluorescent tracer that is injected within the pipeline to test for leaks post pipeline installation or maintenance. Surface and subsurface contamination of petrochemical fuels and products about storage units and during transfer is also a well suited application for intrinsic LIF.[50]

Within the chemical industry, LIF monitoring includes monitoring epoxy curing, pigments, optical brighteners and industrial waste water streams applications. Remote uranium monitoring is the most known application within the nuclear industry and often involves lifetime-based LIF process instruments.[41,51,52] A complete discussion of LIF monitoring applications across industries or sectors is outside the scope of this chapter. Nevertheless, various current and potential applications are summarized in Table 11.1. In addition, fluorescent sensing has and continues to be a significant area of academic research.[2]

11.6.2 Pharmaceutical PAT applications

While LIF monitoring is established in other manufacturing industries, it is relatively new within the pharmaceutical industry. The application of real-time intrinsic LIF for the manufacture of pharmaceuticals in general can be broken in to three areas: (i) classic (small molecule) final drug product manufacturing; (ii) biopharmaceutical product manufacturing; and (iii) factory operational applications.

The first demonstration of solid state fluorescence of API dates back to 1961,[65] while its in-line use for final drug product manufacturing was not demonstrated until recently.[48,66] While in its infancy as a process analytical technology for real-time monitoring and product parametric real-time release, the applications identified and in some instances demonstrated include: (i) blend endpoint API content uniformity detection;[48] (ii) segregation monitoring or API content at various process critical control points; and (iii) at-line tablet content uniformity determination.[66] The fundamentals of solid-state luminescence spectroscopy for pharmaceutical solids has been covered by Brittain.[4]

As already discussed, intrinsic LIF bioreactor monitoring is the most common application within biopharmaceutical manufacturing. Other extrinsic PAT LIF methods are also possible for biopharmaceuticals such as various fluorescence immunoassays[67] facilitated by flow injection analysis[24] or other real-time approaches. As PAT begins to emerge within biopharmaceuticals[68] the wider used of sensitive and precise in-line intrinsic and extrinsic approaches such as optodes[10] and at-line extrinsic methods is likely to occur.[67]

Table 11.1 *Representative industrial and field LIF monitoring applications*

Application	Sector or industry	Reference
Biofouling/biomediation (e.g., oil in water, soils, etc.)	Environmental/Marine science Petrochemical	50, 53, 54
Algal pigments	Lake/reservoir management Water treatment Aquaculture Marine science	55
Industrial waste water management	Industry	56, 55
Ammonium in aquatic environments	Environmental/Marine science	57, 56
Actinides analysis monitoring (e.g., uranium)	Nuclear, remediation	51, 52
Quinine, pigments, polymers, optical brighteners	Chemical, environmental	56, 58
Fuel compositions	Petrochemical	59
Crude palm oil	Food and agriculture	60
• Photosynthesis in phytoplankton and higher plants • Chlorophyll for aquatic monitoring	Marine science	61, 62
Chemical and biological warfare agents	Defense/security	25, 30
Tracer studies: water ways, water flows/dynamics, biofouling, industrial storage retention, water pollution control	All industries/sectors, environmental	53, 52
Bioreactor or fermentation monitoring	Bioprocessing: Biochemical/ biopharmaceuticals	12–14, 38, 39
Epoxy curing	Chemical	63
Chirality sensing	Biopharmaceuticals	64

That is, compared to conventional (small molecule) drug product manufacture, biopharmaceutical manufacturing necessitates an even higher state of PAT for robust process understanding and control. Not only are the processes more complex but the cost of waste is often much higher.

There are also several operational type methods appropriate for LIF monitoring. For example, effluent stream monitoring and control is an increasing concern to the pharmaceutical industry where LIF may also find increased utility.[69] That is, sending too much API to the local waste treatment plant can impact the environment while over or under feeding oxidizing agents to holding tanks can result in a costly out of control condition. LIF has also been used to monitor fluorescent APIs,[56] such as antibiotics[70] in a flowing waste stream to control feeding of chemical oxidants to waste stream holding tanks and verification in the output stream. Cleaning verification is yet another potential application of LIF alongside the current real-time methods such as UV-vis spectroscopy and total organic carbon analysis.

11.7 Conclusions

Fluorescent monitoring in its various forms can be applied to a plethora of real-time process and field analytical problems across various sectors. Given the superior analytical attributes and ease of use, steady state

intrinsic LIF sensing should continue to evolve as a standard PAT measurement technology alongside the other common process spectroscopic techniques such as IR, NIR, UV-vis and Raman. While time-domain and optode approaches are currently more limited, the combined selectivity and sensitivity of these techniques is very attractive, particularly for more complex real-time analytical problems, and thus may appear on a grander scale in the marketplace within the near future.[3] Finally, because fluorescence is particularly sensitive to molecular dynamics associate with a fluorophore and matrix effects, real-time intrinsic LIF may also find potential in detecting more complex molecular phenomena that may occur during product manufacture, such as conformational changes (e.g., protein folding) and guest–host interactions.

References

1. E.L. Wehry, Applications of Fluoresence and phosphoresence techniques in environmental analysis, in *Practical Fluorescence*, G.G. Guilbault (Ed), Marcel Dekker, New York, 1999.
2. G.G. Guilbault, General aspects of luminescence spectroscopy, in *Practical Fluorescence*, G.G. Guilbault (Ed), Marcel Dekker, New York, 1990.
3. J.R. Lakowicz, *Principles of Fluorescence Spectroscopy*, Springer, New York, 2006.
4. H.G. Brittain, Luminesence spectroscopy, in *Spectroscopy of Phamaceutical Solids*, H.G. Brittain (Ed), Taylor & Francis, New York, 2006.
5. C.N.R. Rao, *Ultra-Violet and Visible Spectroscopy*, Butterworths, London, 1961.
6. RxList [Online], Top 300 Prescriptions for 2005, http://www.rxlist.com/top200.htm accessed 20 Nov. 2009.
7. Perkin Elmer Inc. [Online], An Introduction to Fluorescence in Biological Analysis, available http://las.perkinelmer.com/content/manuals/gde_fluorescencebiologicalanalysis.pdf accessed 20 November 2009.
8. S. Udenfriend, *Fluorescence Assay in Biology and Medicine*, Academic Press, New York, 1969.
9. A. Castellan, R. Ruggiero, E. Frollini, L.A. Ramos and C. Chirat, *Holzforschung*, **61**, 504–508 (2007).
10. A.P. Demchenko, *Trends Biotechnol.*, **23**, 456–460 (2005).
11. C. Kaiser, T. Pototzki, A. Ellert and R. Luttmann, *Eng. Life Sci.*, **8**, 132–138 (2008).
12. J.-k. Li and A.E. Humphrey, *Biotechnol. Bioeng.*, **37**, 1043–1049 (1991).
13. S. Marose, C. Lindemann and T. Scheper, *Trends Biotechnol.*, **17**, 499–505 (1999).
14. R. Ulber, J.-G. Frerichs and S. Beutel, *Anal. Bioanal. Chem.*, **376**, 342–348 (2003).
15. K. Soroka, R.S. Vithanange, D.A. Phillips, B. Walker and P.K. Dasgupta, *Anal. Chem.*, **59**, 629–636 (1987).
16. P.K. Dasgupta, K. Soroka and R.S. Vithanage, *Anal. Chem.*, **59**, 629–636 (1987).
17. P. Jones, L. Ebdon and T. Williams, *Analyst*, **113**, 641 (1988).
18. T. Williams and N.W. Barnett, *Anal. Chim. Acta*, **259**, 19 (1992).
19. L.J. Cline-Love, M. Grayeski, J. Noroski and R. Weinberger, *Anal. Chim. Acta*, **170**, 3–12 (1985).
20. A. Coly and J. Aaron, *Anal. Chim. Acta*, **360**, 129–141 (1998).
21. O.J. Francis, G.P. Kirschenheuter, G.M. Ware, A.S. Carman and S.S. Kuan, *J. Assoc. Off. Anal. Chem.*, **71**, 725–728 (1988).
22. J.R. Lakowicz, *Topics in Fluorescence Spectroscopy; Probe Design and Chemical Sensing*, Plenum, New York, 1994.
23. K.G. Carr-Brion and J.R.P. Clarke, *Sampling Systems for Process Analyzers*, Butterworth-Heinemann, Oxford, 1996.
24. M. Trojanowicz, *Flow Injection Analysis Instrumentation and Applications*, World Scientific, Hackensack, 2000.
25. P. Vitta, N. Kurilčik, S. Juršenas, *et al.*, Fluorescence-lifetime identification of biological agents using deep ultraviolet light-emitting diodes, in *Proc. SPIE-Int. Soc. Opt. Eng.*, 2005.
26. A. Žukauskas, N. Kurilcik, P. Vitta, S. Juršenas, E. Bakiene and R. Gaska, Optimization of a UV light-emitting diode based fluorescence-phase sensor, in *Proc. SPIE-Int. Soc. Opt. Eng.*, 2006.
27. G. Boisde and A. Harmer, *Chemical and Biochemical Sensing With Optical Fibers and Waveguides*, Artech House Publishers, Norwood, 1996.
28. Polestar Technologies [Online], available http://www.polestartech.com/ accessed 20 November 2009.
29. J. Janata, *Principles of Chemical Sensors*, Plenum Press, New York, 1989.
30. M. Burnworth, S.J. Rowan and C. Weder, *Chem. Eur. J.*, **13**, 7828–7836 (2007).
31. Y. He, G. Wang, J. Cox and L. Geng, *Anal. Chem.*, **73**, 2302–2309 (2001).
32. J.D. Ingle and S.R. Crouch, *Spectrochemical Analysis*, Prentice Hall, Englewood Cliffs, 1988.

33. H.A. Strobel and W.R. Heineman, *Chemical Instrumentation: A Systematic Approach*, Wiley, New York, 1989.
34. H. Herman, Uv-Vis, fluorescence and chemiluminescence spectroscopies in chemical process analysis, in *Spectroscopy in Process Analysis*, J.M. Chalmers (Ed), CRC Press, Boca Raton, 1999.
35. J. Mukherjee, C. Lindermann and T. Scheper, *Appl. Microbiol. Biotechnol.*, **52**, 489–494 (1999).
36. J.I. Rhee and T.-H. Kang, *Process Biochem.*, **42**, 1124–1134 (2007).
37. B. Tartakovsky, M. Scheintuch, J.M. Hilmer and T. Scheper, *Biotechnol. Prog.*, **12**, 126–131 (1996).
38. S. Marose, C. Lindemann and T. Scheper, *Biotechnol. Prog.*, **14**, 63–74 (1998).
39. DELTA Light & Optics [Online], BioVIEW, available http://www.delta.dk/C1256ED60044AA49/0/3E5C8D00B7451877C1256F970039ED05 accessed 20 November 2009.
40. Carl Zeiss Inc. [Online], On-line Fluorescence Measurements in Biotechnology using a Fiber Optical Probe and a Light Sensitive CCD Spectrometer, available http://www.zeiss.de/c12567bb00549f37/Contents-Frame/b98a7597898099dfc125727d0047c6bc accessed 20 November 2009.
41. HORIBA Jobin Yvon Inc. [Online], Process Monitoring & Control, available http://www.horiba.com/us/en/scientific//?Ovly=1, accessed 10 September 2008.
42. Teledyne Analytical Instruments [Online], Series 6000 – Ultraviolet Photometer, available: http://www.teledyne-ai.com/products/6000.asp accessed 20 November 2009.
43. Custom Sensors and Technology [Online], available http://www.customsensors.com/ or http://www.teledyneanalytical.com/products/photox.asp accessed 10 September 2008.
44. Turner Designs Inc. [Online], available http://www.turnerdesigns.com/t2/instruments/instruments.html accessed 20 November 2009.
45. Sensor Electronic Technology, Inc. [Online], LEDs, available http://www.s-et.com/ accessed 20 November 2009
46. J.E. Dickens, M. Ponstingl, S. Christian, C. Dixon, J. Schroeder and S. Horvath, A Novel Light Induced Fluorescent (LIF) Sensor for Process and Field Applications, *IFPAC Conference*, Baltimore, MD, 2009.
47. A.E. Moe, S. Marx, N. Banani, M. Liu, B. Marquardt and D. M. Wilson, *Sens. Actuators B*, **111–112**, 230–241 (2005).
48. C.K. Lai, D. Holt, J.C. Leung, C.L. Cooney, G.K. Raju and P. Hansen, *AIChE J.*, **47**, 2618–2622 (2001).
49. D. McStay, J. Kerlin and R. Acheson, *J. Physics: Conference Series*, **76**, 1–6 (2007).
50. M. Lemke, R. Fernández-Trujillo and H.-G. Löhmannsröben, *Sens. Actuators B*, **5**, 61–69 (2005).
51. C. Moulin, P. Decambox and P. Mauchien, *J. Radioanal. Nucl. Chem.*, **226**, 135–138 (1997).
52. S.-M. Lee, J.-S. Shin and S.-W. Kang, *J. Korean Nucl. Soc.*, **33**, 166–174 (2001).
53. Turner Designs Inc. [Online], Fluorescent Tracer Studies, available http://www.turnerdesigns.com/t2/doc/appnotes/tracer_dye.html accessed 20 November 2009.
54. K.S. Booksh, A. R. Muroski and M. L. Myrick, *Anal. Chem.*, **68**, 3539–3544 (1996).
55. K.S. Baker, *Limnol. Oceanogr.*, **28**, 1037–1040 (1983).
56. Custom Sensors and Technology, Fluorescence Methods, available http://www.customsensors.com/ accessed 20 Nov. 2009.
57. Turner Designs Inc. [Online], Ammonium Concentration Determination, available http://www.turnerdesigns.com/t2/doc/appnotes/S-0025.pdf accessed 20 November 2009.
58. Turner Designs, Inc. [Online], Optical Brighteners, available http://www.turnerdesigns.com/t2/doc/appnotes/S-0080.pdf accessed 20 November 2009.
59. R. Karpicz, A. Dementjev, V. Gulbinas, Z. Kuprionis, S. Pakalnis, R. Westphal and R. Reuter, Laser fluorosensor for oil spot detection, *Lith. J. Physics*, **45**, 213–218 (2005).
60. Y.A. Tan and C.L. Chong, *J. Sci. Food Agric.*, **67**, 375–379 (1995).
61. E. Fuchs, R.C. Zimmerman and J.S. Jaffe, *Plankton Res.*, **24**, 1221–1229 (2002).
62. J.J. Cullen and E.H. Renger, *Mar. Biology*, **53**, 13–20 (1979).
63. W.B. Dang and N.H. Sung, *Polym. Eng. Sci.*, **34**, 707–715 (1999).
64. Z. Lin, M. Wu and O.S. Wolfbeis, *Chirality*, **17**, 464–469 (2005).
65. W.F. Head, Jr, *J. Pharm. Sci.*, **15**, 1041–1044 (1961).
66. C.K. Lai, A. Zahari, B. Miller, W. E. Katstra, M. J. Cima and C. L. Cooney, *AAPS PharmSciTech*, **5**, 1–10 (2004).
67. B.J. Del-Tito, *J. Ind. Microbiol.*, **8**, 121–126 (1991).
68. Pharmaceutical Formulation and Quality [Online], Biopharma Says Bye Bye to Blackmagic, available http://www.pharmaquality.com/ accessed 20 November 2009.
69. US EPA [Online], Pharmaceuticals, available http://www.epa.gov/epawaste/hazard/wastetypes/universal/pharm.htm accessed 20 November 2009.
70. R. Hirsch, T. Ternes, K. Haberer and K.-L. Kratz, *Sci. Total Environ.*, **225**, (1999).

12

Chemometrics in Process Analytical Technology (PAT)

Charles E. Miller

Merck and Company, West Point, PA

12.1 Introduction

12.1.1 What is chemometrics?

The term 'chemometrics' was coined several decades ago to describe a new way of analyzing chemical data, in which elements of both *statistical* and *chemical* thinking are combined [1]. Since then, chemometrics has developed into a legitimate technical field of its own, and is rapidly growing in popularity within a wide range of chemical disciplines.

There are probably as many definitions of chemometrics as there are those who claim to practice it. However, there appear to be three elements that are consistently used in historical applications of chemometrics:

1. empirical modeling,
2. multivariate modeling, and
3. chemical data.

With this in mind, I ask the reader to accept my humble definition of 'chemometrics': *the application of multivariate, empirical modeling methods to chemical data* [2].

The empirical modeling element indicates an increased emphasis on *data-driven* rather than *theory-driven* modeling of data. This is not to say that appropriate theories and prior chemical knowledge are ignored in chemometrics, only that they are not relied upon completely to model the data. In fact, when one builds a 'chemometric' calibration model for a process analyzer, one is likely using prior knowledge or theoretical relations of some sort regarding the chemistry of the sample or the physics of the analyzer. One example

Process Analytical Technology 2e Edited by Katherine Bakeev
© 2010 John Wiley & Sons, Ltd.

of this in PAT applications is the use of the Beer–Lambert Law relating spectral intensity and concentration.

The multivariate element of chemometrics indicates that more than one response variable of the analyzer is used to build a model. This is often done out of necessity, because no single response variable from the analyzer has sufficient selectivity to monitor a specific property without experiencing interferences from other properties.

The combination of empirical and multivariate modeling elements makes chemometrics both very powerful and very dangerous. The power of chemometrics is that it can be used to model systems that are both unknown and complex. Furthermore, these models are not restricted by theoretical constraints – which can be a big advantage if large deviations from theoretical behavior are known to be present in your system.

However, there are prices to pay for the advantages above. Most empirical modeling techniques need to be fed large amounts of *good* data. Furthermore, empirical models can only be safely applied to conditions that were represented in the data used to build the model (i.e., extrapolation of such models is very dangerous). In addition, the availability of multiple response variables for building a model results in the temptation to overfit models, in order to obtain artificially optimistic results. Finally, multivariate models are usually much more difficult to explain to others, especially those not well versed in math and statistics.

12.1.2 Some history

The use of multivariate statistics to decipher multivariate data dates back to the 1930s (Hotelling) [3], and most early applications of these tools were in the fields of psychology, sociology and economics – fields where large amounts of highly unspecific data are the rule, rather than the exception.

Despite its successes in other fields, chemometrics was not introduced in analytical chemistry until the late 1960s and 1970s. Somewhat surprisingly, the first appearance of chemometrics for analytical chemistry in the literature came from the food industry, rather than from academia. The work of Norris and Hruschka at the USDA starting in the 1960s [4–7] demonstrated the practical feasibility of multiple-wavelength near infrared (NIR) calibrations for rapid nondestructive analyses of foods and agricultural materials. With these practical successes as a background, a few intrepid researchers in the academic realm [8–14], pushed the integration of existing tools with this new field, while also developing new and unique chemometrics tools.

Through the 1980s and 1990s the development of chemometrics for analytical chemistry was largely driven by the increasing popularity and utility of nonspecific NIR spectroscopy technology. Since the work of Norris, it was quickly realized that the great potential of NIR technology for a wide range of industrial applications could not be 'unlocked' unless chemometrics is used to provide sufficient selectivity. At the same time, increased efficiency, and accessibility of computer technology allowed the power of chemometrics to be accessible to the vast majority of scientists and researchers. These forces combined to generate a boom in the publication of chemometrics applied to analytical chemistry starting in the 1990s.

Unfortunately, accessibility does not necessarily lead to effective deployment in practical applications. Despite the voluminous publications on chemometrics applications, there was a noticeable time lag between these and the mass deployment of effective process analytical chemometrics applications in the field. I propose that this time lag was caused by several factors:

- The 'overselling' and lack of user understanding of chemometrics methods during the 'boom' period, which ultimately led to misuse of the methods and failed applications,
- The lack of adequate software tools to develop and *safely* implement chemometric models in a process analytical environment, and
- The lack of qualified resources to develop and maintain chemometrics-based analytical methods in the field.

At the same time, there are many effective chemometrics applications in PAT across a wide range of industries, although relatively few of them are published (for example, see references [15–19]). Nonetheless, these provide the driving force for analytical scientists and engineers to push for new applications in industry.

12.1.3 Some philosophy

Chemometrics is a rather unique technology – for two main reasons. Most technologies are invented and developed at academic or R&D institutions, 'handed over' to industry, and then adapted for practical use by engineers and scientists in industry. Chemometrics technology is very different, in that industry played a very large role in its development, dating back to its infancy. This resulted because many chemometric tools are a product of the desire for *practical* applicability, rather than theoretical rigor. This often leads many in academic institutions to shun the technology as lacking sufficient fundamental basis or ignoring established rules of statistical analysis, even though it has proven itself in many industrial applications.

Second, the emphasis on empirical modeling leads to chemometrics being a highly 'interfacial' discipline, in that specific tools are often developed with specific applications already in mind. For example, specific chemometric tools have been developed to align retention time axes in chromatograms [20] and to preprocess diffuse reflectance data [21]. In contrast, other disciplines, such as statistics, are associated with well-defined 'stand-alone' tools (ANOVA, *t*-test, etc.) that can be applied to a wide array of different applications. One consequence of this interfacial property of chemometrics is that one must often sift through a very large 'toolbox' of application-specific tools in order to find one that suits a particular application.

Martens and Naes [1] provide an interesting perspective on chemometrics as 'a combination of chemical thinking and statistical thinking'. In this context, it is suggested that statisticians are well aware of the limitations and capabilities of statistical tools, but confined to a 'rigid' structure of these tools – not allowing chemical knowledge to aid in the experimental design, data modeling, or the interpretation of results. Conversely, chemists are often not aware of the theoretical limitations on experimental design and statistical analysis techniques, but have the chemical knowledge to optimize experimental designs and detect anomalies in the analyses. Both chemical thinking and statistical thinking are required to generate optimal results.

Finally, the highly-empirical and practical nature of chemometrics leads many outside the field to conclude that chemometrics practitioners do not understand the modeling tools they use, or the models that they develop- thus potentially leading to disastrous results. Although this perception might be accurate for many users, it need not be true for all users. This leads to the mission of this chapter**:** *to enable users to better understand, and thus more effectively implement, chemometrics in process analytical applications*.

12.1.4 Chemometrics in analytical chemistry?

Chemometrics tools can be used for a wide variety of tasks, including experimental design, exploratory data analysis, and the development of predictive models. In the context of analytical chemistry, however, chemometrics has been shown to be most effective for two general functions:

1. *Instrument specialization:* Multivariate calibration models are built in order to *provide selectivity* for a multivariate analytical instrument, or
2. *Information extraction:* Chemometrics tools are used to *'unlock' hidden information* already present in information-rich multivariate analytical instruments

Regarding point 1, the need to improve specificity of an analyzer depends on both the analytical technology and the application. For example, chemometrics is required for most NIR spectroscopy applications, because

the information in NIR spectra is generally nonspecific for most applications. In contrast, chemometrics might not be critical for most ICP atomic emission or mass spectrometry applications, because these techniques provide sufficient selectivity for most applications.

Regarding point 2, the information extraction function of chemometrics is a very valuable one that is often overlooked, especially in the industrial world. It will be mentioned later in this chapter that this function can be used *concurrently* with the instrument specialization function, rather than relying upon additional exploratory data analysis.

12.1.5 Chemometrics in process analytical chemistry?

Process analytical (PAC, PAT) applications differ from laboratory analytical applications in several ways:

- the analyzer hardware and software are in a more 'hostile' environment
- analyzer maintenance and operations personnel in the analyzer area have not undergone lab technician training, and
- the purpose of the analyzer is usually for rapid process control rather than troubleshooting.

As a result, process analytical applications tend to focus more on automation, minimization of maintenance, and long-term reliability. In this more pragmatic context, the instrument specialization function of chemometrics is used even more exclusively than the information extraction function.

An additional feature of chemometrics that is appealing to process analytical applications is the use of *qualitative* models to detect and characterize faults in the analyzer system (calibration, instrument, sampling interface, and sampling systems), sample chemistry, and process dynamics. Such faults can be used to trigger preventive maintenance, and to troubleshoot- thus supporting the long-term reliability of the analyzer system. Specific examples of such fault detection are given in references [15–16].

12.2 Foundations of Chemometrics

All technologies are built upon a foundation of previously-existing technologies, principles and protocols. There are several of these that were critical to the development of chemometrics, and the ones that are most relevant to PAT applications will be discussed in this section.

12.2.1 Notation

In the upcoming discussions, I will use the following convention in expressing numerical values:

scalars, which consist of a single value, are expressed as *lower case*, *italic* letters

$$\text{e.g., } a = [0.5]$$

vectors, which consist of a one-dimensional series of values, are expressed as *lower-case*, *bold-face* letters

$$\text{e.g., } \mathbf{a} = [5\ 4\ 1\ 2]$$

matrices, which consist of a two-dimensional array of values, are expressed as *upper-case*, *bold-face* letters

$$\text{e.g., } \mathbf{A} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \\ 13 & 14 & 15 & 16 \end{bmatrix}$$

In addition, specific letters are used to denote specific quantities that are common for the chemometrics tools discussed in this chapter:

1. The letter '*x*' is reserved for values of *independent variables*. In PAT applications, these variables most often represent spectral intensities at different wavelengths, chromatographic intensities at different retention times, or similar analytical profiles.
2. The letter '*y*' is reserved for values of *dependent variables.* In PAT applications, these variables most often represent constituent concentrations or properties of interest that need to be predicted from the analyzer measurement.
3. The letter '*b*' is reserved for the *regression equation coefficients* in a regression problem
4. The letter '*e*' is reserved for the residuals of the independent variables in a modeling problem, also called the '*x* residuals'.
5. The letter '*f*' is reserved for the residuals of the dependent variables in a regression problem, also called the '*y* residuals'.

Finally, the placement of a caret symbol '^' on top of a letter indicates that the letter represents an *estimated* quantity, rather than a measured or theoretical quantity. For example:

$\hat{\mathbf{X}}$ indicates a matrix of independent variables that have been *estimated by a model*, where
$\mathbf{X}$ indicates a matrix of independent variables that have been *actually measured* on an analytical device.

Finally, some special notation for matrix operations are used throughout the text:
superscript 't' indicates the transpose of a matrix:

$$\text{e.g., if } \mathbf{D} = \begin{bmatrix} 1 & 5 \\ 2 & 6 \\ 3 & 7 \\ 4 & 8 \end{bmatrix}, \text{ then } \mathbf{D}^{\mathbf{t}} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \end{bmatrix}$$

superscript '–1' indicates the inverse of a matrix (note: the matrix must be square and invertible!):

$$\text{e.g., if } \mathbf{M} = \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix}, \text{ then } \mathbf{M}^{-1} = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.25 \end{bmatrix}$$

Two matrices or vectors placed adjacently to one another indicates matrix multiplication. (Note: the dimensionality of the matrices or vectors must match appropriately for multiplication to be possible):

$$\text{e.g., } \mathbf{DM} = \begin{bmatrix} 1 & 5 \\ 2 & 6 \\ 3 & 7 \\ 4 & 8 \end{bmatrix} * \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix} \text{(a } 4 \times 2 \text{ matrix multiplied by a } 2 \times 2 \text{ matrix)}$$

12.2.2 Some basic statistics

Despite the conflicting philosophies between statisticians and chemometricians, it is an unavoidable fact that statistics provides a strong foundation for chemometrics. One could easily argue that, without classical statistics, there could be no chemometrics.

One underlying principle of classical statistics is that any observation in nature has an *uncertainty* associated with it. One extension of this principle is that multiple observations of the same object will result in a *distribution* of values. One common graphical representation of a distribution of values is the histogram, where the frequency of occurrence of a value is plotted versus the value. Many statistical tools are based on a specific type of distribution, namely the Gaussian, or Normal distribution, which has the following mathematical form:

$$f(x) = D\mathrm{e}^{-\frac{(x-\bar{x})^2}{2\sigma^2}} \tag{12.1}$$

where $\bar{x}$ is the mean of the values, and σ is the standard deviation of the values. The mean is the best single value approximation of all of the values, and the variance is a measure of the extent to which the values vary. The mean and variance as defined by the Gaussian distribution can be directly calculated from the multiple values:

$$\bar{x} = \frac{\sum_{i=1}^{n} x}{N} \tag{12.2}$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} (x-\bar{x})^2}{N}} \tag{12.3}$$

Where N is the number of observations.

Although there are many useful statistical tools, there are two that have particular relevance to chemometrics: the t-test and the f-test [22,23]. The t-test is used to determine whether a single value is statistically different from the rest of the values in a series. Given a series of values, and the number of values in the series, the t-value for a specific value is given by the following equation:

$$t = \frac{|\bar{x} - x|}{\left(\sigma / \sqrt{N}\right)} \tag{12.4}$$

The t value is the number of standard deviations that the single value differs from the mean value. This t value is then compared to the critical t value obtained from a t-table, given a desired statistical confidence (i.e., 90%, 95%, or 99% confidence) and the number of degrees of freedom (typically $N-1$), to assess whether the value is statistically different from the other values in the series. In chemometrics, the t test can be useful for evaluating *outliers* in data sets.

The f-test is similar to the t-test, but is used to determine whether two different standard deviations are statistically different. In the context of chemometrics, the f-test is often used to compare distributions in regression model errors in order to assess whether one model is significantly different than another. The f-statistic is simply the ratio of the squares of two standard deviations obtained from two different distributions:

$$f = \frac{\sigma_1^2}{\sigma_2^2} \tag{12.5}$$

As the f statistic deviates further from one, there is a greater probability that the two standard deviations are different. To quantitate this probability, one needs to consult an f-table that provides the critical values of f as a function of the degrees of freedom for the two distributions (typically $N - 1$ for each of the data series), and the desired statistical confidence (90%, 95%, 99%, etc.).

12.2.3 Linear regression

Linear regression [1,22,23] is typically used to build a linear model that relates a single independent variable (x) to a single dependent variable (y) using a series of (x,y) observations. For example, one could make a set of observations of the intensity of a single peak in a Fourier transform infrared (FTIR) spectrum (x), and a matching set of observations of an analyte concentration obtained from an off-line wet-chemistry method (y) for the same set of N samples (see Table 12.1). Given this data, one can use linear regression to develop a predictive model that can be used to estimate the analyte concentration from the intensity of the specified FTIR peak, for any unknown sample.

The model for linear regression is given below.

$$\mathbf{y} = b\mathbf{x} + b_0\mathbf{1} + \mathbf{f} \tag{12.6}$$

where $\mathbf{y}$ is a vector of measured independent variables, $\mathbf{x}$ is a matching vector of measured dependent variables, $\mathbf{1}$ is a vector of ones, and $\mathbf{f}$ is a vector containing the residuals of the linear regression model. All of the vectors mentioned above have N elements. This equation can be simplified to the following form:

$$\mathbf{y} = \mathbf{X}_{\text{aug}}\mathbf{b} + \mathbf{f} \tag{12.7}$$

where

$$\mathbf{b} = \begin{bmatrix} b \\ b_0 \end{bmatrix}, \text{ and}$$

$$\mathbf{X}_{\text{aug}} = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \dots & \dots \\ x_N & 1 \end{bmatrix}$$

Table 12.1 *Six sets of observations*

Analyzer measurement	Analyte concentration
0.56	0.00
0.87	0.98
1.05	2.10
1.31	3.02
1.70	4.99

Note: Each set consists of an analyzer measurement and its matching known analyte concentration obtained from an off-line wet-chemistry method; used for illustrating linear regression.

As the name suggests, the residuals (**f**) contain the variation in y that cannot be explained by the model.

Given the matching sets of measured data, **x** and **y**, it is now possible to estimate the model regression coefficients **b**. Assuming that the model errors (values in **f**) are Gaussian-distributed, it can be proven that the value of **b** that minimizes the sum of squares of the model errors is determined using the least squares method:

$$\hat{\mathbf{b}} = \left(\mathbf{X}_{\mathbf{aug}}^{\mathrm{t}}\mathbf{X}_{\mathbf{aug}}\right)^{-1}\mathbf{X}_{\mathbf{aug}}^{\mathrm{t}}\mathbf{y} \tag{12.8}$$

Graphically, the elements of **b** refer to the slope (b) and intercept (b_0) of the line of best fit through the observed data points. For our example data in Table 12.1, a graph of the linear regression fit is illustrated in Figure 12.1. Once the **b** coefficients are estimated, the model error (**f**) can also be estimated:

$$\hat{\mathbf{f}} = (\mathbf{y} - \hat{\mathbf{y}}) = \left(\mathbf{y} - \mathbf{X}_{\mathbf{aug}}\hat{\mathbf{b}}\right) \tag{12.9}$$

There are several properties of linear regression that should be noted. First, it is assumed that the model errors are normally distributed. Second, the relationship between the x and y variables is assumed to be linear. In analytical chemistry, the first assumption is generally a reasonable one. However, the second assumption might not be sufficiently accurate in many situations, especially if a strong nonlinear relationship is suspected between x and y. There are some nonlinear remedies to deal with such situations, and these will be discussed later.

Another assumption, which becomes apparent when one carefully examines the model (Equation 12.7), is that all of the model error (**f**) is in the independent variable (**y**). There is no provision in the model for errors in the dependent variable (**x**). In process analytical chemistry, this is equivalent to saying that there is *error only in the reference method, and no error in the on-line analyzer responses*. Although this is obviously not true, practical experience over the years has shown that linear regression can be very effective in analytical chemistry applications.

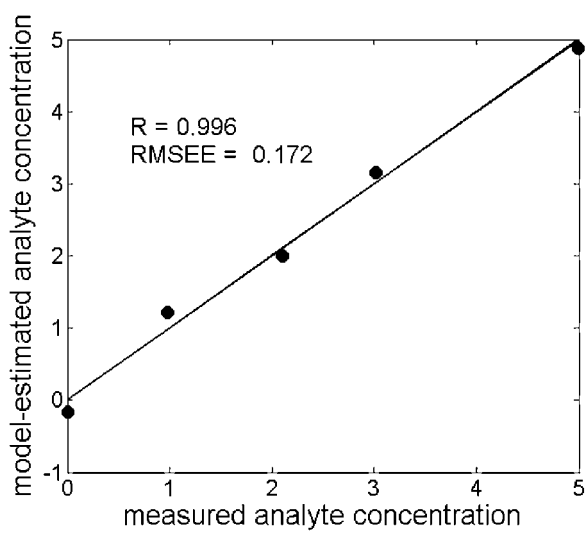


Figure 12.1 *Graph of the data points represented in Table 12.1, along with the line of best linear fit. The correlation coefficient (*r*) and the RMSEE for the model fit are listed on the figure.*

There are several figures of merit that can be used to describe the quality of a linear regression model. One very common figure of merit is the correlation coefficient, r, which is defined as:

$$r = \frac{\left(\sum_{i=1}^{N}\left(\hat{y}_i - \hat{\bar{y}}\right) * \left(y_i - \bar{y}\right)\right)}{\sqrt{\left(\sum_{i=1}^{N}\left(\hat{y}_i - \hat{\bar{y}}\right) * \sum_{i=1}^{N}\left(y_i - \bar{y}\right)\right)}} \tag{12.10}$$

where y_i $\hat{y}_i$ is $\bar{y}$ the mean of the *known y* values and $\hat{\bar{y}}$ is the mean of the *model-estimated y* values. A model with a perfect fit will yield a correlation coefficient of 1.0.

Another figure of merit for the fit of a linear regression model is the root mean square error of estimate (RMSEE), defined as:

$$\text{RMSEE} = \sqrt{\frac{\sum_{i=1}^{N}\left(y_i - \hat{y}_i\right)^2}{N-1}} \tag{12.11}$$

A model with a perfect fit will yield an RMSEE of 0. The r and RMSEE of the linear regression model built from the example data are shown in Figure 12.1.

Comparing the two figures of merit, the correlation coefficient has the advantage of taking into account the *range* of y values used in the regression. It has the disadvantage of being a unitless quantity, which might not provide sufficient meaning to the customer. An advantage of the RMSEE is that it is in the same units as the y variable, and therefore provides a good absolute assessment of the model fit error. However, if one were interested in the model error relative to the range, it would be more relevant to consider the RMSEE as a percentage of the range of y values, or just the correlation coefficient.

12.2.4 Multiple linear regression

An extension of linear regression, multiple linear regression (MLR) involves the use of *more than one independent variable*. Such a technique can be very effective if it is suspected that the information contained in a single dependent variable (x) is insufficient to explain the variation in the independent variable (y). In PAT, such a situation often occurs because of the inability to find a single analyzer response variable that is affected solely by the property of interest, without interference from other properties or effects. In such cases, it is necessary to use more than one response variable from the analyzer to build an effective calibration model, so that the effects of such interferences can be compensated.

The multiple linear regression model is simply an extension of the linear regression model (Equation 12.7), and is given below:

$$\mathbf{y} = \mathbf{bX} + \mathbf{f} \tag{12.12}$$

The difference here is that **X** is a matrix that contains responses from M (>1) different x variables, and **b** contains M regression coefficients for each of the x variables. As for linear regression, the coefficients for MLR (**b**) are determined using the least-squares method:

$$\hat{\mathbf{b}} = \left(\mathbf{X}^{\mathbf{t}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathbf{t}}\mathbf{y} \tag{12.13}$$

At this point, it is important to note two limitations of the MLR method:

- The x variable matrix ($\mathbf{X}$) has the dimensionality of N by M. As a result, *the number of x variables (M) cannot exceed the number of samples (N)*, otherwise the matrix inversion operation $(\mathbf{X}^t\mathbf{X})^{-1}$ in Equation 12.13 cannot be done.
- Second, if any two of the x variables are correlated to one another, then the same matrix inversion cannot be done.

In real applications, where there is noise in the data, it is rare to have two x variables exactly correlated to one another. However, a high degree of correlation between any two x variables leads to an unstable matrix inversion, which results in a large amount of noise being introduced to the regression coefficients. Therefore, one must be very wary of *intercorrelation* between x variables when using the MLR method.

12.2.5 Principal components analysis (PCA)

12.2.5.1 The Need for Data Compression

It was mentioned earlier that empirical multivariate modeling often requires a very large amount of data. These data can contain a very large number of samples (N), a very large number of variables (M) per sample, or both. In the case of PAT, where spectroscopic analytical methods are often used, the number of variables collected per process sample can range from the hundreds to the thousands!

The presence of such a large number of variables presents both logistical and mathematical issues when working with multivariate data. Data compression is the process of reducing data into a representation that uses fewer variables, yet still expresses most of its information. From a logistical standpoint, a compressed representation of the data uses fewer variables, requires less data storage, and requires fewer resources to transmit via hardwired or wireless communication. A mathematical advantage of data compression is that it can be used to reduce unwanted, redundant or irrelevant information in the data, thus enabling subsequent modeling techniques to perform more efficiently.

There are many different types of data compression that can be useful for a wide range of technical fields. For spectral and chromatographic data, which are quite common in PAT applications, and usually have the property of a continuous variable axis scale (wavelength and retention time, respectively), there are several compression methods that can be useful. For example, Fourier transform compression [1,24] can be used to transform such data from a representation involving many wavelength variables into representation involving a smaller number of frequency components. Wavelet transform compression [25,27] does a similar transformation, except that the new representation retains both frequency and some original (wavelength) scale information. Both of these compression methods have been shown to be effective for some analytical problems. However, there is one specific type of data compression that serves as one of the 'foundations' of chemometric methods – PCA.

12.2.5.2 The Details of PCA

PCA is a data compression method that reduces a set of data collected on M variables over N samples to a simpler representation that uses a much fewer number ($A \ll M$) of 'compressed variables', called principal components (or PCs). The mathematical model for the PCA method is provided below:

$$\mathbf{X} = \mathbf{T}\mathbf{P}^t + \mathbf{E} \tag{12.14}$$

where $\mathbf{T}$ is an N by A matrix containing the *scores* of the A principal components, $\mathbf{P}$ is an M by A matrix containing the *loadings* of the principal components, and $\mathbf{E}$ is a $N \times M$ matrix of x residuals. The scores can

be thought of as the 'intensities' of each of the A new 'compressed' variables for all of the N samples. The loadings can be thought of as the 'definitions' of the A 'compressed' variables in terms of the M original variables. In the most commonly used algorithm for PCA [1,28], the loadings for each PC are both normalized and orthogonal to one another, and the scores for each PC are orthogonal to one another. Conceptually, *orthogonality* can be described as two vectors being 'completely uncorrelated' to one another. Mathematically speaking, this translates to the following equations:

$$\mathbf{P}^t\mathbf{P} = \mathbf{I}, \text{ and} \tag{12.15}$$

$$\frac{\mathbf{T}^t\mathbf{T}}{(N-1)} = \text{diag}(\lambda) \tag{12.16}$$

where $\mathbf{I}$ is the identity matrix (a diagonal matrix where each element on the diagonal is equal to 1), and diag(λ) is a square diagonal matrix of dimensionality A, where each element contains the *eigenvalue* of principal component A. The eigenvalue of each PC reflects the amount of variance in the data that each PC explains. In the context of regression (MLR), the orthogonality of PCs is particularly important, because one can *avoid the intercorrelation problem* in MLR by using PCs instead of original x variables.

The most commonly used PCA algorithm involves sequential determination of each principal component (or, each matched pair of score and loading vectors) via an iterative least-squares process, followed by subtraction of that component's contribution to the data. Each sequential PC is determined such that it *explains the most remaining variance in the x data*. This process continues until the number of PCs (A) equals the *minimum* of the number of original variables (M) and the number of samples (N), at which time 100% of the variance in the data is explained. However, data compression does not really occur unless the user chooses a number of PCs that is much lower than the number of original variables ($A \ll M$). This necessarily involves ignoring a small fraction of the variation in the original x data, which is stored in the model x residuals, $\mathbf{E}$. If the calculated PCA scores for the first A PCs are contained in $\hat{\mathbf{T}}$, and the calculated PCA loadings for the first A PCs are contained in $\hat{\mathbf{P}}$, then the residuals can be estimated by the following:

$$\hat{\mathbf{E}} = \mathbf{X} - \hat{\mathbf{T}}\hat{\mathbf{P}}^t \tag{12.17}$$

In practice, the choice of an optimal number of PCs to retain in the PCA model (A) is a rather subjective process, which balances the need to explain as much of the original data as possible with the need to avoid incorporating too much noise into the PCA model (overfitting). The issue of overfitting is discussed later in Section 12.4.

An example of PCA compression is made using the classic Fisher's Irises data set [29], illustrated below. Table 12.2 lists *part* of a data set containing four descriptors (x variables) for each of 150 different iris samples. Note that these iris samples fall into three known classes: Setosa, Verginica, and Versicolor. From observation of Table 12.2 only, it is rather difficult to determine whether the four x variables can be useful for discriminating between the three known classes. The raw data in Table 12.2 are first preprocessed by autoscaling (preprocessing is discussed in Section 12.3.1), so that the data for each of the four variables have a mean of 0 and a standard deviation of 1. When PCA is applied to the autoscaled data, it is found that the first two PCs explain almost 96% of the variation in the original data (see Table 12.3). This result suggests that the four original x variables, due to correlations to one another, really explain only two truly independent effects in the data. These two effects can be expressed more efficiently using the first two PC scores (see the scatter plot in Figure 12.2). Note that this PC scores scatter plot enables one to make a better assessment regarding the ability of the four original measurements to discriminate between the three classes.

Table 12.2 *Part of the Fisher Iris dataset, used to illustrate PCA*

sample	species	petal width	petal length	sepal width	sepal length
1	Setosa	2	14	33	50
2	Verginica	24	56	31	67
3	Verginica	23	51	31	69
4	Setosa	2	10	36	46
5	Verginica	20	52	30	65
6	Verginica	19	51	27	58
7	Versicolor	13	45	28	57
8	Versicolor	16	47	33	63
9	Verginica	17	45	25	49
10	Versicolor	14	47	32	70

Note: This dataset contains four descriptors (*x* variables) for each of 150 different iris samples that can be in one of three known classes: Setosa, Verginica, and Versicolor.

Table 12.3 *Results of PCA on the Fisher Iris dataset: the percentage of X data variance explained by each of the first 4 PCs*

PC number	% X variance explained by PC	Cumulative % X-variance explained
1	72.8	72.8
2	22.9	95.7
3	3.7	99.4
4	0.6	100

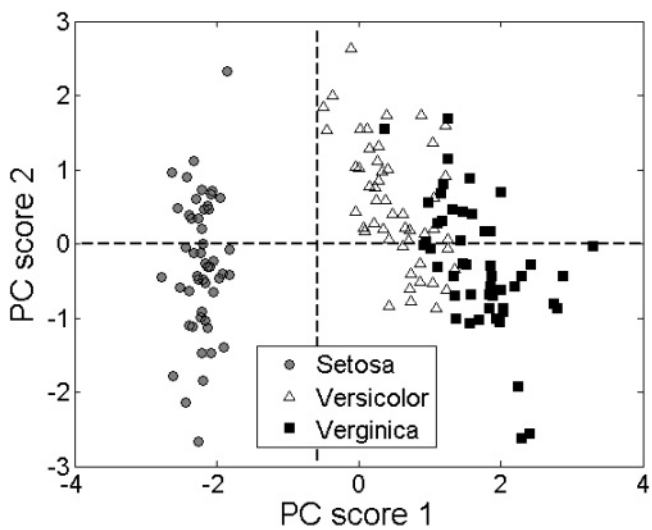


Figure 12.2 *Scatter plot of the first two PC scores obtained from PCA analysis of the Fischer iris data set.*

Figure 12.3 provides a scatter plot of the first two PC loadings, which can be used to roughly interpret the two new 'compressed variables' in terms of the original four variables. In this case, it appears that the first PC is a descriptor between the sepal width and the other three *x* variables, while the second PC is a descriptor of the two sepal measurements only. The plot also shows that the petal width and petal length

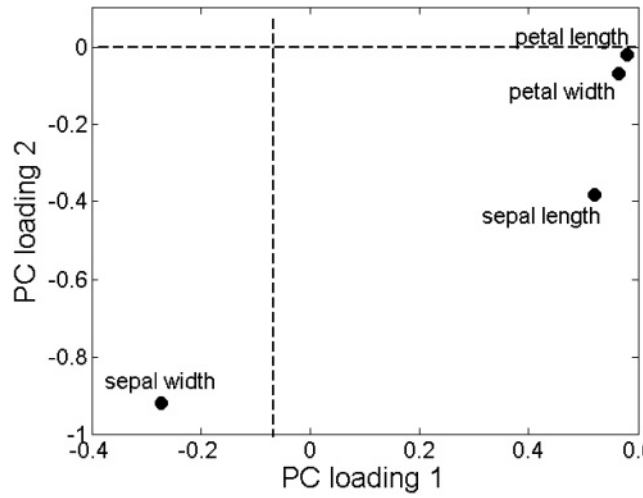


Figure 12.3 *Scatter plot of the first 2 PC loadings obtained from PCA analysis of the Fischer iris data set.*

are highly correlated to one another, for the 150 irises used in this study. Further discussion on the interpretation of PC scores and loadings is found in Section 12.3.4.

As the above example illustrates, PCA can be an effective *exploratory* tool. However, it can also be used as a *predictive* tool in a PAT context. A good example of this usage is the case where one wishes to determine whether newly collected analyzer responses are normal or abnormal with respect to previously collected responses. An efficient way to perform such analyses would be to construct a PCA model using the previously collected responses, and apply this model to any analyzer response ($\mathbf{x}_p$) generated by a subsequently-collected sample. Such PCA model *application* involves first a multiplication of the response vector with the PCA loadings ($\mathbf{P}$) to generate a set of PCA scores for the newly collected response:

$$\hat{\mathbf{t}}_p = \mathbf{x}_p\mathbf{P} \tag{12.18}$$

In linear algebra, this operation is called a *projection* of the response onto the space of the PCA model. Once this is done, the PCA-model-estimated response for the sample ($\hat{\mathbf{x}}_p$) can be calculated:

$$\hat{\mathbf{x}}_p = \hat{\mathbf{t}}_p\mathbf{P}^t \tag{12.19}$$

This then allows calculation of the *residual* for the new sample ($\hat{\mathbf{e}}_p$), which is simply the difference between the PCA-model-estimated response and the actual response:

$$\hat{\mathbf{e}}_p = \mathbf{x}_p - \hat{\mathbf{x}}_p \tag{12.20}$$

Conceptually, the residual is the portion of the measured response that cannot be explained by the PCA model.

The PCA prediction scores and residuals can then be used to generate two common quality metrics for conveying the abnormality of the measured response: namely the Hotelling T^2 statistic and the Q residual statistic, which are defined below:

$$T^2_{\text{samp}} = \hat{\mathbf{t}}_{\mathbf{p}} * \text{diag}(\lambda) * \hat{\mathbf{t}}^{\mathbf{t}}_{\mathbf{p}} \tag{12.21}$$

$$Q_{\text{samp}} = \hat{\mathbf{e}}_{\mathbf{p}} \hat{\mathbf{e}}^{\mathbf{t}}_{\mathbf{p}} \tag{12.22}$$

Conceptually, the T^2 value for a given sample reflects the 'extremeness' of that sample's response *within* the PCA model space, whereas the Q value reflects the amount of the sample's response that is *outside* of the PCA model space. Therefore, both metrics are necessary to fully assess the abnormality of a response. In practice, before one can use a PCA model as a monitor, one must set a confidence limit on each of these metrics. There are several methods for determining these confidence limits [30,31], but these usually require two sets of information: (1) the set of T^2 and Q values that are obtained when the calibration data (or a suitable set of independent test data) is applied to the PCA model, and (2) a user-specified level of confidence (e.g. 95%, 99%, or 99.999%). Of course, the latter is totally at the discretion of the user, and is driven by the desired sensitivity and specificity of the monitoring application.

12.2.6 Design of experiments (DOE)

Design of experiments (DOE) tools are widely considered to be some of the most important tools in the development of calibration models for PAT [22]. In cases where sufficiently relevant calibration samples can be synthetically prepared, DOE can be used to specify an efficient and effective set of calibration mixture standards for calibration development. Even in those PAT problems where synthetic standards cannot be prepared, the concepts behind DOE are useful for understanding the importance of covering the needed composition and instrument response space for a given problem.

The first step in choosing an experimental design for a given application is to specify the *experimental objective*. In the application of chemometrics to PAT, the experimental objective most often consists of two parts:

1. To provide calibration data that sufficiently covers the range of expected analyzer responses during real-time operation, and
2. To provide calibration data that can be used to model or assess any nonlinear effects in the analyzer data.

Depending on the PAT application, there could be additional, or different, objectives to the experiment. For example, one might not want to build a quantitative regression model for an on-line analyzer, but rather perform an exploratory analysis to identify which of many design variables has the greatest affect on the analyzer response. In this case, a set of tools called *screening designs* [22] can be quite useful.

The next step in experimental design is to *identify appropriate design variables***.** For PAT calibration problems, this step strongly depends on prior knowledge of the chemistry, physics and process dynamics of the system, as well as the measurement principles of the analyzer. In the case of multivariate calibration, design variables are often the concentrations of process sample constituents and the values of physical properties (e.g., temperature, pressure) and other process variables that could affect the analyzer response. Selection of design variables is often a 'balancing act': it is important to include all design variables that could affect the ability of the process analyzer to measure the property of interest, but the selection of too many design variables can lead to an unacceptably large number of calibration samples. The optimal set of design variables for a given problem usually depends on several factors, including technical, logistical and fiscal constraints.

After the design variables have been identified, the *number of target levels* for each design variable must be determined. Nominally, one starts with two levels for each design variable. However, in PAT, one often needs to specify more than two target levels for one or more of the design variables, so that any non-

linear relationships between these design variables and the analyzer response can be sufficiently modeled. This is particularly important if nonlinear modeling methods, such as neural networks, will be used on the resulting data. Of course, the number of design points (or the number of calibration standards) increases dramatically as the number of levels increases. However, some design tools allow one to increase the number of levels used for each variable without greatly increasing the number of samples, and these will be discussed below.

Once the number of target levels per variable is determined, the *values* of these targets must be specified. For PAT, this is also a 'balancing act'. For a typical calibration problem, one needs to assign values that cover the ranges of values that are expected to occur during real-time operation of the analyzer. At the same time, however, one must avoid using ranges that are too wide, as this might result in a degradation of model precision and accuracy in the analyzer's relevant operating range, due to local nonlinearities. For process spectrometers, where excessively high concentrations of some chemical species can result in saturated absorption peaks and nonlinear responses, one is particularly susceptible to this trap.

Once the target levels for each design variable are specified, one can choose from several different experimental design types. Figure 12.4 provides spatial representations of four different types of experimental designs that are useful for process analyzer calibration. All of these representations involve only three design variables, each of which has two specified levels. These different design types vary in their ability to characterize various interaction effects between design variables, as well as the total number of design points. Note that all four of the designs shown in Figure 12.4 include a *center point*, which represents a design point that uses central values (usually the median values) of each of the design variables. In PAT, the center point can be very useful for two reasons:

1. It enables an additional level of each design variable to be tested, so that the resulting data can provide *some* assessment of nonlinear effects between design variables and analyzer responses, and

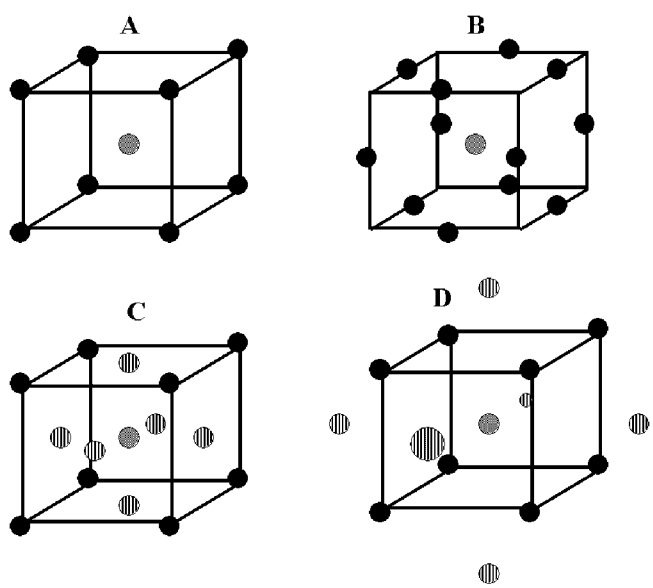


Figure 12.4 *The spatial representations of four different types of experimental designs that are useful for process analyzer calibration: (A) full-factorial, (B) Box–Behnken, (C) face-centered cube, and (D) central composite.*

Table 12.4 *Comparison of four different experimental design types, with respect to the number of samples required and the number of actual design variable levels that are used: for the case of 2 specified levels of 3, 4 and 5 design variables*

Design type	Number of samples required (for 2 ***specified*** levels)			Number of different levels ***actually used*** for each variable
	3 design variables	4 design variables	5 design variables	
Full factorial, no centerpoint	8	16	25	2
Full factorial, with centerpoint	9	17	26	3
Box-Behnken	13	25	41	3
Face-Centered Cube	15	25	43	3
Central Composite	15	25	43	4

2. If replicate center point samples are prepared, they can be used to assess the repeatability of the analyzer response and the sample preparation method.

Table 12.4 provides a comparison of these four design types, with regard to the number of samples that are required. Note that the inclusion of a single center point to the full-factorial design results in an increase in the number of levels actually used by the design from two to three. The Box–Behnken and face-centered cube design types also result in the use of three different levels for each design variable, and they also require the use of more samples than the full-factorial design. However, when one observes the spatial representation of these designs (Figure 12.4), it is clear that they do a better job of covering the design variable space than the full-factorial design, and are better able to detect nonlinear relationships between variables. The central composite design has the added benefit of generating experiments that use four different levels for each variable from only two specified target levels per design variable.

Once the type of design is specified, sufficient information is available to generate a table containing a list of calibration samples, along with the values of each design variable for each sample. However, it is important to note that there are additional considerations which can further alter this experimental plan:

1. *Randomization:* If one is concerned about time-dependent effects (for example, analyzer response drift, or chemical aging effects in the calibration standards), then the order of preparation and analysis of calibration samples in the design should be randomized.
2. *Blocking:* If there is a particular design variable that is difficult or costly to change, one might consider a sampling order in which this variable is blocked, so that it does not have to be changed frequently.
3. *Truncation:* If prior knowledge indicates that certain samples in the original design refer to combinations of design variables that are either irrelevant to the problem or not physically attainable, then these samples can be removed from the design.

12.3 Chemometric Methods in PAT

With the foundations of chemometrics as a background, this section provides some detailed discussion of several specific chemometric tools. A truly complete discussion of *all* chemometric methods would require at least a book's worth of material. As a result, this discussion will focus on the tools that have been found

to be most useful in PAT applications. These tools include data preprocessing methods, model development methods, and model optimization schemes.

12.3.1 Data preprocessing

In PAT applications, it is often the case that a large amount of the analyzer response is irrelevant for solving the problem at hand. For example, NIR diffuse reflectance spectroscopy can be very effective for on-line interrogation of solid materials, but the spectra invariably contain variations in baseline offset and multiplicative effects that are usually irrelevant for determining the property of interest. This irrelevant portion of the response can greatly hinder the ability to develop an effective method. The goal of preprocessing is to reduce or eliminate these irrelevant responses, thus reducing the burden on the modeling mathematics to develop an effective method. Although several preprocessing methods have been widely demonstrated to be very effective in PAT, it is critically important to understand the inherent assumptions of these methods, so that they can be used optimally.

Figure 12.5 shows a hierarchy of preprocessing methods that are common in PAT applications. Note that these preprocessing methods can be classified into five different categories: Offsetting, variable-wise scaling, sample-wise scaling, filtering, and compression. Some preprocessing methods from each category will be discussed in detail below.

12.3.1.1 Mean-centering

For a data array $\mathbf{X}$ containing responses of M variables over N different samples, this method involves the calculation of the mean response of each of the M variables over the N samples, and, for each of the M variables, subsequent subtraction of this mean from the original responses. Mean-centering can be represented by the following equation:

$$\mathbf{X}_{\mathrm{mc}} = \mathbf{X} - \bar{\mathbf{x}} * \mathbf{1}_N \tag{12.23}$$

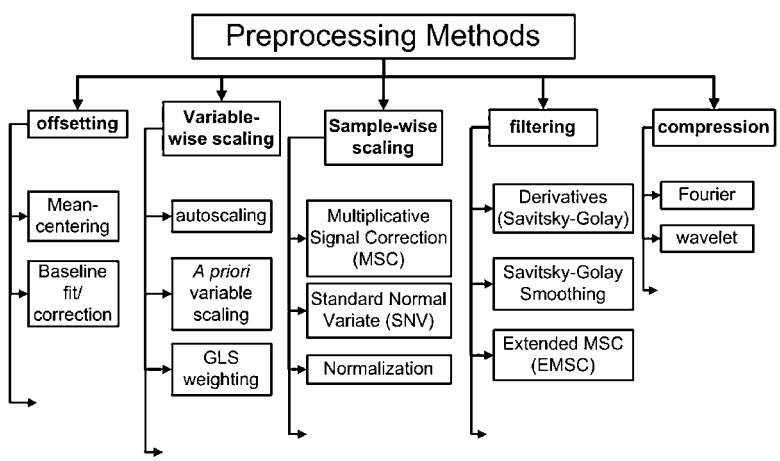


Figure 12.5 *Hierarchical representation of preprocessing methods that are common in PAT applications.*

where $\mathbf{X}_{mc}$ is the mean-centered data, $\bar{\mathbf{x}}$ is a vector that contains the mean response values for each of the M variables, and $\mathbf{1}_N$ is a vector of ones that is N elements long. For spectral data, the $\bar{\mathbf{x}}$ vector can be thought of as the mean spectrum of the specific data set, and the mean centering operation can be thought of as subtraction of the mean spectrum from each individual sample spectrum.

The mean-centering operation effectively removes the absolute intensity information from each of the variables, thus enabling subsequent modeling methods to focus on the response *variations* about the mean. In PAT instrument calibration applications, mean-centering is almost always useful, because it is almost always the case that relevant analyzer signal is represented by *variation* in responses at different variables, and that the absolute values of the responses at those variables are not relevant to the problem at hand.

It should be noted that there are a few cases in PAT where mean-centering *might not* be desired. In general, these are cases where the absolute values of the analyzer signal have meaning or, similarly, where the origin (i.e., where the values of all variables are equal to zero) represents a state that is relevant to the problem. In some spectrometry applications, depending on how spectral referencing is done, the origin can represent a state where there are no chemical absorbers in the sample, and such a state can be a useful 'baseline' for applications that require quantitation or detection of chemical species at levels close to zero concentration. Also, for such cases where it is also known that the path length (or effective path length) of the samples varies, mean-centering can actually hinder effective estimation and correction of the path length by other preprocessing methods (such as multiplicative scatter correction, or MSC, which is discussed below).

12.3.1.2 Baseline Correction

This actually refers to a family of similar spectral pretreatment methods that involve the subtraction of a baseline component **d** from an individual sample's spectrum **x**:

$$\mathbf{x}_{BC} = \mathbf{x} - \mathbf{d} * k \tag{12.24}$$

Different baseline correction methods vary with respect to the both the properties of the baseline component **d** and the means of determining the constant k. One of the simpler options, baseline *offset* correction, uses a 'flat-line' baseline component (**d** = vector of 1s), where k can be simply assigned to a single intensity of the spectrum **x** at a specific variable, or the mean of several intensities in the spectrum. More elaborate baseline correction schemes allow for more complex baseline components, such as linear, quadratic or user-defined functions. These schemes can also utilize different methods for determining k, such as least-squares regression.

12.3.1.3 Autoscaling

This common variable-wise scaling method consists of mean-centering followed by division of the resulting mean-centered intensities by the variable's standard deviation:

$$\mathbf{X}_{as} = (\mathbf{X} - \bar{\mathbf{x}} * \mathbf{1}_N)(\mathrm{diag}(\mathbf{S}_M))^{-1} \tag{12.25}$$

where $\mathbf{X}_{as}$ is the autoscaled data, and $\mathbf{S}_M$ is a square diagonal matrix containing the standard deviations for each of the M variables over the N samples in the data. The autoscaled data set $\mathbf{X}_{as}$ has the unique characteristic that each of the variables has a 0 mean and a standard deviation of 1.

The mean-centering part of autoscaling removes absolute intensity information in the x variables. However, the additional division by the standard deviation also removes total variance information in each of the vari-

ables. Conceptually, this means that each of the variables in autoscaled data $\mathbf{X}_{as}$ is 'forced' to have the same variance.

In many types of instrumental data, such as spectra and chromatograms, there can be some variables that do not exhibit much variance for a given problem. In these cases, autoscaling has the effect of amplifying the importance of noise in these variables relative to the potentially useful variance in other variables. However, there can be other cases where high-variance variables are not relevant to the problem, while there are some lower-variance variables that are relevant. In these cases, the 'equalizing' effect of autoscaling can improve the possibility of subsequent modeling methods being able to effectively utilize the relevant information in the lower-variance variables. Furthermore, autoscaling is necessary in cases when the units of measurement are not the same for all of the x variables (e.g., ppm, %, °C, pH units). In such cases, if autoscaling is not done, then those variables with the largest absolute range will tend to 'mask' those with the lowest absolute range in subsequent modeling work.

In exploratory chemometric analyses, one must also be careful when interpreting modeling results from autoscaled data. For PAT problems, one can do exploratory analyses using PCA and other chemometric methods to assess the relative sensitivities of different analyzer responses to a property of interest. If autoscaling is done, however, such an assessment cannot be done, as relative sensitivity (i.e., variance) information for the variables has been removed.

12.3.1.4 A Priori Variable Scaling

Although autoscaling is a convenient means of scaling variables, it might not be the most optimal means for a given application. In some cases, prior knowledge regarding the properties of the variables is available, such as nominal noise levels, theoretical relevance to the problem, or response linearity. In such cases, this information can be used to explicitly scale individual variables such that the variables that are most relevant and/or effective for the problem at hand are given the largest variance, and thus the largest importance in subsequent modeling work.

12.3.1.5 Derivatives (Savitsky–Golay)

This 'filtering' preprocessing method can be used whenever the variables are expressed as a continuous physical property. One example is dispersive or Fourier-Transform spectral data, where the spectral variables refer to a continuous series of wavelength or wavenumber values. In these cases, derivatives can serve a dual purpose: (1) they can remove baseline offset variations between samples, and (2) they can improve the resolution of overlapped spectral features.

Savitsky–Golay filters were developed several years ago [32–33] to enable both smoothing and derivative calculation on discrete, digitized data. These filters consist of a predefined set of coefficients, the number and values of which are defined by three parameters: (1) window width, (2) derivative order, and (3) polynomial order. Application of such a filter involves sequential multiplication of these coefficients to local spectral windows in a 'moving-window' manner across the variable axis, where the resulting products are the evaluations of the derivative at each variable. The original Savitsky–Golay work [32] showed that such a moving-window approach effectively approximates the local fitting of a polynomial and subsequent evaluation of the derivative of that polynomial.

Regarding the three adjustable parameters for Savitsky–Golay derivatives, the window width essentially determines the amount of smoothing that accompanies the derivative. For rather noisy data, it can be advantageous to use higher window widths, although this also deteriorates the resolution improvement of the derivative. The polynomial order is typically set to two, meaning that the derivative is calculated based on the best fits of the local data windows to a second-order polynomial. The derivative order, of course, dictates

whether a 1st, 2nd, 3rd or higher-order derivative is to be calculated. It should be noted that use of a derivative order of zero results in smoothing of the data only, with no derivative preprocessing.

In spectroscopy applications, a first derivative effectively removes baseline *offset* variations in the spectral profiles. Second-derivative pretreatment results in the removal of both baseline offset differences between spectra *and differences in baseline slopes between spectra*. Its historical effectiveness in NIR diffuse reflectance applications suggests that baseline slope changes are common in these applications, although there is no theoretical basis for such variations.

Figure 12.6 shows the effects of first-derivative and second-derivative pretreatments on a set of NIR diffuse reflectance spectra.

12.3.1.6 Standard Normal Variate (SNV)

This method is commonly used on spectral data to correct for multiplicative variations between spectra. In spectroscopy, such variations often originate from unintended or uncontrolled differences in sample path length (or 'effective' path length, in the case of reflectance spectroscopy), caused by variations in sample physical properties (particle size, thickness), sample preparation, sample presentation, and perhaps even variations in spectrometer optics. Such variations can be particularly problematic because they are confounded with multiplicative effects from changes in component concentrations, which often constitute the signal in quantitative applications. It is important to note that *multiplicative variations cannot be removed by derivatives, mean-centering or variable-wise scaling*.

SNV is the sample-wise equivalent to autoscaling, which was discussed earlier. The SNV correction parameters for a single spectrum **x** are simply the mean and standard deviation of the variable intensities in that spectrum. With these parameters determined, SNV correction involves subtraction of the mean intensity from each of the variable intensities, followed by division of the resulting values by the standard deviation:

$$\mathbf{x}_{\text{corr}} = \frac{\mathbf{x} - \bar{x}\mathbf{1}_M}{\sigma} \tag{12.26}$$

where $\bar{x}$ and σ are the mean and standard deviation of the variable intensities in the spectrum **x**, respectively. Figure 12.7 shows the result of SNV preprocessing of the same NIR diffuse reflectance spectra that were shown in Figure 12.6. Note that the y axis scale for the SNV-corrected data runs through 0. SNV can be thought of as a normalization process on a single spectrum, so that its resulting intensities have a mean of 0 and a standard deviation of 1.

12.3.1.7 Multiplicative Signal Correction (MSC)

Like SNV, this pretreatment method [1,21] is a sample-wise scaling method, which has been effectively used in many spectroscopic applications where multiplicative variations are present. However, unlike SNV, the MSC correction parameters are **not** the mean and standard deviation of the variables in the spectrum **x**, but rather the result of a linear fit of a 'reference spectrum' $\mathbf{x}_{\mathbf{ref}}$ to the spectrum. The MSC model is given by the following equation:

$$\mathbf{x} = a\mathbf{x}_{\text{ref}} + b \tag{12.27}$$

where $\mathbf{x}_{\text{ref}}$ is a reference spectrum, a is a multiplicative correction factor and b is an additive correction factor. For most applications, $\mathbf{x}_{\text{ref}}$ is simply assigned as the mean spectrum of the data set, although this need

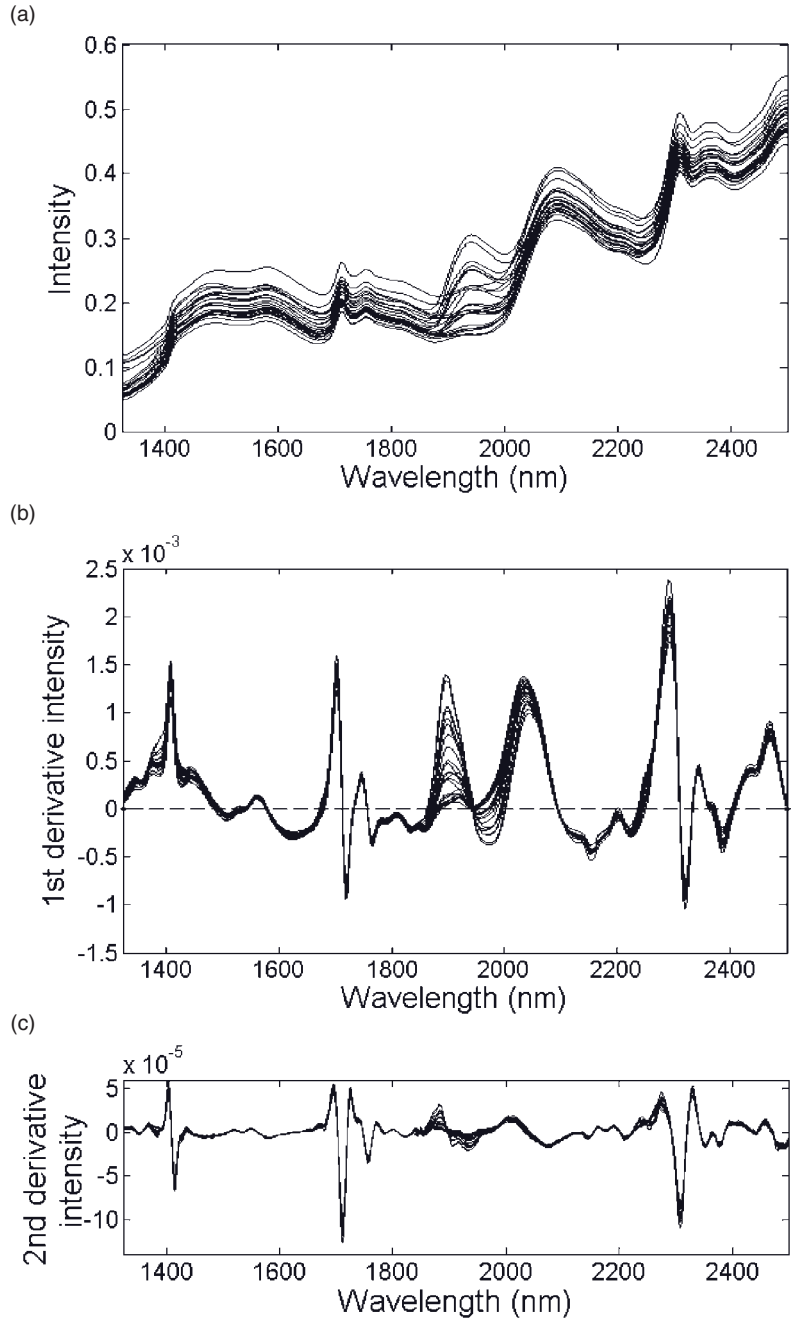


Figure 12.6 *The effect of Savistky–Golay first and second derivative preprocessing on a set of NIR diffuse reflectance spectra: (A) the raw (uncorrected) spectra, (B) spectra after 1st derivative preprocessing, (C) spectra after 2nd derivative preprocessing. In both cases the window width was 15 points (7.5 nm), and the polynomial order was 2.*

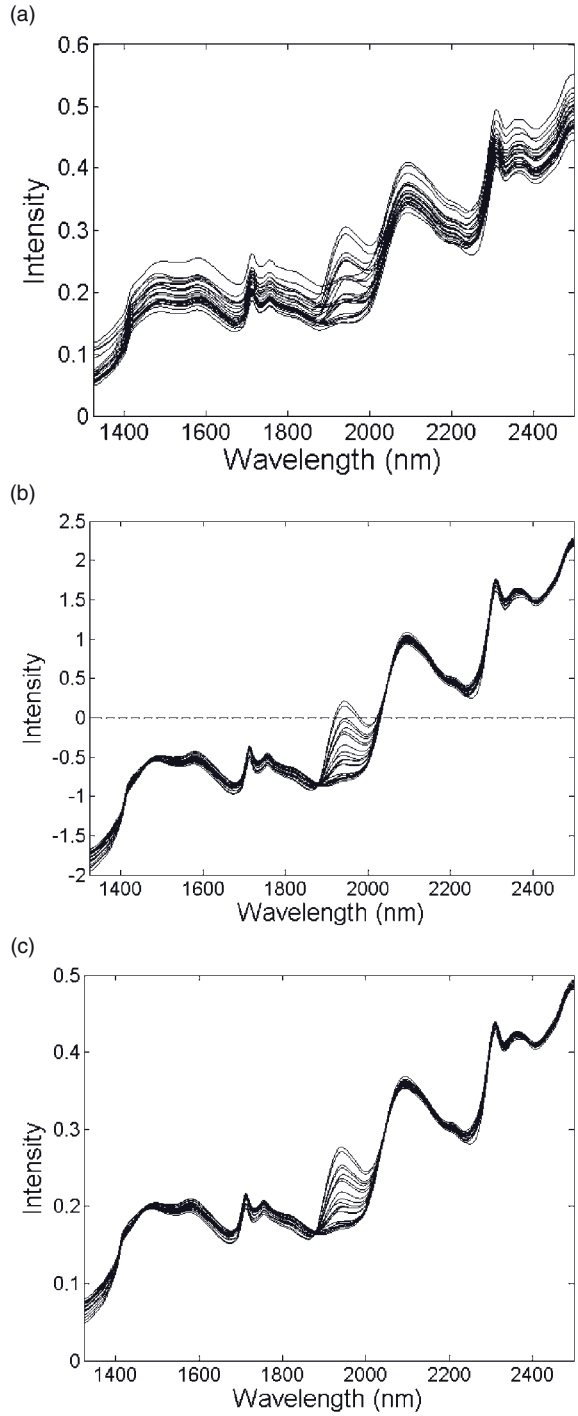


Figure 12.7 *The effect of SNV and MSC preprocessing on the same set of NIR diffuse reflectance spectra that were shown in Figure 12.6: (A) the raw (uncorrected) spectra, (B) spectra after SNV preprocessing, (C) spectra after MSC preprocessing.*

not always be the case. Given a sample's spectrum *and the reference spectrum*, the MSC multiplicative and additive correction factors (a and b) are estimated using the linear regression method (Section 12.2.3). Once they are estimated, the MSC-corrected spectrum can be calculated:

$$\mathbf{x}_{\text{corr}} = \frac{\left(\mathbf{x} - \hat{b}\mathbf{1}_M\right)}{\hat{a}} \tag{12.28}$$

where $\mathbf{1}_M$ is a vector of 1s with M elements.

When one carefully examines the MSC model above (Equation 12.27), it becomes apparent that this model assumes that any sample spectrum can be simply estimated as a multiple of the reference spectrum, plus an offset. This underscores a very important limitation of the MSC method: *it assumes that offset and multiplicative spectral effects are much larger than effects from changes in sample chemistry*. As a result, uses of this method in applications where chemical-based variations in the data are much greater than the additive and multiplicative variations can lead to poor modeling results.

Figure 12.7 shows the results obtained when MSC preprocessing (using the mean as the reference spectrum) is applied to the same NIR diffuse reflectance spectra that were shown previously. Note that both SNV and MSC accomplish some degree of correction of multiplicative variations, although these corrections are somewhat different. Although these two methods make different assumptions about the multiplicative variations in the spectral data, in most practical cases they perform rather similarly when these variations are present in the data.

12.3.1.7 Extended Multiplicative Signal Correction (EMSC)

In the previous section, it was noted that the MSC method is expected to perform poorly in cases where spectral offset and multiplicative variations are very small relative to those obtained from chemistry-based variations. In response to this, several methods were developed to enable more accurate multiplicative corrections through better modeling of chemistry-based variations in the data [34–35]. One such method is an extension of the MSC method, appropriately called extended multiplicative signal correction (EMSC) [35–37].

As the name implies, EMSC involves an extension of the MSC model in Equation 12.27, to include explicit terms for effects other than pure offset and multiplicative variations. One representation of the EMSC model is given below:

$$\mathbf{x} = a\mathbf{x}_{\text{ref}} + \mathbf{b}_H\mathbf{H} + \mathbf{c}_S\mathbf{S} + \mathbf{d}_P\mathbf{P} \tag{12.29}$$

where $\mathbf{H}$ is a matrix containing different baseline profiles to be filtered out of the data (for example baseline offset, linear baseline slope, and/or curved baseline components), $\mathbf{S}$ contains profiles corresponding to components or properties that one does *not* wish to filter out of the data, and $\mathbf{P}$ contains profiles corresponding to components or properties that one *does* wish to filter out of the data. In the special case where no profiles are specified in $\mathbf{S}$, no profiles specified in $\mathbf{P}$, and only the baseline offset profile (vector $\mathbf{1}_M$) is specified in $\mathbf{H}$, EMSC converges to MSC (Equation 12.27).

However, the benefit of EMSC over MSC is the ability to explicitly use prior knowledge of the spectroscopy and chemistry of the problem to both (a) enable better estimates of the baseline offset and multiplicative effects in the spectra, and (b) enable filtering/removal of spectral effects that are known or suspected to be irrelevant to the problem. In practice, the challenge in using EMSC effectively in PAT applications is often in the determination of spectral profiles to use in $\mathbf{H}$, $\mathbf{P}$ and $\mathbf{S}$. For example, $\mathbf{S}$ and $\mathbf{P}$ can be populated with measured, estimated, or even 'library' spectra corresponding to relevant pure components and irrelevant

interferences, respectively. **H** can be populated with various arbitrary baseline functions, or a baseline profile calculated from measured 'blank' samples. With careful selection of profiles for **H**, **P** and **S**, EMSC can be a very effective preprocessing method.

12.3.1.8 Generalized Least-Squares (GLS) Weighting

While EMSC can be used as a *filter*, to remove precharacterized irrelevant spectral contributions from a data set, GLS works slightly differently: as a *down-weighting* method to reduce the influence of predetermined multivariate directions in a data set. There are several references to GLS [1,38,39] that are quite useful, so only the general principles of GLS will be discussed here.

The first step in GLS preprocessing is the identification of a set of samples that *should* have identical responses (**x** vectors) but do not. In a typical PAT regression problem, such data can come from replicate sample measurements, or from different samples that happen to have the identical property (y) value. Analyzer data obtained from these samples can then be used to determine a difference matrix $\mathbf{X}_d$ that expresses *all* of the possible spectral differences between samples that have the *same* property value. Singular value decomposition (SVD) [1,31] is then performed on $\mathbf{X}_d$ to identify the 'multivariate directions' (linear combinations of x variables) corresponding to effects that are irrelevant to predicting the property y. This information, along with the extent to which the user wishes to down-weight these directions, is used to generate an $M \times M$ GLS weighting matrix that can then be applied to any subsequent sample spectrum.

GLS preprocessing can be considered a more elaborate form of variable scaling, where, instead of each variable having its own scaling factor (as in autoscaling and variable-specific scaling), the variables are scaled to de-emphasize multivariate directions that are known to correspond to irrelevant spectral effects. Of course, the effectiveness of GLS depends on the ability to collect data that can be used to determine the difference effects, the accuracy of the measured difference effects, and whether the irrelevant spectral information can be accurately expressed as linear combinations of the original x variables.

12.3.1.9 Compression Methods for Preprocessing

It was mentioned earlier that PCA is a useful method for 'compressing' the information contained in a large number of x variables into a smaller number of orthogonal principal components that explain most of the variance in the x data. This particular compression method was considered to be one of the 'foundations' of chemometrics, because many commonly used chemometric tools are also focused on explaining variance and dealing with colinearity. However, there are other compression methods that operate quite differently than PCA, and these can be useful as both compression methods and preprocessing methods.

One such method is Fourier compression, where the basis set of functions are predefined trigonometric functions of the original variables. In Fourier compression, each spectral profile (**x**) is essentially decomposed into a linear combination of sine and cosine functions of different frequency. The coefficients of the linear combinations indicate the contributions of different frequency components to the profile **x**. Preprocessing can then be achieved by reconstructing the original profile after excluding specific (high or low) frequency components. This can be useful for high-frequency noise filtering or reduction of low-frequency baseline fluctuations. In addition, some researchers have shown that the use of Fourier coefficients as x variables in multivariate calibration can be effective in some cases [24].

A disadvantage of Fourier compression is that it might not be optimal in cases where the dominant frequency components vary across the spectrum, which is often the case in NIR spectroscopy [40,41]. This leads to the *wavelet compression* [26,27] method, which retains both position and frequency information. In contrast to Fourier compression, where the full spectral profile is fit to sine and cosine functions, wavelet compression involves variable-localized fitting of basis functions to various intervals of the spectrum. The

'core' of the wavelet transform is the 'mother wavelet', which is a scalable oscillatory-like function (similar *but not identical* to the sine or cosine) that defines the wavelet. Once this function is chosen, it is then 'stretched' to match different x axis scales, the largest scale being 1 to M, the next largest scale from 1 to $M/2$, the next largest from 1 to $M/4$, and so on. These differently scaled wavelet functions can then be fit to localized segments of the spectrum (using linear regression) in order to obtain the wavelet coefficients. The resulting wavelet coefficients represent spectral intensities of different frequency components at different positions in the spectrum. Typically, these wavelet coefficients are then used in place of the original x variables for developing chemometric models [42–44]. However, a truncated set of wavelets can also be used to more efficiently store NIR spectral data [45].

12.3.2 Quantitative model building

In PAT, the most common purpose of quantitative model building is to convert a nonselective analyzer into a selective one, to enable its effective use for a specific application. There are several different quantitative model building tools that have been shown to be effective for PAT applications, and these will be discussed in this section.

Figure 12.8 displays an organization chart of various quantitative methods, in an effort to better understand their similarities and differences. Note that the first 'discriminator' between these methods is the *direct* versus *inverse* property. Inverse methods, such as MLR and partial least squares (PLS), have had a great deal of success in PAT over the past few decades. However, direct methods, such as classical least squares (CLS) and extensions thereof, have seen a recent resurgence [46–51]. The criterion used to distinguish between a direct and an inverse method is the general form of the model, as shown below:

$$\text{Direct: } \mathbf{X} = \mathbf{C}\mathbf{S}^{\mathbf{t}} + \mathbf{E}_{\mathrm{x}} \tag{12.30}$$

$$\text{Inverse: } \mathbf{C} = \mathbf{X}\mathbf{B} + \mathbf{E}_{\mathbf{c}} \tag{12.31}$$

Where $\mathbf{C}$ is a matrix of component concentrations or sample properties, $\mathbf{S}$ is a matrix of basis vectors (pure component spectra, or spectral profiles reflecting a pure sample property), and $\mathbf{E}_{\mathrm{x}}$ and $\mathbf{E}_{\mathrm{c}}$ are model residuals. The direct model expresses the analyzer responses ($\mathbf{X}$) as a function of concentrations, whereas the inverse model expresses concentrations as a function of the analyzer responses. Because the former is more in line with the Beer–Lambert Law (absorbance = concentration × absorptivity), it is given the label 'direct'.

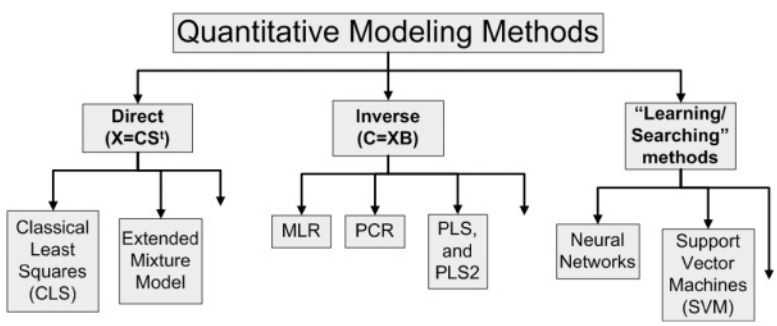


Figure 12.8 *Organization chart of various quantitative modeling methods that are commonly used for PAT applications.*

There are some general differences between direct and inverse models that are worth noting before specific modeling methods are discussed. First of all, note that the model residuals for the direct model ($\mathbf{E}_x$) are associated with the x variables, whereas the model residuals for the inverse model ($\mathbf{E}_c$) are associated with the concentration or property values. This reflects the fact that direct and inverse methods assume that all measurement error is in the x variables only and the property (y) variables only, respectively. Although neither of these assumptions is strictly accurate, as both analyzer responses (x data) and reference values (y data) have error, this does not prevent successful usage of both types of modeling for PAT applications. A second distinction between direct and inverse methods concerns the ability to build predictive models for nonconcentration properties, such as octane number or compressibility of tablets. Direct methods are not conducive for building models for such properties because they attempt to explicitly model the complete analyzer ($\mathbf{X}$) response in terms of discrete components. However, inverse methods only attempt to explain the property values ($\mathbf{C}$) in terms of the analyzer responses (X), and as a result, $\mathbf{C}$ for inverse modeling can contain concentration and/or nonconcentration properties.

During the upcoming discussion of quantitative calibration methods, a specific example involving NIR transmission spectroscopy of 70 different styrene–butadiene copolymers [52] will be used help illustrate these methods. In this example, complete laboratory NMR analyses of all of the copolymers were available, which for each sample provided the concentrations of the four known chemical components: *cis*-butadiene, *trans*-butadiene, 1,2-butadiene, and styrene. The NIR spectra of these copolymer samples over the spectral range 1570–1850 nm are shown in Figure 12.9.

Finally, the specific methods discussed below were chosen based on their utility for PAT applications and their availability in current commercial software packages. For those who are interested in a more complete list of quantitative model building methods, further details can be found in the following references [1,46–56].

12.3.2.1 Multiple Linear Regression (MLR)

MLR is an inverse method that uses the multiple linear regression model that was discussed earlier [1,46]:

$$\mathbf{y} = \mathbf{b}\mathbf{X}_{\text{sel}} + \mathbf{f} \tag{12.32}$$

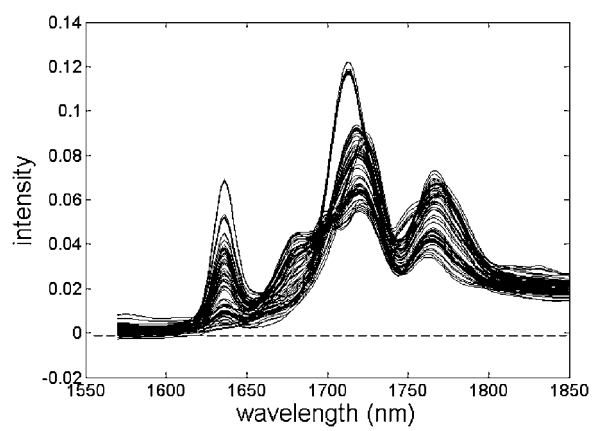


Figure 12.9 *The NIR transmission spectra of 70 different styrene-butadiene copolymers over the range 1570–1850 nm, for use in demonstrating the quantitative model building methods.*

where $\mathbf{X}_{sel}$ contains the responses of the x variables to be used in the MLR model. If the total number of available x variables (M) is less than the number of samples used to build the model (N), then $\mathbf{X}_{sel}$ can contain all of the x variables. More often, though, $\mathbf{X}_{sel}$ contains the data from a *selected subset* of m of the M x variables.

There are several means by which the x variables can be selected for a given MLR application, and these vary by three main attributes: (1) the selection criterion, (2) the selection scheme, and (3) the 'stop condition'. The selection criterion is often either the fit error or the *cross-validation* error of the MLR model. Cross validation is discussed later in Section 12.4.3, and is a means of estimating the performance of a model when it's applied to samples *not* used to build the model. Regarding selection schemes, the step-wise scheme [57,58] is rather efficient and was quite popular decades ago, when computing power was significantly more limited than it is now. This scheme involves sequential selection and retention of variables in the MLR equation until the selection criterion no longer improves. In contrast, the *all possible combinations* scheme involves the selection of the best possible subset of m variables, once the user specifies m. This scheme can be rather time-consuming if M is very large. Other selection and searching schemes, including genetic algorithms (GA) (Section 12.7.2) can also be useful and more efficient in cases where selection from a very large number of variables is necessary. Finally, the stop criterion is used to determine when to stop selecting variables for the MLR model. Several stop criteria are possible, including f-test of the distributions of fit errors before and after selecting an additional variable, the Mallows Cp [59], and user judgment based on the fit or cross-validation error before and after the selection of an additional variable. Defining an appropriate stop criterion is important, because the MLR method is particularly susceptible to overfitting through the use of too many variables, and overfitting can render a model ineffective by incorporating too much noise and spurious correlation, as discussed later in Section 12.4.1.

Once the m variables have been identified, the MLR regression coefficients are determined from a matching set of analyzer ($\mathbf{X}$) and reference ($\mathbf{y}$) calibration data using least squares:

$$\hat{\mathbf{b}}_{\mathrm{MLR}} = \left(\mathbf{X}_{\mathrm{sel}}^{\mathbf{t}}\mathbf{X}_{\mathrm{sel}}\right)^{-1}\mathbf{X}_{\mathrm{sel}}^{\mathbf{t}}\mathbf{y} \tag{12.33}$$

These coefficients can then be used to determine the property (y value) of an unknown sample from the values of its selected x variables ($x_{\mathrm{sel,p}}$):

$$\hat{y}_{p,\mathrm{MLR}} = \hat{\mathbf{b}}_{\mathbf{MLR}}\mathbf{x}_{\mathbf{sel,p}} \tag{12.34}$$

The quality of fit of an MLR model can be assessed using both the RMSEE and the correlation coefficient, discussed earlier (Equations 12.11 and 12.10).

MLR has several distinct advantages in PAT. MLR methods are relatively simple to construct, and simplicity carries a very high price in PAT, where ease of automation and long-term reliability are critical. MLR models are also rather intuitive, and easy to communicate to customers, since each individual x variable used in the equation can often be related directly to a specific chemical or physical property of the process sample. However, the biggest disadvantages of MLR are the possibility of colinearity between the selected x variables and the possibility of overfitting the model through the use of too many variables. Granted, some selection schemes by their nature *should* avoid the colinearity problem, although this is not guaranteed. The overfitting 'trap' can be avoided by using a rigorous stop criterion, *cross-validation* methods for assessing model quality (Section 12.4.3), and by adhering to the parsimony principle (i.e., 'strive to use as few variables as possible').

To illustrate the MLR method, the *step-wise* MLR calibration method is used to build a model for the *cis*-butadiene content in our styrene–butadiene example case. In this case, four variables are specified for selection, based on prior knowledge that there are four major chemical components that are varying inde-

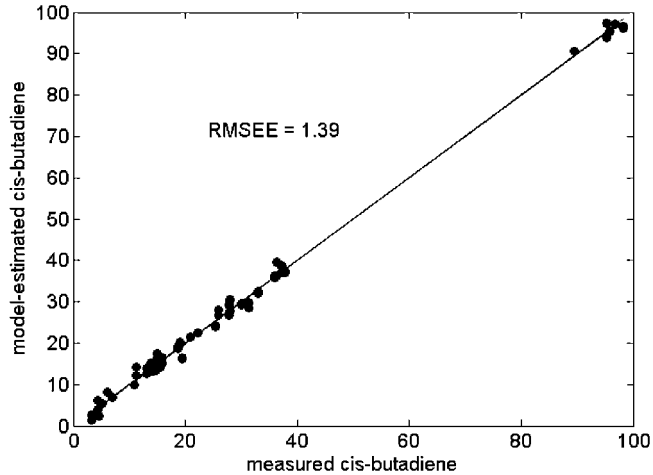


Figure 12.10 *The fit of the MLR model for* cis-*butadiene content in styrene-butadiene copolymers that uses four selected* x-*variables. The four variables were chosen by the stepwise method.*

Table 12.5 *The model fit (RMSEE) values of 4 different MLR models for predicting cis-butadiene content in styrene-butadiene copolymers by NIR spectroscopy: using 1, 2, 3 and 4* x *variables that were selected in a step-wise manner*

Wavelength(s) chosen for MLR model (nm)	RMSEE (in % cis-butadiene)
1706	4.69
1706, 1824	2.19
1706, 1824, 1670	1.90
1706, 1824, 1670, 1570	1.39

pendently in the calibration samples. The step-wise MLR method chooses the four x variables 1706 nm, 1824 nm, 1670 nm, and 1570 nm, in that order. These four selected variables are then used to build an MLR regression model for *cis*-butadiene content, the fit of which is shown in Figure 12.10. Table 12.5 lists the variables that were chosen by the stepwise MLR method, as well as the calibration fit errors (in RMSEE) obtained for the MLR models that use one, two, three and four variables. Note that the fit error continually improves as the number of variables used in the model increases.

12.3.2.2 Classical Least Squares (CLS)

In contrast to MLR, CLS is a direct calibration method that was designed specifically for use with spectroscopic data, and whose model formula is a reflection of the classical expression of the Beer–Lambert Law for a system with multiple analytes:

$$\mathbf{X} = \mathbf{CK} + \mathbf{E} \tag{12.35}$$

where $\mathbf{X}$ contains the spectra of N samples, $\mathbf{C}$ contains the pure concentrations of P chemical components for the N samples, $\mathbf{K}$ contains the pure component spectra of the P components, and $\mathbf{E}$ contains the model error [1,46]. This model indicates that any sample's spectrum is simply a linear combination of the spectra

of the pure components (in **K**). In order to implement the CLS method for quantitative modeling, *the total number of analytes present in the samples must be known beforehand*.

A CLS model can be built using either *experimentally measured* spectra or *estimated* spectra to populate **K**. Use of the latter option requires that a series of standards be obtained for which the known concentrations (**C**) of *all components* are known. The spectra of such standards ($\mathbf{X_{std}}$) could then be used to estimate **K** by least squares:

$$\hat{\mathbf{K}} = \left(\mathbf{C}^{\mathbf{t}}\mathbf{C}\right)^{-1}\mathbf{C}^{\mathbf{t}}\mathbf{X}_{\mathrm{std}} \tag{12.36}$$

For most PAT applications, this option for populating **K** is often preferred, because it is rare that high-quality pure component spectra of all components can be physically obtained. Furthermore, even if they could be experimentally obtained, they might not sufficiently represent the spectra of the pure components *in the process mixture*. Once **K** is determined, the concentrations for all analytes in a future sample ($\mathbf{c}_p$) can be estimated from the spectrum of that sample ($\mathbf{x}_p$) by least squares:

$$\hat{\mathbf{c}}_{\mathbf{p}} = \mathbf{x}_{\mathbf{p}}\hat{\mathbf{K}}^{\mathbf{t}}\left(\hat{\mathbf{K}}\hat{\mathbf{K}}^{\mathbf{t}}\right)^{-1} \tag{12.37}$$

This equation represents a simple curve fit of the pure component spectra to the unknown's spectrum, where $\hat{\mathbf{c}}_p$ contains the coefficients of the curve fit.

The same calibration fit metrics, RMSEE and *r*, can be used to express the fit of a CLS model. However, one must first calculate the model-estimated component concentrations for the calibration standards:

$$\hat{\mathbf{C}} = \mathbf{X}\hat{\mathbf{K}}\left(\hat{\mathbf{K}}\hat{\mathbf{K}}^{\mathbf{t}}\right)^{-1} \tag{12.38}$$

Then, separate RMSEE and *r* statistics can be derived for each component, using each column of the measured and estimated concentration matrix (**C** and $\hat{\mathbf{C}}$, respectively).

The CLS method has several appealing advantages for PAT. First, a single model can be used to estimate multiple analytes. Second, because the spectral responses of all analytes are being modeled simultaneously, it can be shown that the sum of the CLS predictions (sum of elements in $\mathbf{c}_p$) can be used to correct for multiplicative variations in the on-line spectra, such as those from changes in sample thickness and path length. Furthermore, when CLS is implemented using *estimated* pure analyte spectra ($\hat{\mathbf{K}}$), it can be a useful exploratory tool for extracting qualitative information regarding the spectroscopy of the analytes, or even chemical interactions between analytes, in the process sample. The 'pure analyte spectrum' basis used in CLS is also much easier to understand, and much easier to communicate to the PAT customers, especially those well-versed in spectroscopy.

Unfortunately, the CLS method has some practical and technical disadvantages and limitations. From a practical viewpoint, it is only applicable for concentration properties, rather than nonconcentration properties (such as viscosity, octane number, etc.). In addition, one must be sure that *all* of the spectrally active analytes that *could* be present in a process sample have been identified, in order to build a sufficiently relevant model. Furthermore, if one wants to use estimated pure component spectra as a basis for the CLS method, one must be able to obtain or prepare calibration standards in which the concentrations of *all* spectrally active analytes in *all* of the calibration standards are known. This requirement can make the CLS method rather resource-intensive for many PAT applications.

From the technical viewpoint, the matrix inversion $(\mathbf{C}^{\mathbf{t}}\mathbf{C})^{-1}$ in Equation 12.36 can be very unstable if any two of the analyte concentrations in the calibration standards happen to be highly correlated to one another. This translates to the need for careful experimental design in the preparation of calibration standards for CLS modeling, which is particularly challenging because multiple constituents must be considered. In addi-

tion, the CLS model assumes perfect linearity and additivity of the spectral contributions from each of the analytes in the sample. In many practical PAT applications, there can be significant nonlinearity of analyzer responses and strong spectral interaction effects between analytes, which can render the CLS method ineffective. Finally, the direct CLS model assumes that all of the noise is in the spectral (**X**) data, and there is no noise in the concentration (**C**) data, which might be grossly inaccurate in PAT applications, where the reference analytical data are often noisier than the spectral data.

Although the CLS model can be considered rather 'rigid' and limited in its scope of application, its advantages can be considerable in cases where it is applicable. Furthermore, recent work [47–51] has shown that extensions of the CLS model can reduce some of this 'rigidity', thus enabling the power of direct calibration methods to be applied to a wider scope of practical applications. Such extensions of CLS will be discussed in the following section.

In the styrene–butadiene polymer example, it is not physically possible to collect spectra of the pure butadiene components (*cis*-, *trans*- and 1,2-butadiene) experimentally, because they cannot be synthesized. As a result, only the CLS method that uses estimated pure component spectra basis can be used. Using the concentrations of all four major analytes for all of the calibration samples (obtained by ^{13}C NMR), it was possible to estimate the pure component spectra of the four known components (see Figure 12.11). When these estimated pure component spectra are then used to estimate the concentrations of the four constituents in the samples (Equation 12.38), it is found that the RMSEE for *cis*-butadiene is 1.53.

12.3.2.3 Extended Mixture Model

Although the CLS method is very rigid, it is possible to 'extend' it in order to improve its flexibility and thus open it up to more practical applications. While many analytical and PAT practitioners avoided the use of the rigid and resource-intensive CLS method in favor of more flexible and user-friendly inverse methods (such as MLR and PLS), the possibility of such *extended mixture models* had already been discussed some

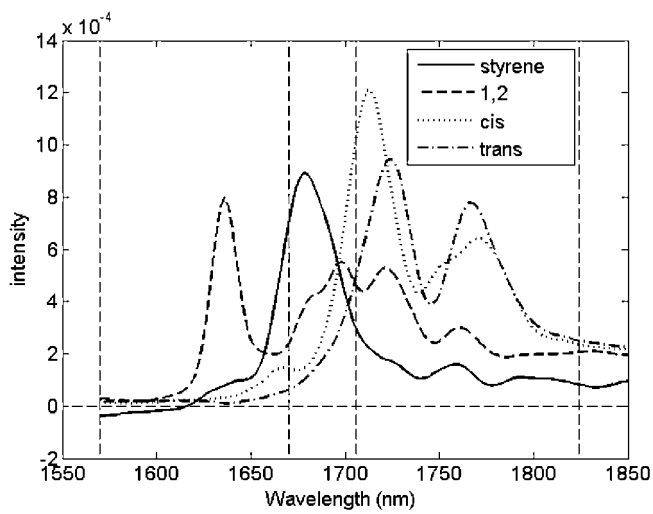


Figure 12.11 *Estimates of the pure component spectra for styrene, 1,2-butadiene,* cis*-butadiene and* trans*-butadiene units in styrene–butadiene polymers, obtained using the CLS method. The vertical dashed lines indicate the four wavelengths chosen by the stepwise MLR method.*

time ago [1], and such methods were being utilized extensively in other fields such as remote sensing [60]. In more recent years, resurgence in these types of models is starting to make its way back to the PAT community [47–51].

Appropriately, extended mixture models involve an extension of the CLS model:

$$\mathbf{X} = \mathbf{CK} + \mathbf{DL} + \mathbf{E} \tag{12.39}$$

where the term **DL** is specified to capture any systematic spectral variation that is not captured by the known components. Like **K**, **L** contains basis spectra, or profiles, of interfering components or effects that are expected to be present in the process spectra, and **D** contains the 'intensities' of these interfering components. Such interfering components can include baseline functions, peak shift functions, spectra of unknown chemical components not included in **K**, or other profiles that account for nonidealities in the CLS model.

In practice, unlike CLS, the basis vectors of the interfering effects (**L**) cannot be determined by least squares because their 'concentrations' or intensities in the calibration samples are seldom known *a priori*. Instead, these basis vectors can either be arbitrarily assigned to specific functions, such as a baseline offset or quadratic function, or they can be calculated from an independent set of auxiliary experiments that are specifically geared towards assessing interferences in the analysis. For example, some of the profiles in **L** could be loadings from an independent PCA analysis of spectra of replicate samples, and another profile could be assigned to the spectrum of an empty sample cell. In such a case, the model extension might enable better modeling of replicate sampling and baseline variation effects.

Another very appealing property of extended mixture models in PAT is the ability to explicitly account for interferences or interfering effects that cannot be present in the calibration standards. In such a case, the challenge is determining a sufficiently accurate spectral basis for the interference(s), although these can be estimated from specialized experiments or library spectra. One extension of the CLS method, prediction augmented CLS (PACLS) [50] uses results from actual predictions on process samples to determine such an interference spectral basis.

At this point, one might wonder: 'Why can't I just "load up" a CLS model with a large number of extensions covering a plethora of possible interferences?' It is true that such extensions help the model better-account for the specific interferences, but one must also keep in mind the concept of the net analyte signal (NAS) [61,62]. In short, the NAS is a vector that is the component of the pure analyte response vector that is orthogonal to the vectors of all pure interference responses. The magnitude of the NAS vector indicates the sensitivity of a direct calibration model to the component. As one adds more and more interference components to a direct calibration model, the length of the NAS vector *must* decrease because it must become orthogonal to more and more different interference vectors, with the net result of decreasing the sensitivity of the model to the analyte of interest. This is why one must be careful to only add extensions to a CLS model that are both accurate and known to be needed for process sample monitoring.

12.3.2.4 Principal Components Regression (PCR)

Like MLR, PCR [63] is an inverse calibration method. However, in PCR, the *compressed variables* (or PCs) from PCA are used as variables in the multiple linear regression model, rather than selected original x variables. In PCR, PCA is first done on the calibration x data, thus generating PCA scores ($\hat{\mathbf{T}}$) and loadings ($\hat{\mathbf{P}}$) (see Section 12.2.5), then a multiple linear regression is carried out according to the following model:

$$\mathbf{y} = \hat{\mathbf{T}}\mathbf{q} + \mathbf{f} \tag{12.40}$$

where **y** contains the known values of the property of interest, **q** contains the *y* loadings for the PCR model, and **f** contains the error in the *y* values. The estimated y loadings ($\hat{\mathbf{q}}$) are then obtained using the least-squares procedure:

$$\hat{\mathbf{q}} = \left(\hat{\mathbf{T}}^{\mathbf{t}}\hat{\mathbf{T}}\right)^{-1}\hat{\mathbf{T}}^{\mathbf{t}}\mathbf{y} \tag{12.41}$$

Unlike MLR and CLS, where the matrix inversions $(\mathbf{X}^{\mathbf{t}}\mathbf{X})^{-1}$ and $(\mathbf{C}^{\mathbf{t}}\mathbf{C})^{-1}$ can be very unstable in some situations, PCR involves a very stable matrix inversion of $(\mathbf{T}^{\mathbf{t}}\mathbf{T})^{-1}$, because the PCA scores are orthogonal to one another (recall Equation 12.16).

The PCR model can be easily condensed into a set of regression coefficients ($\hat{\mathbf{b}}_{\mathrm{PCR}}$) using the following equation:

$$\hat{\mathbf{b}}_{\mathbf{PCR}} = \hat{\mathbf{P}}\hat{\mathbf{q}} \tag{12.42}$$

The model can then be used to provide the *y* value of an unknown sample from its array of *x* values ($\mathbf{x_p}$):

$$\hat{y}_{\mathrm{p,PCR}} = \hat{\mathbf{b}}_{\mathbf{PCR}}\mathbf{x_p} \tag{12.43}$$

The calibration fit metrics previously discussed, RMSEE and *r*, can be used to express the fit of a PCR model. In this case, the RMSEE can be calculated directly from the estimated *y* residuals ($\hat{\mathbf{f}}$), which are calculated according to Equation 12.44:

$$\hat{\mathbf{f}} = \mathbf{y} - \hat{\mathbf{T}}\hat{\mathbf{q}} \tag{12.44}$$

The main advantage of PCR over inverse MLR is that it accounts for covariance between different *x* variables, thus avoiding any potential problems in the model computation mathematics, *and* removing the 'burden' on the user to choose variables that are sufficiently independent of one another. From a practical viewpoint, it is much more flexible than CLS, because it can be used to develop regression models for any property, whether it is a concentration property or otherwise. Furthermore, one needs only the values of the property of interest for the calibration samples (and not the concentrations of all components in the samples).

Like MLR, however, one must be careful to avoid the temptation of overfitting the PCR model. In this case, overfitting can occur through the use of too many principal components, thus adding unwanted noise to the model and making the model more sensitive to unforeseen disturbances. Model validation techniques (discussed in Section 12.4) can be used to avoid overfitting of PCR models.

For the styrene–butadiene example, the use of the PCR method to develop a calibration for *cis*-butadiene is summarized in Table 12.6. It should be mentioned that the data were mean-centered before application of the PCR method. Figure 12.12 shows the percentage of explained variance in both *x* (the spectral data) and *y* (the *cis*-butadiene concentration data) after each principal component. After four principal components, it does not appear that the use of any additional PCs results in a large increase in the explained variance of *x* or *y*. If a PCR regression model using four PCs is built and applied to the calibration data, a fit RMSEE of 1.26 is obtained.

12.3.2.5 Projection to Latent Structures (PLS) Regression

The inverse calibration method of Projection to Latent Structures (PLS, also known as 'partial least squares'), is very similar to PCR, and has been a highly utilized tool in PAT [1]. Like the PCR method, PLS uses the

Table 12.6 *The model fit (RMSEE) values of 10 different PCR models for predicting cis-butadiene content in styrene-butadiene copolymers by NIR spectroscopy: using 1 to 10 PCs*

Number of principal components in the PCR model	RMSEE (in % cis-butadiene)
1	17.00
2	5.80
3	3.18
4	1.26
5	1.25
6	1.25
7	1.23
8	1.23
9	1.15
10	1.15

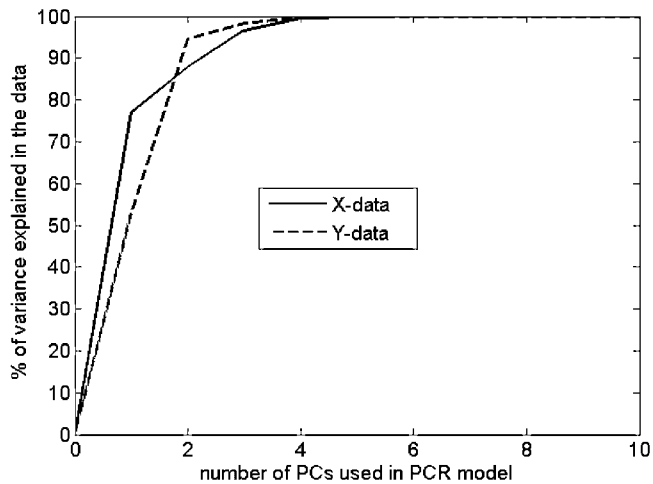


Figure 12.12 *The percentage of explained variance in both the* x *data (solid line) and y data (dotted line), as a function of the number of PCs in the PCR regression model for* cis-*butadiene content in styrene–butadiene copolymers).*

exact same mathematical models for describing the **X** data (Equation 12.14) and the y data (Equation 12.40). Like PCR, the compressed variables in PLS have the mathematical property of orthogonality, and the technical and practical advantages thereof.

The difference between PLS and PCR is the manner in which the x data are compressed. Unlike the PCR method, where x data compression is done solely on the basis of explained variance in **X** followed by subsequent regression of the compressed variables (PCs) to **y** (a simple 'two-step' process), PLS data compression is done such that the most variance in both x and y is explained. Because the compressed variables obtained in PLS are different from those obtained in PCA and PCR, they are *not* principal components (or PCs)! Instead, they are often referred to as *latent variables* (or LVs).

There are several different PLS algorithms, the most common of which are the NIPALS algorithm [1,64], the SIMPLS algorithm [65], and the Bidiagonalization algorithm [66]. These algorithms are somewhat more

complex than the PCR algorithm, so the reader is referred to the above references for details. However, all of them strive for the same goal: to find basis vectors in **X** that sequentially explain the most variance in both **X** and **y**. In the case of the NIPALS algorithm, several PLS model parameters are estimated: the *x* scores and loadings ($\hat{\mathbf{T}}_{\mathbf{PLS}}$ and $\hat{\mathbf{P}}_{\mathbf{PLS}}$), the *y* loadings ($\hat{\mathbf{q}}$) and the *loading weights* ($\hat{\mathbf{W}}_{\mathbf{PLS}}$). The loading weights have the same dimensionality as the x loadings ($\hat{\mathbf{P}}_{\mathbf{PLS}}$). Without going into much detail, the distinction between PLS loadings and weights is that the loadings are used to calculate spectra (**X**) from scores ($\mathbf{X} = \mathbf{TP}^{\mathbf{t}}$), whereas weights are used to calculate scores from spectra ($\mathbf{T} = \mathbf{XW}$).

Like PCR, a PLS model can be condensed into a set of regression coefficients ($\hat{\mathbf{b}}_{\mathbf{PLS}}$). For the NIPALS algorithm discussed above, the regression coefficients can be calculated by the following equation:

$$\hat{\mathbf{b}}_{\mathbf{PLS}} = \hat{\mathbf{W}}_{\mathbf{PLS}}\left(\hat{\mathbf{P}}_{\mathbf{PLS}}^{\mathbf{t}}\hat{\mathbf{W}}_{\mathbf{PLS}}\right)^{-1}\hat{\mathbf{q}} \tag{12.45}$$

The model can then be applied to provide the *y* value of an unknown sample from its array of *x*-values ($\mathbf{x_p}$):

$$\hat{y}_{p,PLS} = \hat{\mathbf{b}}_{\mathbf{PLS}}\mathbf{x_p} \tag{12.46}$$

As the above discussions show, the PLS and PCR methods are very similar, and generally perform quite similarly in typical PAT applications. It has often been said that PLS is slightly more susceptible to overfitting than PCR, especially if the *y* data are rather noisy. In such situations, one could easily run into a situation where the addition of a PLS factor helps to explain noise in the *y* data, thus improving the model fit without an improvement in real predictive ability. However, there might be a small advantage of PLS for qualitative interpretation. Even though the latent variables in PLS are still abstract, and rarely express pure chemical or physical phenomena, they are at least more relevant to the prediction of *y* than the PCs obtained from PCR.

In the styrene–butadiene copolymer example, Figure 12.13 shows the explained variance in both *x* and *y* as a function of the number of PLS latent variables. When this explained variance graph is compared to the

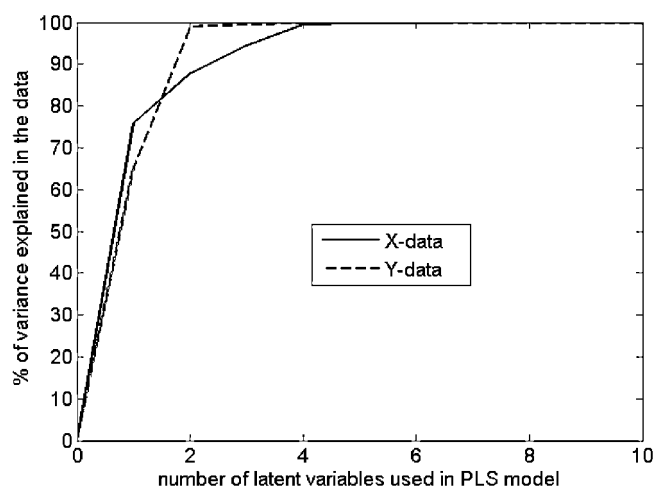


Figure 12.13 *The percentage of explained variance in both the* x *data (solid line) and y data (dotted line), as a function of the number of latent variables in a PLS regression model for cis-butadiene content in styrene–butadiene copolymers.*

same graph for the PCR model (Figure 12.12) it becomes clear that more of the *y* data are explained in the first few PLS latent variables than in the first few PCR PCs. This result reflects the fact that the PLS method chooses its compressed variables with explained *y* data variance in mind, whereas the PCR method does not. However, note that the fit of the optimal PLS model is virtually the same as that of the optimal PCR model: the fit of the PLS model that uses four latent variables is 1.25 (versus 1.26 for PCR).

12.3.2.6 PLS2: PLS for Multiple y Variables

The PLS regression method can be extended to accommodate problems where multiple *y* variables must be predicted. This extension, commonly called *PLS2*, operates on the same principle as PLS, where the goal is to find compressed variables (latent variables) that sequentially describe the most variance in both the *x* and *y* data [1]. However, the algorithm for the PLS2 method is slightly different than that of the PLS method, in that one must now account for covariance between different *y* variables as well as covariance between *x* variables.

Many PAT problems involve the need to predict multiple properties from the same analyzer's data. In such cases, it is often asked whether it would be more effective to develop a single PLS2 model than to develop a series of PLS, PCR or MLR models for each individual property. In general, there are two cases where the single-model PLS2 approach might be preferred to the multimodel approach: (1) where the goal is efficient exploratory analysis of the *x*–*y* relationship, and (2) where the *y* variables are known to be highly correlated to one another. Regarding case (2), it has been mentioned that PLS2 can exploit the correlation structure between the *y* variables to stabilize the determination of the PLS weights (**W**), especially if there is also a lot of random noise in the *y* variables [1]. However, even this improvement is limited for PLS2, as the *y* loadings (*q*) are still estimated for each *y* variable separately. Furthermore, PLS2 can encounter difficulties when nonlinearities exist in the *x*–*y* relationships, and this is especially the case if these nonlinear relationships are different for different *y* variables. Finally, a more practical advantage of the multimodel approach is that each model can be optimized individually with respect to sample and variable subsets. As a result, it is more often the case that the multimodel approach is used for PAT deployment applications.

12.3.2.7 Artificial Neural Networks (ANN)

Artificial neural networks (ANNs) [53,67,68] simulate the parallel processing capabilities of the human brain, where a series of processing units (aptly called 'neurons') are used to convert input variable responses into a concentration (or property) 'output'. As a chemometric quantitative modeling technique, ANN is rather different than all of the regression methods mentioned previously, for three main reasons: (1) the model structure is best expressed using a 'map' or architecture, rather than a simple mathematical expression; (2) the model parameters are not determined by least squares or any other closed-form mathematical operation, but rather by a 'searching' algorithm, which starts with a random selection of starting parameters and performs sequential alteration of the parameters ('training') by testing with individual calibration samples; and (3) ANNs allow for nonlinear modeling through the use of nonlinear transfer functions in the model structure.

Regarding model structure, Figure 12.14 shows one specific case of a feed-forward network with four inputs, three 'hidden nodes' (a layer of processing units that lies between the inputs and outputs), and one output. Regarding model parameter determination, one of the most common methods used in ANN training is called back propagation [69]. Naes *et al*. [67] provide a good comparison of back propagation parameter estimation versus the more straightforward least-squares parameter estimation method that is used in other regression techniques. Finally, the nonlinear modeling capability of the ANN method means that it can be effective for modeling unknown and nonlinear relationships between the *x* variables and the *y* variable.

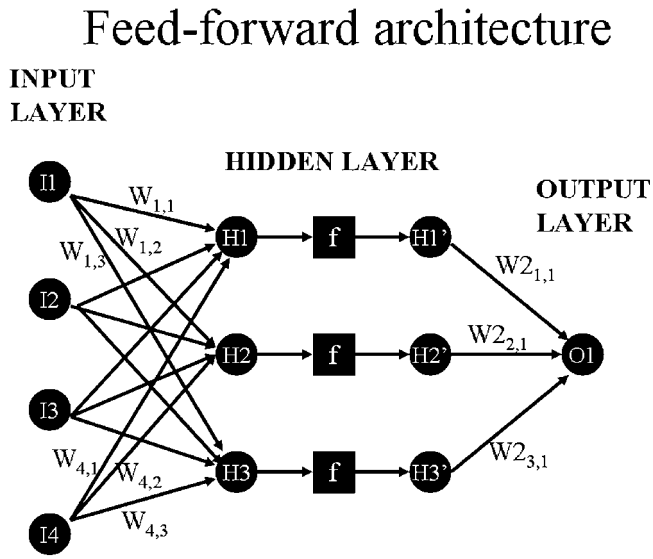


Figure 12.14 *The structure of a typical feed-forward neural network.*

Table 12.7 *The model fit (RMSEE) values of 7 different neural net (ANN) models for predicting cis-butadiene content in styrene-butadiene copolymers by NIR spectroscopy: using 1 to 6 nodes in the hidden layer*

Number of nodes in the hidden layer	RMSEE (in % cis-butadiene)
1	2.70
2	1.52
3	1.13
4	1.05
5	0.96
6	1.07

It is interesting to note that the data processing that occurs during the *operation* of a PCR model is just a special case of that of a ANN feed-forward network, where the input weights (W) are the PC loadings (**P**, in Equation 12.19), the output weights (W2) are the *y* loadings (**q**, in Equation 12.30), and there is no non-linear transfer function in the hidden layer [67].

Unfortunately, the ANN method is probably the most susceptible to overfitting of the methods discussed thus far. For similar N and M, ANNs require many more parameters to be estimated in order to define the model. In addition, cross validation can be very time-consuming, as models with varying complexity (number of hidden nodes) must be trained individually before testing. Also, the execution of an ANN model is considerably more elaborate than a simple dot product, as it is for MLR, CLS, PCR and PLS (Equations 12.34, 12.37, 12.43 and 12.46). Finally, there is very little, or no, interpretive value in the parameters of an ANN model, which eliminates one useful means for improving the confidence of a predictive model.

For the styrene–butadiene copolymer application, fit results of ANN models using one to six hidden nodes are shown in Table 12.7. Based on these results, it appears that only three, or perhaps four, hidden nodes are required in the model, and the addition of more hidden nodes does not greatly improve the fit of the

model. Also, note that the model fit error is slightly less for the ANN model that uses three hidden nodes than for the PLS model that uses four latent variables. This result could be an indicator of the improved ability of the ANN method to model nonlinear relationships between the x data and y data. It could be the case that one of the four PLS latent variables is used primarily to account for such nonlinearities, whereas the ANN method can more efficiently account for these nonlinearities through the nonlinear transfer function in its hidden layer.

12.3.2.8 Support Vector Machines (SVMs)

SVMs were first introduced by Vapnik [70,71] several decades ago as a machine learning tool, and have recently gained a great deal of attention in analytical chemistry. Although the original SVM method was used to build classifiers [71,72], rather than quantitative regression models, a regression version of SVM, commonly called least-squares SVM (LS-SVM) has been developed more recently [73,74]. This method has three attributes that make it quite different from any of the previously discussed regression methods: (1) they can use nonlinear transform functions, called *kernels*, to enable more efficient modeling of nonlinear or complex data structures; (2) the model coefficients are not determined using the standard least-squares minimization criterion (Equation 12.8), but using a more complex criterion; and (3) once developed, an SVM model is expressed in terms of a series of vectors, called *support vectors*, rather than a single regression coefficient vector. Regarding distinction (2), for the standard least-squares criterion, all calibration samples contribute to the assessment of model error to various extents, depending on their estimation error (known minus model-estimated property value). However, for SVM regression, such a 'cost penalty' is assessed only for those calibration samples that have errors above a predetermined threshold. A consequence of this distinction is that SVM regression models do not rely on all calibration samples equally, but rely more heavily on a subset of samples that are most effective for modeling the property of interest.

Like ANNs, SVMs can be useful in cases where the x–y relationships are highly nonlinear and poorly understood. There are several optimization parameters that need to be optimized, including the severity of the 'cost penalty', the threshold fit error, and the nature of the nonlinear kernel. However, if one takes care to optimize these parameters by cross-validation (Section 12.4.3) or similar methods, the susceptibility to overfitting is not as great as for ANNs. Furthermore, the deployment of SVMs is relatively simpler than for other nonlinear modeling alternatives (such as local regression, ANNs, nonlinear variants of PLS) because the model can be expressed completely in terms of a relatively low number of support vectors. More details regarding SVMs can be obtained from several references [70–74].

12.3.3 Qualitative model building

There are several circumstances where qualitative models can be very useful in PAT:

1. One needs to deploy an outlier- or fault-detection system, based on the analyzer outputs
2. Poor precision of reference (y) data causes unacceptably large errors in quantitative models, and
3. The inherent quantitative relationships between the x variables and the y variable are very nonlinear and very complex.

Cases (2) and (3) reflect an inherent advantage of qualitative models over quantitative ones: that it is not necessary to assume a quantitative relationship between the x variables and the y variable. This results in less 'burden' on the regression modeling procedure, and more likelihood of generating an effective model. The disadvantages of qualitative models are that they generate less-specific results, and that they tend to be somewhat more complicated to implement in a real-time environment.

This section will focus on classification methods, or *supervised learning* methods, where a method is developed using a set of calibration samples and complete prior knowledge about the class membership of the samples. The development of any supervised learning method involves three steps:

1. Defining the space,
2. Specifying a distance measure in the space, and
3. Developing a classification rule, using objects with known class membership.

12.3.3.1 Qualitative Model Building Procedure

Defining the space A simple classification space could be defined by a few of the individual x variables. One could also define the dimensions of the space using the first A principal components obtained from a PCA analysis of the calibration data. The selected x variable space is relatively easy to understand, and easy to implement in real-time, because the original x variables are used. However, one must be careful to avoid selecting redundant variables, or too many variables, which could lead to overfitting of the classification model. Some classification methods use empirical modeling tools to avoid selecting redundant or too many x variables. However, it is usually up to the user to limit the number of original or compressed x variables that are used to define the space.

The principal component space does not have the redundancy issue discussed above, because the PCs are orthogonal to one another. In addition, because each PC explains the most remaining variance in the x data, it is often the case that fewer PCs than original x variables are needed to capture the relevant information in the x data. This leads to simpler classification models, less susceptibility to overfitting through the use of too many dimensions in the model space, and less noise in the model.

Specifying a distance measure Once the classification space has been defined, it is necessary to define the distance in the space that will be used to assess the proximity of samples in the space. The most straightforward distance that can be used for this purpose is the *Euclidean distance* between two vectors ($D_{\mathbf{ab}}$), which is defined as:

$$D_{\mathbf{ab}} = \sqrt{(\mathbf{x_a} - \mathbf{x_b})(\mathbf{x_a} - \mathbf{x_b})^t} \tag{12.47}$$

in the case where original x variables are used to define the space. In the case where principal components are used to define the space, $\mathbf{x}_a$ and $\mathbf{x}_b$ can be replaced by $\mathbf{t}_a$ and $\mathbf{t}_b$.

Although the Euclidean distance is rather intuitive, it does not take into account the *variation* of each of the x variables (or PCs) in the calibration data. This can be particularly important in many analytical chemistry applications (such as spectroscopy), where it is often the case that samples from individual classes have different variations in the different dimensions of the space. Geometrically speaking, these classes tend to form elliptical clusters, rather than spherical clusters, in the space. To illustrate this, Figure 12.15 shows the representation of a cluster of calibration samples (that belong to a single class) and two unknown prediction samples, in a two-dimensional space. It is important to note that both of the axes in this plot use an identical scale of −1 to +1. In this situation, it is apparent that the Euclidean distances of the two unknown samples from the mean of the calibration samples are exactly the same. However, visually speaking, it appears the unknown two is more likely to belong to the class of samples than unknown one.

To address this issue, another type of distance measure, called the *Mahalanobis distance*, has been used. This distance is defined as:

$$M_{ab} = \sqrt{(\mathbf{x_a} - \mathbf{x_b})\mathbf{S}^{-1}(\mathbf{x_a} - \mathbf{x_b})^t} \tag{12.48}$$

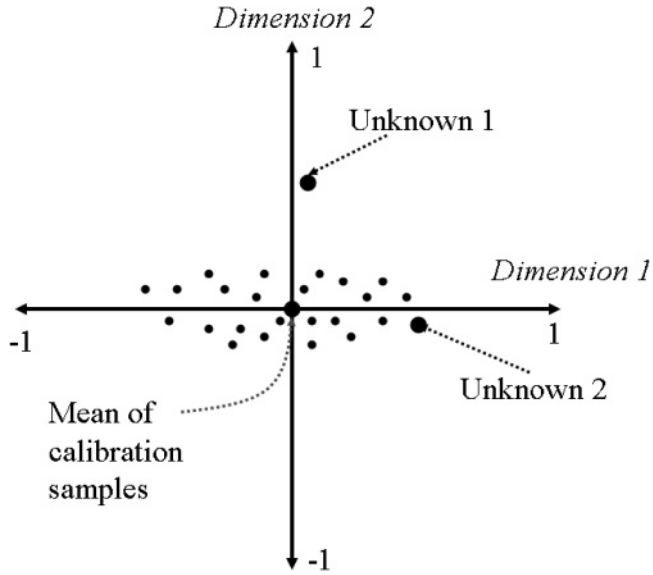


Figure 12.15 *Spatial representation of a set of calibration samples (denoted by small circles) and two unknown samples (denoted by large circles). The mean of the calibration samples is at the origin.*

in the case where original x variables are used to define the space. The same substitution of **t** for **x** is done in the case where PCs are used to define the space. **S** is the covariance matrix of the selected x variables or PCs for the calibration data, which essentially describes the total variance described by each dimension, as well as interactions between the different dimensions. In the case where PCs are used to define the space, there are no interactions between the dimensions, and **S** is simply a diagonal matrix containing the sum of squared scores for each of the PCs.

The Mahalanobis distance is simply a *weighted* Euclidean distance, where each of the dimensions is inversely weighted by its overall variance in the calibration data. As a result, deviations in a dimension that has high variance in the calibration data will not be weighted as much as deviations in a dimension that has low variance in the calibration data. With this in mind, it becomes clear that the Mahalanobis distances of the two unknowns shown in Figure 12.15 from the origin are not equal. Furthermore, it becomes qualitatively apparent that unknown 2 becomes closer to the class mean than unknown 1, if Mahalanobis distance is used instead of Euclidean distance.

Although Euclidean and Mahalanobis distances are the ones most commonly used in analytical chemistry applications, there are other distance measures that might be more appropriate for specific applications. These are discussed in reference [75].

Developing a classification rule This step requires the known class membership values for all calibration samples. Classification rules vary widely, but they essentially contain two components:

1. A set of classification parameters, and
2. The classification logic.

Typical parameters are the means and standard deviations of each class in the specified space, and the number of samples used to define each class. However, some methods might require other parameters, such as the number of nearest neighbors (KNN), the set of PCA loadings associated with each class (SIMCA), or specific

threshold values or confidence levels used to separate classes in the space. The classification logic is simply the 'recipe' or instructions that the specific method uses to assign class membership for a given unknown sample, given the parameters and the distance(s) associated with the unknown sample.

12.3.3.2 Quality Metrics for Qualitative Models

Many qualitative methods are used for what I'll call 'binary problems', in which the analyzer must generate one of only two possible results for a given observation: such as 'threat/no threat', 'presence/absence', or 'sick/not sick'. In such cases, there are precisely four possible outcomes for any individual analysis:

- a positive result for a true positive sample (true positive, TP)
- a negative result for a true positive sample (false negative, FN)
- a positive result for a true negative sample (false positive, FP), and
- a negative result for a true negative sample (true negative, TN).

Of course, the first and last outcomes are desired, whereas the second and third outcomes are not. Quality assessment of any 'binary' method requires simultaneous assessment of two method properties: (1) its ability to alarm when a true positive is encountered, and (2) its ability to *avoid* alarming when a true negative is encountered. For a given set of N test samples of known class used to evaluate the method, these properties are most commonly conveyed using the metrics *sensitivity* and *specificity*, respectively:

$$\text{sensitivity} = N_{TP}/(N_{TP} + N_{FN}) \tag{12.49}$$

$$\text{specificity} = N_{TN}/(N_{TN} + N_{FP}) \tag{12.50}$$

Where N_{TP} is the number of true positives, N_{FN} is the number of false negatives, N_{TN} is the number of true negatives, N_{FP} is the number of false positives, and $N = N_{TP} + N_{FN} + N_{TN} + N_{FP}$. Sensitivity and specificity are often expressed as either fractional quantities ranging from 0 to 1, or percentage quantities ranging up to 100%. In either case, 1.0 or 100% for *both* metrics indicates perfect method performance.

For qualitative methods in which more than two (J) possible outcomes or classifications per sample are possible, it is common to simply report the fraction or percentage of correct classifications (CC) for each individual class, as well as an overall correct classification rate.

$$\%\text{CC}_{\text{total}} = 100 \times (N_{\text{correct,total}}/N) \tag{12.51}$$

A more complete set of results for such a problem would need to be contained in a $J \times J$ matrix, where each element in the matrix [row a = 1 to J, column b = 1 to J] indicates the number of classifications of a true class a sample as class b. This would allow one to detect whether there are certain specific misclassifications that the method tends to generate more frequently.

12.3.3.3 Classification Methods – Example Data Set

To illustrate some commonly encountered classification methods, a data set obtained from a series of polyurethane rigid foams will be used [76]. In this example, a series of 26 polyurethane foam samples were analyzed by NIR diffuse reflectance spectroscopy. The spectra of these foams are shown in Figure 12.16. Each of these foam samples belongs to one of four known classes, where each class is distinguished by

Table 12.8 *Summary of the polyurethane rigid foam samples used to illustrate various classification methods*

Sample class number	Number of samples in class	Description of class
1	9	Aliphatic-based polyol
2	5	PET-based polyol
3	4	PET-based and amine-based polyols
4	8	PET-based and aliphatic-based polyols

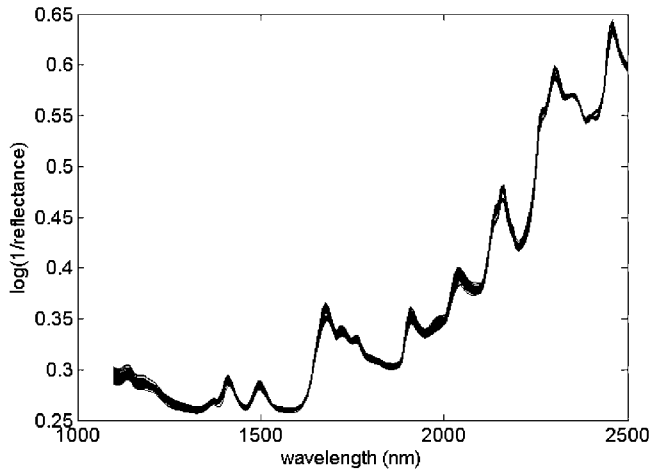


Figure 12.16 *NIR diffuse reflectance spectra of 26 poly(urethane) foams, used to demonstrate different classification methods.*

different chemistry in the hard block parts of the polymer chain. Twenty-four of the 26 samples are selected as calibration samples, and two samples are selected as prediction samples. Prediction sample A is known to belong to class number 2, and prediction sample B is known to belong to class number 4. Table 12.8 provides a summary of the samples used to produce this data set.

When the PCA method is applied to the NIR spectra in Figure 12.16, it is found that three PCs are sufficient to use in the model, and that PCs 1, 2 and 3 explain 52.6%, 20.5%, and 13.6% of the variation in the spectral data, respectively. Figure 12.17 shows a scatter plot of the first two of the three significant PC scores of all 26 foam samples. Note that all of the calibration samples visually agglomerate into four different groups in these first two dimensions of the space. Furthermore, it can be seen that each group corresponds to a single known class.

12.3.3.4 K-Nearest Neighbor (KNN)

The KNN method [77] is probably the simplest classification method to understand. Once the model space and distance measure are defined, its classification rule involves rather simple logic:

- A parameter representing the number of 'nearest neighbors', K ($<<N$), is selected.
- The distances of the N calibration samples from the prediction sample in the space are ranked from least to greatest.

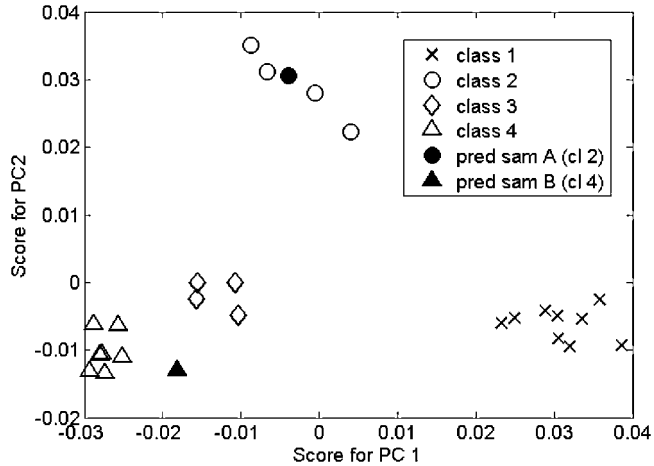


Figure 12.17 *Scatter plot of the first two PC scores obtained from PCA analysis of the polyurethane foam spectra shown in Figure 12.16. Different symbols are used to denote samples belonging to the four known classes. The designated prediction samples are denoted as a solid circle (sample A) and a solid triangle (sample B).*

- Each of the K closest calibration samples submits a 'vote' for the class to which it belongs.
- The class containing the greatest number of 'votes' is chosen as the unknown sample's class.
- If there is a tie between two or more classes, the class whose samples have the lowest combined distance from the prediction sample is selected.

The KNN method has several advantages aside from its relative simplicity. It can be used in cases where few calibration samples are available. In addition, it does *not* assume that the classes are separated by linear partitions in the space. As a result, it can be rather effective at handling highly nonlinear separation structures.

One disadvantage of the KNN method is that it does not provide an assessment of confidence in the class assignment result. In addition, it does not sufficiently handle cases where an unknown sample belongs to *none* of the classes in the calibration data, or to *more than one* class. A practical disadvantage is that the user must input the number of nearest neighbors (K) to use in the classifier. In practice, the optimal value of K is influenced by the total number of calibration samples (N), the distribution of calibration samples between classes, and the degree of 'natural' separation of the classes in the sample space.

In the example data set, KNN is illustrated using the PCA score space and the Mahalanobis distance measure. When only the first two of three PCs are viewed (Figure 12.17), prediction sample A appears to be placed well within the cluster of class two samples, and prediction sample B appears to be somewhat between the group of class three samples and the group of class four samples, although it appears to be closer to the class four samples. If the KNN classification rule is applied to the two prediction samples using values of K from 1 to 10, correct classification of both samples is achieved in all cases. However, it should be noted that one must avoid specifying a value of K that is much greater than the number of samples of the least-represented class in the calibration data, as this could ultimately eliminate the possibility that any test sample could be assigned to this class.

12.3.3.6 PLS-Discriminant Analysis (PLS-DA)

This classification method [78,79] actually uses the *quantitative* regression method of PLS (described earlier) to perform *qualitative* analysis. This is done by populating one or more y variables not with reference property values, but rather with zeros or ones, depending on the known class membership of the calibration samples. For example, if there are only two possible classes and four calibration samples, a single y variable can be constructed as follows:

$$\mathbf{y} = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix}$$

The above **y** vector illustrates the case where the first and last samples belong to the same class and the second and third samples belong to the other class. If more than two classes are involved in the problem, then it is necessary to have more than one y variable, where each y variable reflects the sample's membership to a single class. For example, for a three-class case with four calibration samples, the y matrix might look something like this:

$$\mathbf{y} = \begin{pmatrix} 010 \\ 010 \\ 001 \\ 100 \end{pmatrix}$$

This **y** matrix indicates that the fourth sample belongs to the first class, the first two samples belong to the second class, and the third sample belongs to the third class.

For PLS-DA, calibration is achieved by first applying the PLS or PLS2 algorithm to the x data and an appropriately constructed **y** matrix, to obtain a matrix of regression coefficients (**B**). At this point, the model can only generate a prediction value for each possible class, but does not have the logic to make any class assignments. Therefore, this must be followed up by the determination of a discriminant threshold score value for each class, which is used to determine membership status for each class. There are several ways in which these thresholds can be determined, some of which use Bayes' Theorem, as well as distribution assumptions regarding the prediction values. Note, however, that PLS-DA determines membership status for each class individually and therefore can assign a single sample to multiple classes or to no class at all.

Unlike other classification methods, the PLS-DA method explicitly determines relevant multivariate directions in the data (the PLS latent variables) that optimize the separation of known classes. Second, unlike KNN, the classification rule for PLS-DA is based on statistical analysis of the prediction values, which allows one to apply prior knowledge regarding the expected analytical response distributions of the different classes. Furthermore, PLS-DA can handle cases where an unknown sample belongs to more than one class, or to no class at all.

However, the PLS-DA method requires sufficient calibration samples for each class to enable effective determination of discriminant thresholds, and one must be very careful to avoid overfitting of a PLS-DA model through the use of too many PLS factors. Also, the PLS-DA method does not explicitly account for response variations *within* classes. Although such variations in the calibration data can be useful information for assigning and determining uncertainty in class assignments during prediction, it will be treated essentially

as 'noise' in the PLS model, thus rendering it not only useless in prediction but also destructive in calibration. Furthermore, this method assumes that the classes can be divided using linear partitions in the classification space, which would be problematic if the class separation is strongly nonlinear, or if there is strong natural clustering of samples in a manner that is *not* consistent with any of the known classes (i.e., irrelevant clustering).

12.3.3.7 Linear Discriminant Analysis (LDA)

LDA uses a space that is defined by a basis set of vectors, called *linear discriminants* (LDs) that are similar to the PCs obtained from PCA analysis. Like PCs, LDs are linear combinations of the original M variables in the x data that are also orthogonal to one another. However, they are determined using a quite different criterion: where the *ratio of between-class variability and within-class variability* in the calibration data is maximized.

With the LD space defined and a distance measure chosen, there are several different types of LD classifiers that can be developed [80]. A simple example would be the use of the means and standard deviations of each of the class groups in the calibration data to generate probabilities that a given test sample, when projected into the space, belongs to each known class.

Like PLS-DA, the LDA method defines a relevant classification space using known class information. In fact, the LD space might even be more relevant than the PLS-DA space, because LDA considers the both *between-group* variance and *within-group* variance. This can lead to more parsimonious LDA models versus PLS-DA and other methods given the same calibration data. In addition, the LDA method can also explicitly use *within-class* variance information in the classifier logic, most often in the form of assigning confidence levels on predicted classifications. As a result, this method can handle cases where different classes have very different magnitudes and directions of within-class variations in the space.

Like the PLS-DA method, the LDA method is susceptible to overfitting through the use of too many LDs. Furthermore, as in PLS-DA, it assumes that the classes can be linearly separated in the classification space, and its performance suffers if this is not the case.

12.3.3.8 Soft Independent Modeling of Class Analogies (SIMCA)

Unlike the methods discussed above, which strive to find directions in a common space that separate known classes, the SIMCA method [81] works on a quite different principle: define a unique space for each class, define class-specific models using each of these spaces, and then apply any unknown sample to *all* of these models in order to assess class membership.

A SIMCA model is actually an assembly of J class-specific PCA models, each of which is built using only the calibration samples of a single class. At that point, confidence levels for the Hotelling T^2 and Q values (recall Equations 12.21 and 12.22) for each class can be determined independently. A SIMCA model is applied to an unknown sample by applying its analytical profile to each of the J PCA models, which leads to the generation of J sets of Hotelling T^2 and Q statistics for that sample. At this point, separate assessments of the unknown sample's membership to each class can be made, based on the T^2 and Q values for that sample, and the previously determined confidence levels.

Although the development of a SIMCA model can be rather cumbersome, because it involves the development and optimization of J PCA models, the SIMCA method has several distinct advantages over other classification methods. First, it can be more robust in cases where the different classes involve discretely different analytical responses, or where the class responses are not linearly separable. Second, the treatment of each class separately allows SIMCA to better handle cases where the within-class variance structure is

quite different for each class. Furthermore, the SIMCA method has the flexibility to generate results that assign a sample to one, more than one, or to none of the known classes. Its class-specific modeling approach is also relatively easy to explain to customers. Finally, it can provide a quantitative assessment of the confidence of the class assignments for the unknown sample.

Although the SIMCA method is very versatile, and a properly optimized model can be very effective, one must keep in mind that this method does not use, or even calculate, between-class variability. This can be problematic in special cases where there is strong 'natural' clustering of samples that is *not* relevant to the problem. In such cases, the inherent *inter*class distance can be rather low compared to the *intra*class variation, thus rendering the classification problem very difficult. Furthermore, from a practical viewpoint, the SIMCA method requires that one must obtain sufficient calibration samples to fully represent each of the J classes. Also, the on-line deployment of a SIMCA model requires a fair amount of overhead, due to the relatively large number of parameters and somewhat complex data processing instructions required. However, there are several current software products that facilitate SIMCA deployment.

The SIMCA method is particularly well suited for the polyurethane foams example, if the objective is to build a classifier to distinguish between different materials based on the hard segment of the polymer. This is because there is significant 'natural' clustering of the samples in the PCA space (Figure 12.17) that is directly relevant to the problem. When the SIMCA method is actually applied to this data, it is found that two PCs are optimal for each of the four local PCA class models. When this SIMCA model is the applied to the prediction sample A, it correctly assigns it to class 2. When the model is applied to prediction sample B, it reports that the sample does not belong to any class, but that it is closest to class 4. Although we know that this prediction sample actually belongs to class 4, the SIMCA result is consistent with our visual assessment of the placement of sample B in the PCA space (Figure 12.17): the sample is located between classes 3 and 4, but is closest to class 4.

12.3.3.9 Other Classification Methods

Although the classification methods discussed above can cover many types of PAT applications, they certainly do not cover all methods that are available to the practicing chemometrician. For example, the support vector machines (SVM) method, which was discussed earlier in the quantitative method section, was originally developed for classification problems [70–72]. Like the SVM method, neural networks can also be used for highly nonlinear classification problems [82,83]. Bayesian classification networks [84], have also been shown to be quite effective in cases when one has prior knowledge regarding the probability of samples belonging to particular classes. The reader is encouraged to consult the above references for more detailed information regarding these and other methods.

12.3.4 Exploratory analysis

Some of the most useful applications of chemometric methods involve the extraction of hidden information in large, complex databases. At first glance, such applications might not seem to be relevant for PAT, until one considers the following:

1. PAT is called upon to support not only manufacturing operations, but also product and process research and development efforts;
2. The highly empirical nature of chemometric tools can lead to customer suspicion regarding the efficacy of chemometric models; and
3. Some PAT applications allow for, or require, analyzer hardware optimization.

With these considerations in mind, it becomes clear that the information extraction capability of chemometrics can be quite useful in several PAT-relevant situations, especially during the research and development phase of a project, or when one is trying to improve customer confidence in the capability of a calibration. As a result, a short discussion of several common exploratory tools will be made here.

12.3.4.1 Model Parameter Interpretation

In the earlier sections, the different model parameters generated by PCA, PCR, PLS and other methods were briefly mentioned, for the purpose of better understanding the mechanics of these methods. However, these same parameters can be quite useful for providing subject-matter-specific information regarding the manner in which a model operates. Such information can be critical for improving customer confidence in the efficacy of a model, as well as for enabling model optimization.

In order to illustrate the utility of model parameter interpretation, a data set containing NIR transmission spectra of a series of polymer films will be used [85]. In this example, films were extruded from seven different polymer blends, each of which was formulated using a different ratio of high-density polyethylene (HDPE) and low-density polyethylene (LDPE): where the blend compositions were 0, 2.5, 5, 10, 25, 50 and 100% HDPE. NIR spectra were then obtained for four or five replicate film samples at each blend composition. Figure 12.18 shows the NIR spectra that were obtained.

12.3.4.2 Scores and Loadings

Earlier it was mentioned, and demonstrated using the Fisher Iris example (Section 12.2.5), that the PCA scores (**T**) can be used to assess relationships between samples in a data set. Similarly, the PCA loadings (**P**) can be used to assess relationships between variables in a data set. For PCA, the first score vector and the first loading vector make up the first principal component (PC), which represents the most dominant source of variability in the original x data. Subsequent pairs of scores and loadings ([score vector 2, loading vector 2], [score vector 3, loading vector 3]…) correspond to the next most dominant sources of variability,

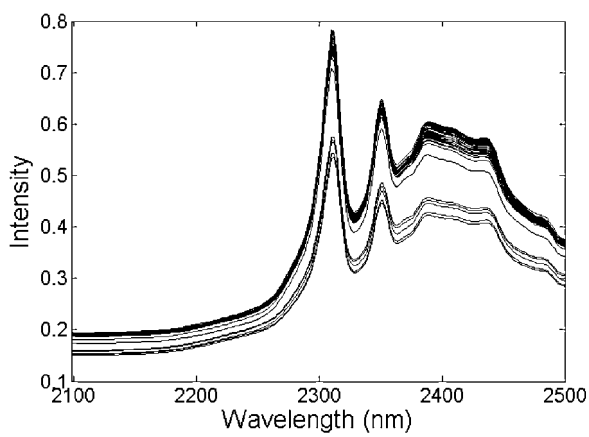


Figure 12.18 *NIR spectra of a series of 34 films obtained from polymer blends with varying ratios of high-density and low-density polyethylene.*

and so on. As a result, one can interpret matching pairs of scores and loadings vectors to better understand the dominant sources of variability in the x data. For PLS, these same pairs of scores and loadings vectors correspond to each LV, each of which sequentially expresses the most (remaining) variation in the x data that is relevant for predicting the property of interest.

A tempting, but dangerous proposition that is often considered by users of chemometrics tools is to assign each PC or LV to a *pure* chemical effect or phenomenon. This is dangerous because of the hard mathematical constraint of orthogonality on PCA and PLS scores and loadings, which almost always results in each PC or LV representing a *mixture* of two or more fundamental effects in the data. In some cases, however, there is sufficient information available to rotate the PCs and LVs to improve their interpretability, by aligning them with known phenomena in the data [85–87].

In the polyethylene example, it is important to first note the baseline offset and multiplicative variations that are readily observable in the raw film spectra (Figure 12.18), which are most likely caused by variations in film thickness, opacity, and sample placement. As a result, the MSC pretreatment method was applied to these spectra before PCA, in order to reduce these effects. A scatter plot of the first two PCA scores vectors, and a spectral plot of the first two PCA loadings vectors obtained from this analysis are shown in Figure 12.19. Note that there are two clear sample trends in the scores plot: (1) the 100% HDPE blend samples are distinguished from the rest of the blend samples, and (2) there is a clear spread in the replicate samples within each blend composition. Also, note that neither of these trends happen to be aligned with the horizontal (PC1) or vertical (PC2) direction. If these PCs are rotated by +38°, the set of scores and loadings in Figure 12.20 are obtained. Note that the sample trends are now better aligned with the horizontal (rotated PC1) and vertical (rotated PC2) directions.

At this point, the rotated loadings (Figure 12.20a) can be used to better understand the spectral nature of the two trends identified above. For a classical spectroscopist, these loadings can appear rather abstract, as they contain positive and negative features. However, they reflect true spectral variability in the data, and specifically in this case they reflect the replicate sample variability (rotated PC1) and the uniqueness in the 100% HDPE sample (rotated PC2). In the case of the former, this interpretation work was extended through an auxiliary experiment, in which the spectra of a single film sample before and after uniaxial stretching were subtracted from one another (see Figure 12.21). The striking resemblance of the stretching effect (Figure 12.21b) and the rotated PC1 loading spectrum (Figure 12.20b) indicates that the replicate sample variability is related to the nature and degree of molecular orientation in the film samples.

12.3.4.3 Residuals

The residuals, as discussed in earlier sections, contain the variation in the original x data that is *not* explained by the model. As a result, residuals can be used to assess the nature of unmodeled information in a calibration data set. It should be emphasized that, although this information typically represents a very small fraction of the information in a calibration data set, it can still be quite interesting and useful. For example, in the polyethylene film example, a spectral plot of some of the calibration sample residuals for a three PC model reveals nonrandom, sinusoidal fluctuations. When one recalls that these samples are thin films, it becomes apparent that these residuals indicate interference fringe effects that are not accounted for in the PCA model. Although these optical effects are significant, they are much weaker in magnitude than the composition and molecular orientation effects that are accounted for in the three PCs that are used in the model. Furthermore, these effects are not mathematically consistent for every sample in the data set, because they vary in shape due to differences in sample thickness and placement in the instrument. This is likely the reason why a single ‘interference fringe’ component was not identified when the model was built, and why these effects were relegated to the residuals.

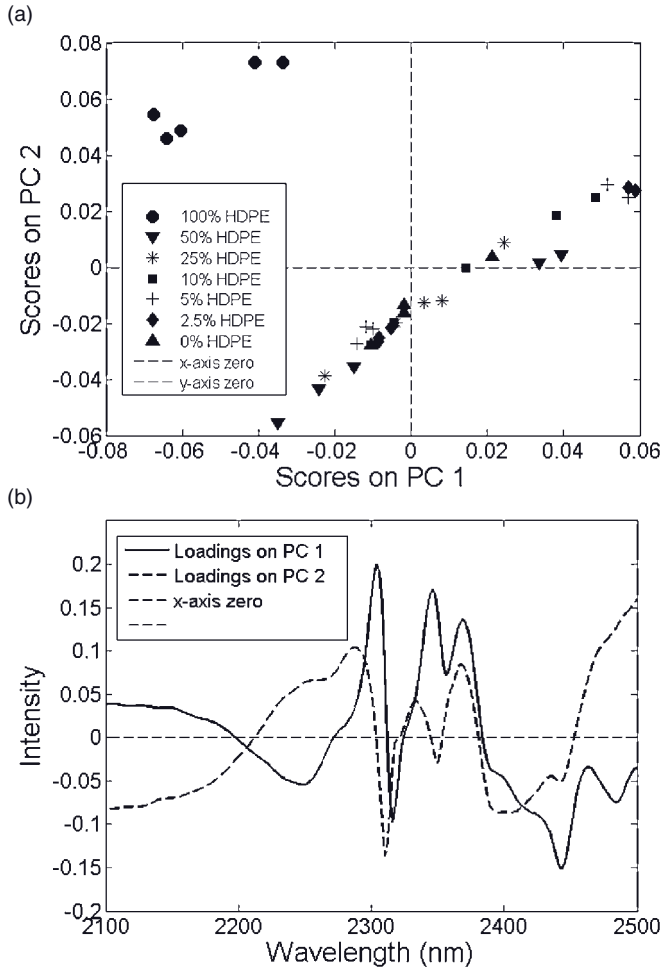


Figure 12.19 *Scores and loadings obtained from PCA analysis of the NIR spectra of high-density/low-density polyethylene blend films shown in Figure 12.18: (A) scatter plot of first 2 PC scores, (B) overlayed spectral plot of first two PC loadings. In the scores plot, different symbols are used to denote different blend compositions.*

12.3.4.4 Regression Coefficients

The regression vector ($\hat{\mathbf{b}}$) obtained from the PCR, PLS, MLR and other regression techniques can also provide some useful interpretive insight. For a given PCR or PLS modeling problem, these coefficients have the same dimensionality as a single loading vector, and thus can also be plotted as a spectral profile for spectroscopic PAT applications. However, unlike the loadings, which reveal the 'spectral signatures' of the dominant sources of variability in the x data, the coefficients indicate something quite different: namely the relative importance of each x variable for determining the property of interest. This means that variables with strong positive or negative regression coefficients do not necessarily have a high correlation to the property of interest, although this might occur in some cases. Rather, all coefficients must be considered together as a *predictor* for the property. This is because the coefficient value at a given variable must take into account spectral effects from phenomena that interfere with the property of interest, which can often

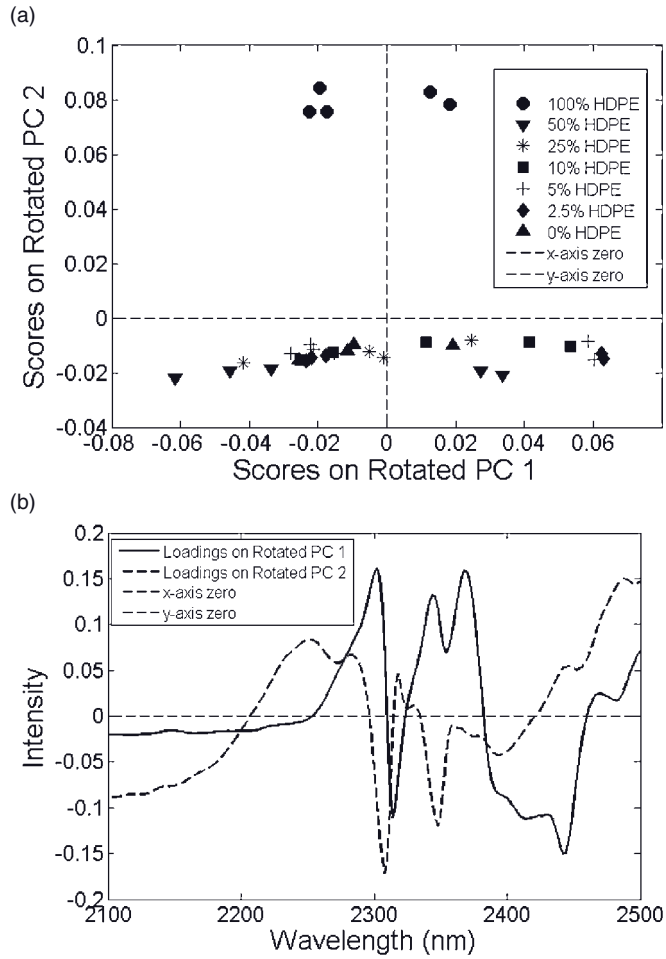


Figure 12.20 *Scores and loadings obtained from PCA analysis of the NIR spectra of high-density/low-density polyethylene blend films shown in Figure 12.18, after rotation of the PCs by +38°: (A) scatter plot of first two rotated PC scores, (B) overlayed spectral plot of first two rotated PC loadings. In the scores plot, different symbols are used to denote different blend compositions.*

be much stronger than spectral effects that are truly generated by the property of interest. In light of this, it is probably more accurate to say that the regression coefficients indicate those variables that are most relevant for determining the property of interest within the specific context of the calibration data set.

Another interesting property of the regression coefficient is that it incorporates more and more noise and increases in overall amplitude as the number of PCs or LVs increases. This is a mathematical consequence of the fact that the regression vector must become orthogonal to more and more multivariate directions as the number of PCs and LVs increases, and the fact that more noise from the x and y data is incorporated into the model as the number of PCs and LVs increases (recall the Net Analyte Signal discussion in Section 12.3.2.3). As a result, one can roughly assess the degree of overfitting through qualitative assessment of the 'noisiness' or random behavior of the regression vector.

In the polyethylene example, Figure 12.22 shows the regression coefficient spectrum for the PCR calibration to the percentage HDPE content in the polyethylene blends. Two particularly recognizable features in

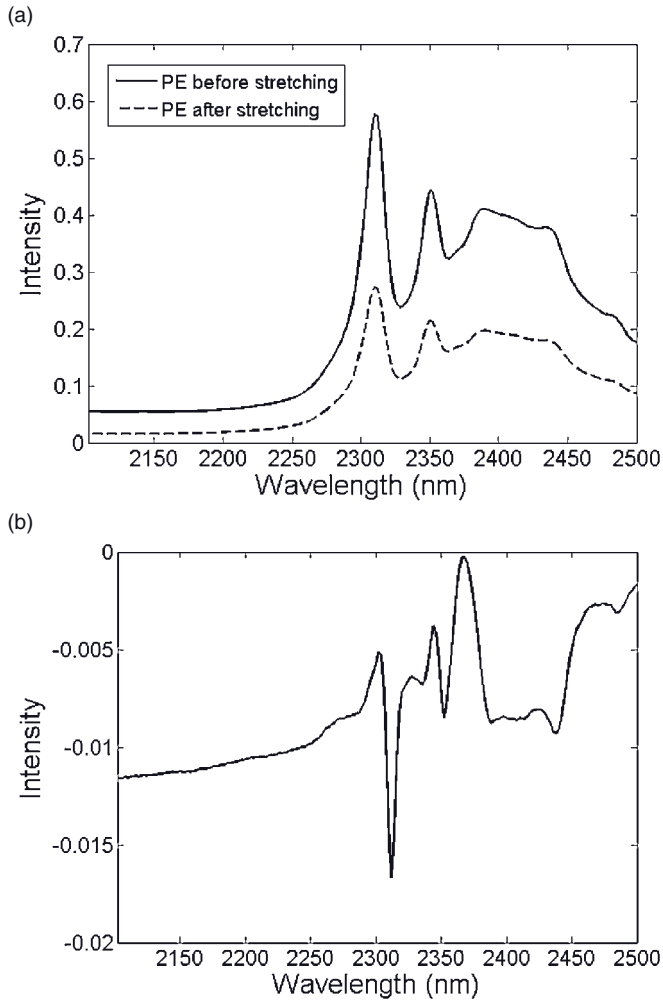


Figure 12.21 *(A) NIR spectra of a 0% HDPE film before and after uniaxial stretching; (B) weighted difference of the two spectra in A, corresponding to the uniaxial stretching effect.*

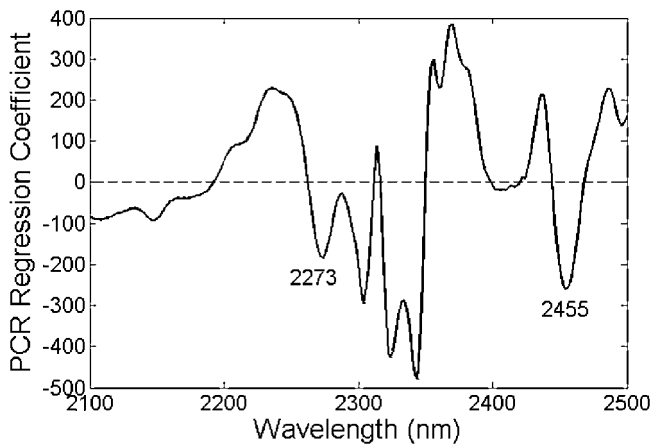


Figure 12.22 *The regression coefficient spectrum for the PCR model of high-density polyethylene content in the polyethylene blend films.*

this plot are the negative peaks at approximately 2273 nm and 2455 nm, which are locations of prominent methyl ($-CH_3$) combination bands. Because the concentration of methyl groups in the polymer is expected to be inversely related to the percentage HDPE content, the presence of these negative methyl peaks in the regression coefficient spectrum makes sense. However, please note that there are many other stronger features in the regression coefficient spectrum, indicating that other interfering effects must be accounted for in order to build an effective predictor of percentage HDPE content.

12.3.4.5 Multivariate Curve Resolution

Earlier, it was mentioned that due to the orthogonality constraints of scores and loadings, as well as the variance-based criteria for their determination, it is rare that PCs and LVs obtained from a PCA or PLS model correspond to pure chemical or physical phenomena. However, if one can impose specific constraints on the properties of the scores and or loadings, they can be rotated to a more physically meaningful form. The multivariate curve resolution (MCR) method attempts to do this for spectral data.

MCR methods [88–92] use the exact same Beer's Law model as the CLS method (Equation 12.35). However, unlike the CLS method, the MCR method can be used without knowing the concentrations of all of the constituents in the calibration samples. Given only the spectral data ($\mathbf{X}$), an estimate of the number of chemically varying and spectrally active components represented in the data (A), and some *constraints* regarding the nature of the expected structures of the pure component profiles ($\mathbf{C}$) and pure component spectra ($\mathbf{K}$), this method provides estimates of the pure component spectra ($\hat{\mathbf{K}}$) and the pure component concentrations ($\hat{\mathbf{C}}$).

The MCR algorithm involves an iterative process, and the constraints put on $\mathbf{C}$ and $\mathbf{K}$ during this iterative process are critical, and not always sufficient, to enable the algorithm to converge to a useful solution. Common constraints that are put on the concentration profiles ($\mathbf{C}$) include non-negativity, closure (i.e., all constituent concentrations for a given sample must add up to 100%), and unimodality. The pure component spectra ($\mathbf{K}$) are often constrained to non-negativity, and can even be constrained to a specific profile based on laboratory-collected pure component spectra or library spectra. In fact, one has the flexibility to constrain *any portion* of $\mathbf{C}$ or $\mathbf{K}$ to have a specific structure, and this makes the MCR method particularly accommodating to any small amount of prior knowledge that the user has available.

Although the MCR method can be a very effective exploratory method, several warnings are appropriate. One must be careful interpreting MCR-determined $\hat{\mathbf{K}}$ and $\hat{\mathbf{C}}$ as 'absolute' pure component spectra and pure component time profiles, respectively, due to the possible presence of intercorrelations between chemical components and spectral features, as well as nonlinear spectral interaction effects. Furthermore, the optimal number of components to be determined (A) must be specified by the user, and is usually not readily apparent *a priori*. In practice, A is determined by trial and error, and any attempts to overfit an MCR model will often result in $\hat{\mathbf{K}}$ or $\hat{\mathbf{C}}$ profiles that either contain anomalous artifacts, or are not physically meaningful. The application of inaccurate or inappropriate constraints can also lead to misleading or anomalous results with limited or no information content.

In a specific example, the MCR method is applied to a set of 210 FTIR spectra obtained sequentially during a chemical reaction of the type $A + B \rightarrow C$. Figure 12.23 shows the original spectra of the samples obtained throughout the reaction, and the MCR-determined spectral and composition profiles, obtained using non-negativity and closure constraints on the concentration profiles, and the non-negativity constraint on the spectral profiles. In this particular case, four components were calculated by the MCR method, based on trial and error and cross-validation. The concentration profiles show a typical 'evolving' pattern, where the reactants are consumed early in the reaction, intermediates are formed and consumed in the middle of the reaction, and the product is formed at the end of the reaction. The pure spectral profiles corresponding to each of the components can be used to provide further information regarding the nature of the pure

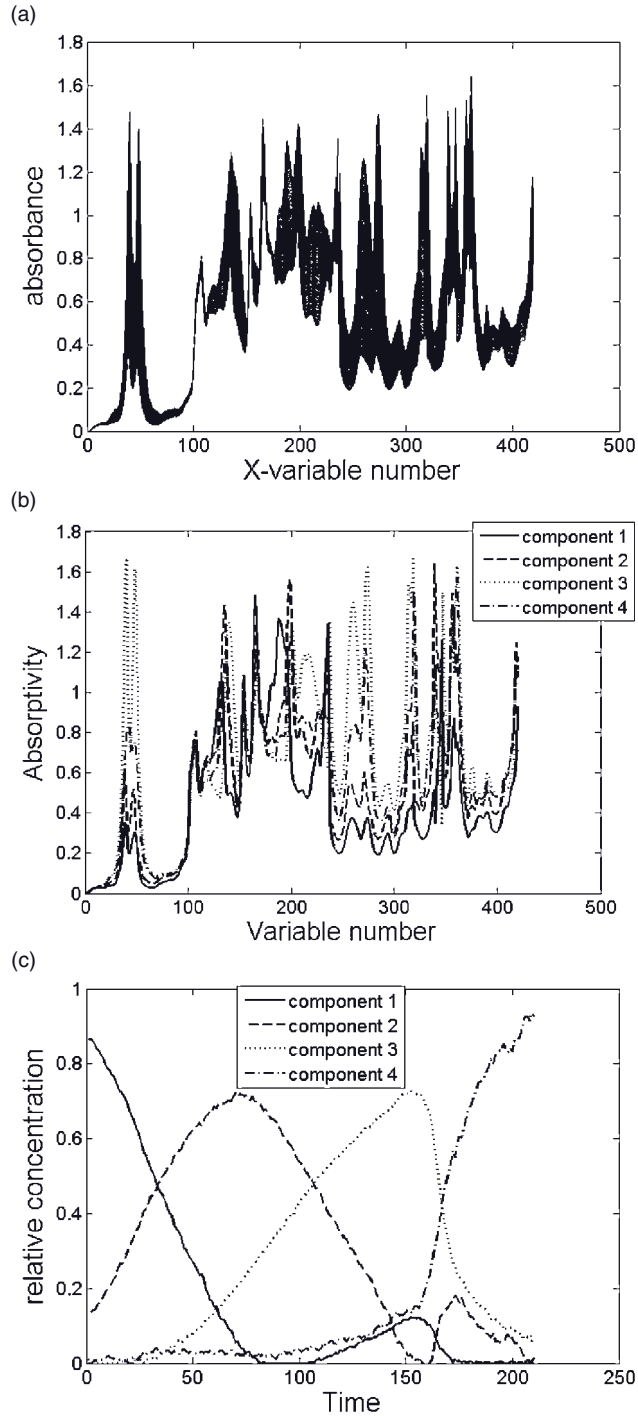


Figure 12.23 *Example of multivariate curve resolution (MCR). (A) A series of 210 FTIR spectra obtained during the course of a chemical reaction; (B) the pure spectral profiles obtained using MCR; and (C) the corresponding pure concentration profiles obtained using MCR.*

components that are represented in the concentration profiles, especially if one has prior knowledge of spectroscopic band assignments for specific functional groups in the process material.

In PAT applications, MCR methods have been found to be particularly useful for scouting studies of new products and new processes, where little is known about the chemistry and process dynamics, and reliable reference analytical methods have not yet been developed.

12.3.4.6 Hierarchical Cluster Analysis (HCA)

HCA is a common tool that is used to determine the 'natural' grouping of objects, based on their multivariate responses [75]. In PAT, this method can be used to determine natural groupings of samples or variables in a data set. Like the classification methods discussed above, HCA requires the specification of a space and a distance measure. However, unlike those methods, HCA does not involve the development of a classification rule, but rather a *linkage rule*, as discussed below. For a given problem, the selection of the space (e.g., original x variable space, PC score space) and distance measure (e.g., Euclidean, Mahalanobis) depends on the specific information that the user wants to extract. For example, for a spectral data set, one can choose PC score space with Mahalanobis distance measure to better reflect separation that originates from both strong and weak spectral effects.

The HCA method starts with each of the objects belonging to its own cluster. Each step of the algorithm involves joining two of the clusters together, using a previously specified linkage rule. Several linkage rules are possible, and some of these are described below:

- nearest neighbor: uses the distance between the *two closest samples* in the different clusters
- furthest neighbor: uses the *greatest distance between any two samples* in the different clusters
- pair-group average: uses the *average distance between all possible pairs of samples* in the two clusters.

The algorithm proceeds until a single cluster containing all samples is reached, but keeps track of the distances and clustering sequence that occurred. HCA results are commonly displayed using a dendrogram plot [75], which provides the user with information regarding both the subsets of samples that cluster together, and the threshold distances at which each sample joins a cluster. If the user then specifies a specific threshold distance, or a specific number of clusters, then a specific grouping of objects can be assigned, which can be effectively conveyed in either a dendrogram plot or a scatter plot of the objects in the specified space, although the scatter plot can only show the clustering of samples in two dimensions of the space at a time.

Comparing the different linkage rules listed above, the nearest-neighbor method tends to generate clusters that are elongated in the space, while the furthest-neighbor method tends to generate more 'rounded' clusters. The pair-group average linkage rule requires more computation time, but is equally appropriate to use in cases where natural sample clustering is 'elongated' or 'rounded'.

HCA is illustrated using the poly(urethane) foam data, where the PC score space, Mahalanobis distance measure, and nearest-neighbor linkage rule are used. Two HCA analyses are done: one using only the first two PCs (73.1% explained variance), and one using the first five PCs (95.8% explained variance). The results of these analyses are shown in the dendrograms in Figure 12.24. Dendrograms, which are commonly used to illustrate HCA results, typically have the y axis as a sample identifier axis (where samples are sorted along the axis according to their proximity in the defined space), and the x axis corresponding to the distance value. In these plots, the sample identifiers consist of four characters, where the third character corresponds to the known class number (Table 12.8) and the fourth character identifies the sample within a class. For a fixed distance measure, the number of clusters is revealed as the number of 'branches' in the dendrogram

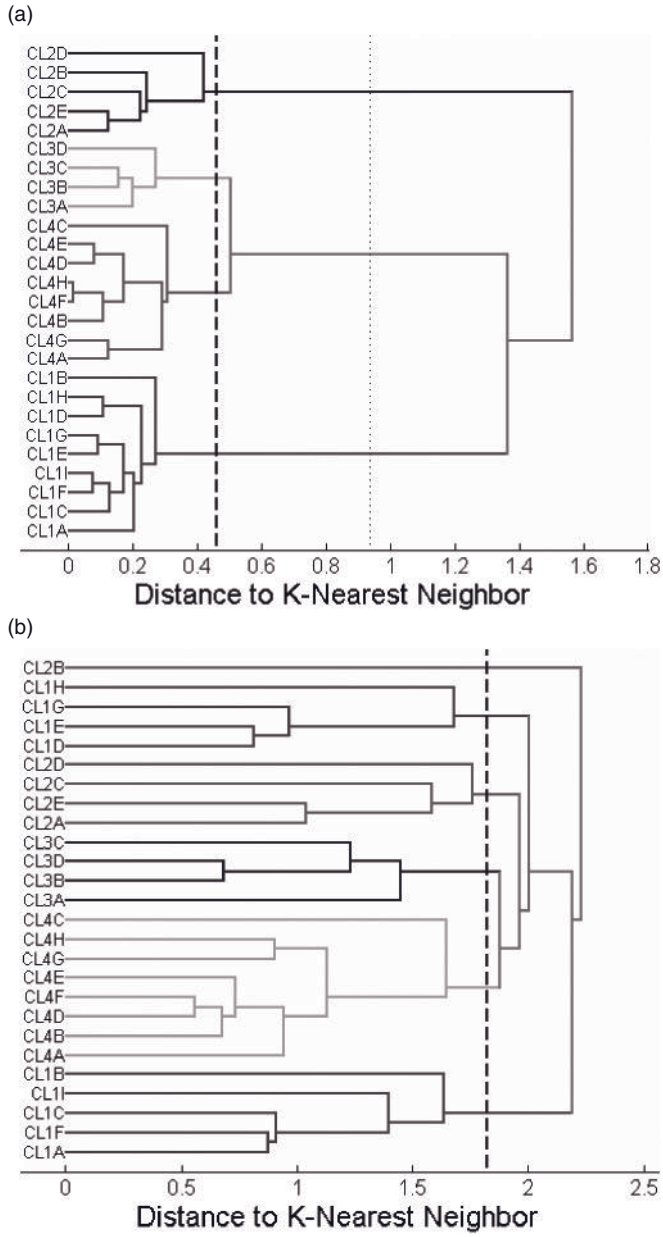


Figure 12.24 *Dendrograms obtained from hierarchical cluster analysis (HCA) of the NIR spectra of the poly(urethane) foam samples (shown in Figure 12.16), (A) using the first two PCA scores as input, (B) using the first five PCA scores as input. In both cases, the Mahalanobis distance measure and the nearest-neighbor linkage rule were used.*

that intersect with a vertical line at the specified distance measure, and the membership of each of these clusters is revealed by the samples that connected to each of these 'branches'. With this in mind, note that the HCA done using only two PCs (Figure 12.24A) results in 100% correct clustering of the four classes into their known groups for a specified distance measure of approximately 0.46. However, the same cannot be said for the HCA that uses five PCs (Figure 12.24B). In this case, the samples in classes 3 and 4 tend to cluster based on class, but some of the class 1 and 2 samples tend to generate inter-class clusters. For a distance measure of approximately 1.8, classes 3 and 4 form their own cluster, which means there are still four different clusters that cover classes 1 and 2. Recall that the PCA scatter plot in Figure 12.17, which shows only the first two PCs, visually suggests that only two PCs are necessary to discriminate between all four classes. Consequently, the HCA results obtained here suggest that PCs 3, 4 and 5 are not relevant for discriminating between all four classes. Furthermore, the use of the Mahalonobis distance measure in this case puts these irrelevant PCs at 'equal footing' with the more relevant first two PCs, thus contributing to the unfavorable result.

12.3.4.7 Partitional Cluster Analysis – K-Means

The HCA methods discussed above can also be considered *agglomerative* clustering methods, because they start with each of the objects belonging to its own cluster, and progress by 'agglomerating' or combining existing clusters together. However, the opposite approach to clustering can be taken: that of starting with a single cluster of all objects with sequential division of groups based on distance. Such methods are called *partitional* clustering, and one of the more common of these methods is called K-Means clustering. The K-Means method involves the initial selection of K cluster 'targets', which usually are individual objects. All objects are then assigned to a cluster based on their distance to the target, and then K new targets are determined as the centroids of the resulting clusters. This process is repeated until convergence is obtained: that is, no samples change groups between iterations.

The K-Means method can be rather efficient for a large number of objects, but requires *a priori* selection of the number of clusters K, and thus is less exploratory in nature than the HCA methods discussed earlier. In addition, its outcome depends on the initial selection of targets. As a result, one can do a 'smart selection' of the initial targets based on their uniqueness in the data set, for example those objects with the highest Hotelling T^2 value (Section 12.2.5) obtained from a PCA analysis of the data.

12.4 Overfitting and Model Validation

12.4.1 Overfitting and underfitting

Throughout the above discussions of modeling tools, the dangers of *overfitting* a model were repeatedly stressed. These dangers cannot be understated, not only because it is so tempting to overfit a model, due to the outwardly appealing model fit statistics that are obtained from doing so, but also because an overfit model carries with it a series of problems that are detrimental to any PAT application. Specifically:

- an overfit model incorporates more of the noise from the analyzer and reference data, and
- an overfit model is more specific to the exact data used to build it.

Another way of stating the second detriment is that an overfit model is much more sensitive to any condition that deviates from the conditions used to collect the calibration samples.

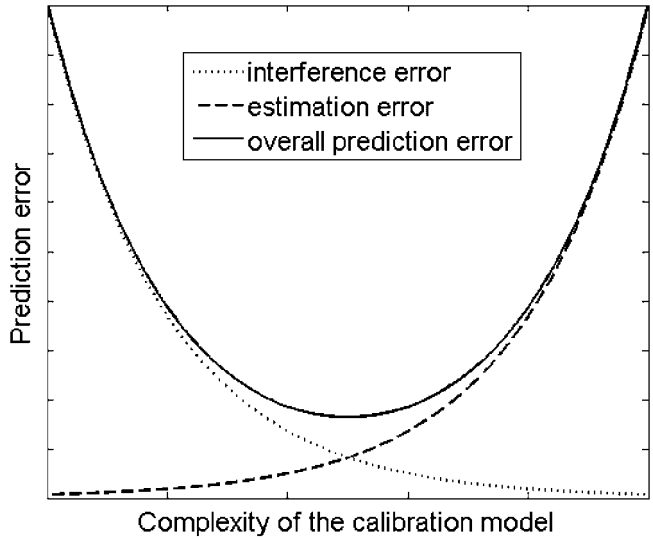


Figure 12.25 *Conceptual plot of interference error (dotted line), estimation error (dashed line), and overall prediction error (solid line) as a function of model complexity.*

A less tempting alternative, which is equally dangerous, is to *underfit* a model. In this case, the model is not sufficiently complex to account for interfering effects in the analyzer data. As a result, the model can provide inaccurate results even in cases where it is applied to conditions that were used to build it!

Figure 12.25 provides a graphical explanation of the phenomena of over and underfitting [1]. It shows that the overall prediction error of a model has contributions from two sources: (1) the interference error and (2) the estimation error. The interference error continually decreases as the complexity of the calibration model increases, as the added complexity enables the model to explain more *interferences* in the analyzer data. At the same time, however, the estimation error of the model increases with the model complexity, because there are more independent model parameters that need to be *estimated* from the same limited set of data. These competing forces result in a conceptual minimum in the overall prediction error of a model, where the combination of interference error and estimation error are minimized. It should be noted that this explanation of a model's prediction error assumes that the calibration data are sufficiently representative of the data that will be obtained when the model is applied.

In PAT, one is often faced with the task of building, optimizing, evaluating, and deploying a model based on a limited set of calibration data. In such a situation, one can use *model validation* and *cross-validation* techniques to perform two of these functions: namely to *optimize* the model by determining the optimal model complexity and to perform preliminary *evaluation* of the model's performance before it is deployed. There are several validation methods that are commonly used in PAT applications, and some of these are discussed below.

12.4.2 Test set validation

The most straightforward means of assessing a model built from a set of calibration data is to apply it to a set of independently collected data. Of course, this method directly simulates the case where the model is to be deployed and applied to future samples, except for the obvious fact that the properties of these test samples are known *a priori*, thus enabling an assessment of the model performance.

The main figure of merit in test set validation is the root mean square error of prediction (RMSEP):

$$RMSEP = \sqrt{\frac{\sum_{i=1}^{N} (\hat{y}_i - y_i)^2}{N}} \tag{12.52}$$

where $\hat{y}_i$ is the model-estimated property value for test set sample i, y_i is the measured property value of test set sample i, and N is the number of test set samples. This equation is very similar in form to the equation for RMSEE (Equation 12.11), with one notable exception: the N samples represented in the equation refer to the *test set samples,* rather than the calibration samples.

The optimal complexity of a model can be determined by plotting the RMSEP versus the complexity of the model. In the case of PCR and PLS regression, complexity is conveyed by the number of PCs and LVs, respectively. In the styrene–butadiene copolymer example, a test set of seven samples was extracted from the data set, and the remaining 63 samples were used to build ten different PLS models for *cis*-butadiene, using 1 to 10 latent variables. These models were then used to predict the *cis*-butadiene of the seven samples in the test set. Figure 12.26 shows both the calibration fit error (RMSEE) and the validation prediction error (RMSEP) as a function of the number of latent variables in the PLS model. Note that the test set validation error in this example increases after four latent variables are added to the PLS model, whereas the calibration fit error continually decreases, as it must. As a result, it appears that the use of four latent variables in this PLS model is optimal, and that the addition of more latent variables causes model overfitting. This example illustrates the fact that the use of validation procedures can provide a more definitive estimate of the optimal complexity of a calibration model, as well as a more realistic estimate of the prediction error of this model. It should be noted, however, that such plots do not always provide definitive results for the optimal model complexity, and one must often make a judgment call based on prior knowledge of the system,

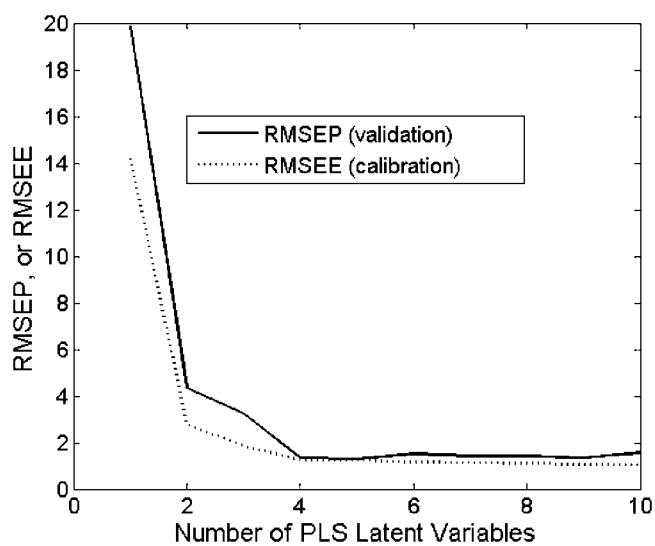


Figure 12.26 *Plot of the calibration error (RMSEE) and the validation error (RMSEP) as a function of the number of latent variables, for the case where 63 of the styrene–butadiene copolymer samples were selected for calibration, and the remaining seven samples were used for validation.*

or the demands of the application. One can also rely on automated methods for determining the optimal complexity, such as those that are based on an f test of the model residuals [93,94].

The test set method of validation is rather straightforward, but requires some caution. First, the test set samples must be sufficiently representative of samples that the model will be applied to in the future. Otherwise, external validation can provide misleading results – in either an optimistic or pessimistic manner. Under-representing the anticipated sample states could lead to optimistic results, while the use of sample states not represented in the calibration data could lead to pessimistic results. A practical disadvantage of the test set method is that it requires that the reference analytical method be performed on an additional set of samples beyond the original calibration samples, which can be an issue if the cost of each reference analysis is high.

However, if the test samples are sufficiently 'balanced' to cover the sample states used to build the model without extending beyond those states, and the cost of obtaining and analyzing these test samples is not prohibitive, then the test set validation method can be very effective. In fact, in some high-stakes applications where model evaluation is the main objective, such validation is required, often with the stipulation that the known reference values of the test samples are withheld from the method developers by the testing authority – a procedure also known as *blind validation*.

12.4.3 Cross validation

Unlike test set validation methods, cross-validation methods attempt to validate a model using the calibration data only, without requiring the preparation and analysis of an additional test set of samples. This involves the execution of one or more internal validation procedures (hereby called *subvalidations*), where each subvalidation involves three steps:

- removal of a subset of the calibration samples
- use of the remaining calibration samples to build a series of calibration models of different complexities
- subsequent application of the models to the removed samples.

After all of the subvalidations are performed, the RMSEP metrics obtained from each of them are combined to generate a composite RMSECV value:

$$RMSECV = \sqrt{\frac{\sum_{i=1}^{N} (\hat{y}_{i,CV} - y_i)^2}{NCV}} \tag{12.53}$$

where $\hat{y}_{i,CV}$ is the property value for test set sample i that is estimated in the subvalidation procedure in which the model data did not contain sample i, and NCV refers to the total number of removed samples over all subvalidation experiments.

Although every subvalidation experiment draws from the same set of calibration data, each individual subvalidation involves the application of a model to a set of data *not* used to build the model. Consequently, cross validation tends to generate more realistic results than the model fit (RMSEE, Equation 12.11).

Cross-validation methods differ in how the sample subsets are selected for the subvalidation experiments. Several methods that are typically encountered in chemometrics software packages are listed below

- *Custom subset:* a single subvalidation is done, where a manually-selected subset of the calibration data is selected.

- *Leave-one-out:* a series of exactly N subvalidations are done, where each sample in the calibration data is selected once.
- *Random:* a prespecified number of subvalidations (A) are done, where a random selection of a prespecified number of samples ($N_{CV} < N$) are removed from the calibration data.
- *Contiguous block:* a prespecified number of subvalidations (A) are done, where a contiguous block of calibration samples are removed.
- *Venetian blinds:* a prespecified number of subvalidations (A) are done, by removing every Ath sample in the calibration data, where the A different subsets result from starting at sample 1, sample 2, … sample A.

The factors that influence the optimal cross-validation method, as well as the parameters for that method, are the number of calibration samples (N), the arrangement order of the samples in the calibration data set, whether the samples arise from a design of experiments (DOE, Section 12.2.6), the presence or absence of replicate samples, and the specific objective of the cross-validation experiment. In addition, there are two traps that one needs to be aware of when setting up a cross-validation experiment.

- The 'external subset selection trap': This occurs when the selected subset for a particular subvalidation is not representative of the samples in the rest of the data. This usually leads to overly *pessimistic* cross-validation results, because the selected subset is outside of the space of the remaining samples.
- The 'replicate sample trap': This occurs when replicates of the same physical sample are present in both the selected subset and the rest of the samples. This usually leads to overly-*optimistic* cross-validation results, because 'representatives' of the same samples are present in both the model and test sets.

Table 12.9 summarizes the utility of the different cross-validation methods mentioned above for several different types of data, including time-series data, batch data, and data from designed experiments. In general, the leave-one-out method is the easiest to use, the Venetian blinds and contiguous block methods are rather easy and quick, and the custom subset method is the most flexible. For small data sets, the leave-one-out method is generally favored, unless the data is generated by a design of experiments, in which case one or more custom subsets should probably be used to avoid the external subset selection trap. The random method can be thought of as a quicker alternative to the leave-one-out method for large data sets, where the user specifies both the number of samples to extract for each subvalidation and the total number of subvalidations to perform. It should also be mentioned that the venetian blinds and contiguous block methods differ from the random method only if the data profiles in the calibration data are arranged in a specific order (i.e., by time, or by a known attribute).

For time-series data, the contiguous block method can provide a good assessment of the temporal stability of the model, whereas the venetian blinds method can better assess nontemporal errors. For batch data, one can either specify custom subsets where each subset is assigned to a single batch (i.e., 'leave one *batch* out' cross-validation), or use Venetian blinds or contiguous blocks to assess within-batch and between-batch prediction errors, respectively. For blocked data that contains replicates, one must be very careful with the venetian blinds and contiguous block methods to select parameters such that the replicate sample trap and the external subset traps, respectively, are avoided.

It should be mentioned that another validation technique, called leverage correction [1], is available in some software packages. This method, unlike cross validation, does not involve splitting of the calibration data into model and test sets, but is simply an altered calculation of the RMSEE fit error of a model. This alteration involves the weighting of the contribution of the root mean square error from each calibration

Table 12.9 *The properties of different cross-validation approaches, for different data analysis scenarios*

		Cross-Validation Method				
		Venetian Blinds	Contiguous Blocks	Random Subsets	Leave-One Out	Custom
Data set properties	General Properties	+ Easy + Relatively quick	+ Easy + Relatively quick	+ Easy – Can be slow, if m or number of iterations large – Selection of subsets unknown	+ Easiest! (Only one parameter) – Avoid using if N > 20	+ Flexible – Requires time to determine/ construct cross validation array
	Small data sets (<~20 objects)			+ OK, if many iterations done	+ Good choice …. – …. unless designed/DOE data	+ often needed to avoid the *external subset selection trap*
	randomly-distributed objects	+ Good choice	+ Good choice	+ Good choice – Can take a while with large N, many iterations	+ OK, but …. – Can take a while with large N	
	time-series data	+ Useful for assessing ***NON-temporal*** model errors – Can be optimistic with low number of data splits	+ Useful for assessing ***temporal*** stability of model			
	Batch data	+ Useful for assessing predictability ***within*** batches/parts of batches	+ Useful for assessing predictability ***between*** batches/ parts of batches			+ Can manually select "batch-wise" test sets
	Blocked data (replicates)	– Beware the *replicate sample trap* (optimistic results)!	+ Good way to avoid *replicate sample trap* – Beware the *external subset selection trap*!	+ Can use to avoid *replicate sample trap* (high number of splits, high iterations preferable)	– overly optimistic results, due to *replicate sample trap*	
	Designed Experiment (DOE) data	– Dangerous, unless object order is randomized	– Dangerous, unless object order is randomized		– Not recommended (*external subset selection trap*)	+ often needed to avoid the *external subset selection trap*

Note: + refers to a favorable property; – refers to an unfavorable property.

sample by that sample's leverage in the calibration set. The leverage is related to the Hotelling T^2 statistic (Equation 12.21), and reflects the 'extremeness' of the sample within the model space.

12.5 Outliers

12.5.1 Introduction to outliers

An outlier can be defined as 'any observation that does not fit a pattern'. In a typical PAT calibration problem, one can encounter three different types of outliers:

- an *x-sample* outlier, based on the sample's analytical (spectral) profile
- an *x-variable* outlier, based on the behavior of a single x variable relative to the other x variables, and
- A *y-sample* outlier, based on the sample's property value(s) (for regression problems only).

It is very important to note that the term 'outlier' does *not* imply 'incorrect'. An outlier could be caused by an error or an incorrect action, but it could just as easily be caused by a real phenomenon that is relevant to the problem.

Outliers demand special attention in chemometrics for several different reasons. During model *development*, their extremeness often gives them an unduly high influence in the calculation of the calibration model. Therefore, if they represent erroneous readings, then they will add disproportionately more error to the calibration model. Furthermore, even if they represent informative information, it might be determined that this specific information is irrelevant to the problem. Outliers are also very important during model *deployment*, because they can be informative indicators of specific failures or abnormalities in the process being sampled, or in the measurement system itself. This use of outlier detection is discussed in the Model Deployment section (12.10), later in this chapter.

12.5.2 Outlier detection and remediation

The fact that not all outliers are erroneous leads to the following suggested practice of handling outliers during calibration development: (1) detect, (2) assess and (3) remove if appropriate. In practice, however, there could be hundreds or thousands of calibration samples and x variables, thus rendering individual detection and assessment of all outliers a rather time-consuming process. However, time-consuming as it may be, outlier detection is one of the most important processes in model development. The tools described below enable one to accomplish this process in the most efficient and effective manner possible.

The most obvious outliers in a calibration data set can be detected by simply plotting the data in various formats. In the case of process spectroscopy, one can simply generate an overlaid plot of multiple spectral profiles as a function of wavelength. Such plots can effectively reveal highly unique samples (spectral profiles), which could originate from an empty or badly-fouled sample cell, or a poorly-functioning instrument. The same plots could also indicate highly unique spectral variables that, for example, exhibit high noise or erratic behavior for all samples In all cases, though, prior knowledge of the analyzer and its application is required to assess the root cause(s) of the outlier(s), and subsequent action (keep or remove?). Similarly, y-sample outliers can be detected by simply plotting the y values for all samples in the data set as a function of sample number, time, or other attribute. Subsequent assessment of any y-sample outliers that are revealed can involve several different information sources, such as production records for the samples, the reference analytical records for the samples, and prior knowledge of the standard deviation of the reference analytical method.

Once the obvious outliers are addressed through simple plotting of the data, more subtle outliers can be detected using some of the same modeling methods that were discussed earlier: namely, PCA and PLS [15,16]. Because the PCA method only operates on the x data, it can be used to detect x-sample and x-variable outliers, using the PCA-estimated scores ($\hat{\mathbf{T}}$), loadings ($\hat{\mathbf{P}}$), and x residuals ($\hat{\mathbf{E}}$) (see Section 12.2.5). More specifically, the Hotelling T^2 statistic (Equation 12.21) and the Q residual statistic (Equation 12.22) can be used to assess objects that are 'extreme' *within* the model space and *outside* of the model space, respectively. Figure 12.27A shows the T^2 and Q statistics for the samples in a process spectroscopy calibration data set, after the obvious outliers were removed as described above. The dashed lines represent the 95% confidence

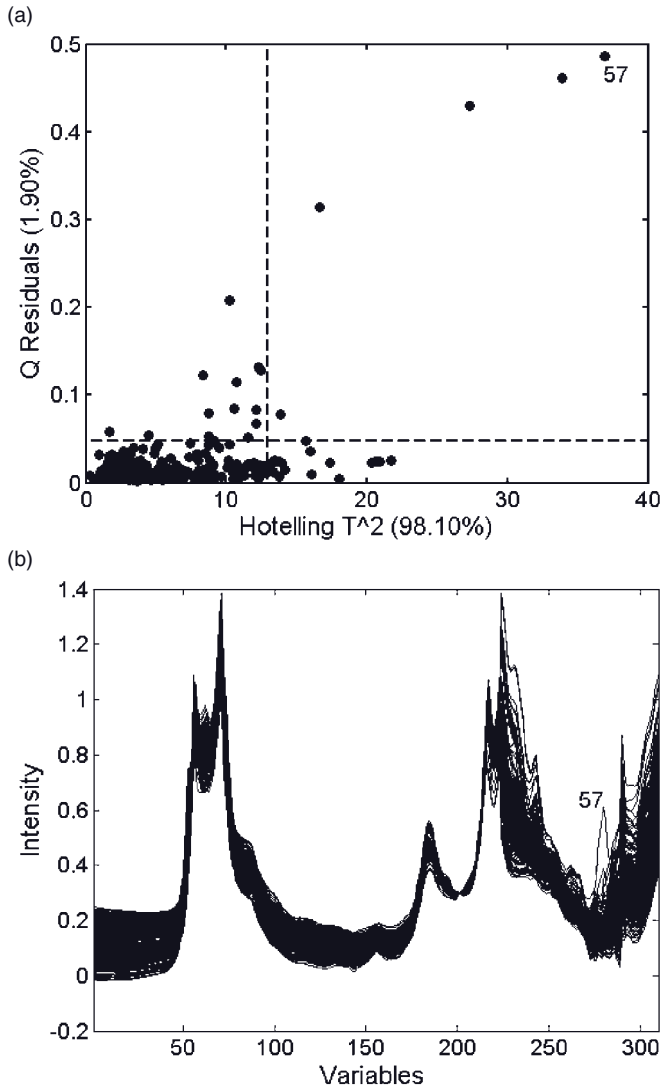


Figure 12.27 *(A) Scatter plot of the Hotelling* T^2 *and* Q *residual statistics associated with the samples in the process spectroscopy calibration data set, obtained from a PCA model built on the data after obvious outliers were removed. The dashed lines represent the 95% confidence limit of the respective statistic. (B) The spectra used to generate the plot in (A), denoting one of the outlier samples.*

limits on the T^2 and Q values, based on the assumption of a normal distribution of these statistics for the calibration data. In a normally-distributed population, one would expect 5% of the observations to exceed the 95% confidence limit. Therefore, when doing outlier detection during calibration development, one is specifically searching for observations that either greatly exceed one of these limits, or significantly exceed both limits. In this example, note that there are five samples that lie above the 95% confidence limit for both T^2 and Q, indicating that they might be outliers. In the most extreme case, that of sample 57, this appears to be caused by the presence of a unique absorption band at variable 280 (Figure 12.27B).

It should be noted that Equations 12.21 and 12.22 refer to the T^2 and Q statistics for *samples* in the x data, not for *variables*. However, the T^2 and Q values for the *variables* in the x data are calculated in a very similar manner. The T^2 values for x variables are calculated using the PCA *loadings*, rather than the scores:

$$T_m^2 = \hat{\mathbf{p}}_m diag(\lambda)\hat{\mathbf{p}}_m^t \tag{12.54}$$

where $\mathbf{p}_m$ are the PCA loadings of the mth variable for all A PCs. The Q residual values for the x variables in the data set are calculated using the spectral residuals ($\hat{\mathbf{e}}_p$):

$$Q_m = \hat{\mathbf{e}}_m \hat{\mathbf{e}}_m^t \tag{12.55}$$

where $\mathbf{e}_m$ are the PCA residuals for variable m over all N samples. In contrast to the sample residuals (Equation 12.22), note that the calculation of the x-variable residuals involves taking the sum of squares of spectral residuals over *samples*, rather than over variables.

Figure 12.28 shows a scatter plot of the Q and T^2 values for the *variables* in the process spectroscopy data set, after the removal of some of the sample outliers identified above. Note that several variables have high Q values, high T^2 values, or both. Some of these variables, including variables 56 and 71, refer to spectral regions of high absorbance, where a very low level of light reaches the detector, which often leads to noisy and/or nonlinear responses. However, this was *not* the case for all of the variables identified as outliers in this plot: some of them (including variables 285 and 286) coincided with wavelengths associated

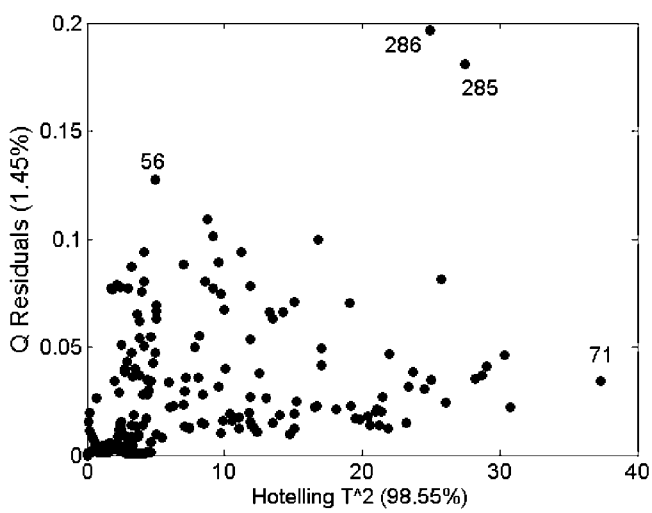


Figure 12.28 *Scatter plot of the Hotelling* T^2 *and Q residual statistics associated with the variables in the process spectroscopy calibration data set, obtained from a PCA model built on the data after obvious outliers were removed. Specific outlier variables are denoted.*

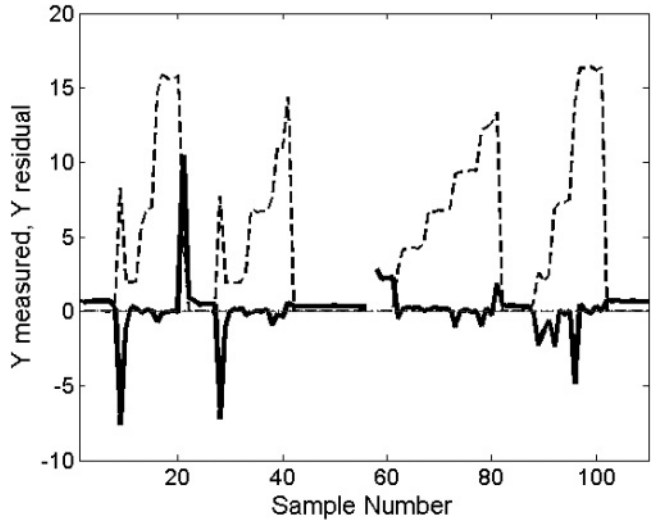


Figure 12.29 *Time-series plot of the y-residuals obtained from a PLS model developed using the process spectroscopy calibration data set (solid line), after removal of sample and variable outliers as discussed earlier. The measured y-values (dashed line) are also provided for reference.*

with the absorption of an interfering component that is known to 'foul' the sample cell windows over time. As a result, the removal of these variable outliers would enable the method to be less-affected by this 'fouling' material [19].

Just as PCA can be effective for detecting subtle x-variable and x-sample outliers, PLS or PCR can be effective for detecting y-sample outliers. This is mainly done through the use of the y residuals ($\hat{\mathbf{f}}$) (see Equation 12.44). The squares of the individual N elements of the y-residual vector ($\hat{\mathbf{f}}$) can then be observed to detect the presence of y-sample outliers. In a manner similar to that described above for Q and T^2 values, the squares of each of the elements in $\hat{\mathbf{f}}$ can be compared to the 95% confidence level of these values in order to assess whether one or more samples might be outliers due to their y data.

Figure 12.29 shows a time series plot of the y residuals that were obtained from a PLS model that was built using the same analyzer data after the removal of all outliers identified above, along with the measured y values. There are two samples (9, 28) with large negative y residuals that, based on knowledge of the sampling protocol used to collect the data, are likely caused by errors in the y-data collection system. However, note that many of the other samples that have large positive or negative y residuals (including 21, 81 and 96) occur when the measured y value is changing. These times correspond to transitions between product grades, when the sample composition is known to change quickly with time. It is also known from the sampling protocol used to collect this calibration data that the measured y values could be subject to significant errors at such times. As a result, it could very well be appropriate to exclude these samples from the calibration data.

12.6 Calibration Strategies in PAT

'Garbage in, garbage out': a ubiquitous phrase in chemometrics. As much as academics might debate the subtleties of different model building tools and methodologies, the difference between success and failure for chemometrics in PAT more often boils down to the quality of the data that is fed into the modeling

methods. This, in turn, is dictated by the specific *strategy* that is used to collect the calibration data. A given PAT problem could allow access to many different strategies, or be restricted to only a few options due to chemistry, physics, economics, safety issues, regulatory issues, or even politics! Furthermore, inverse modeling methods (MLR, PLS) have quite different experimental requirements than direct modeling methods (CLS, ELS). The purpose of this section is to address the critical attributes of calibration strategies, list some common types of strategies in PAT, and discuss how strategy selection might depend on the type of chemometric modeling being considered.

12.6.1 The 'calibration strategy space'

In any chemometrics application, there are two critical attributes of calibration data that are desired:

- *accuracy:* data that has a minimum of random or systematic noise
- *relevance:* data that is sufficiently representative of the data that the model will ultimately be applied to.

Of course, the ultimate goal is to generate calibration data that is both highly accurate and highly relevant. However, in PAT this can be a rather difficult goal, as it seems that one often trades off accuracy for relevance, and vice versa. For example, consider two of the most common calibration strategies in PAT:

1. Use of on-line spectra collected on the field analyzer, with reference values obtained from lab analyses of extracted process samples, and
2. Use of laboratory-synthesized calibration standards, analyzed on a laboratory version of the field analyzer.

Strategy 1 has the advantage of using actual process samples (high relevance), but suffers from possible errors in the reference analysis due to sampling issues or high reference method variability (low accuracy). Strategy 2 has the advantage of well-controlled preparation and characterization of standards (high accuracy), but potentially suffers from the fact that lab-collected spectra of lab-prepared mixtures are not sufficiently representative of field-collected spectra of process material (low relevance). I like to refer to this condition as the *calibration sampling paradox* of PAT.

A useful means for discussing this paradox, as well as representing any calibration strategy, is the *calibration strategy space*, which is a conceptual plot of relevance vs. accuracy. Figure 12.30 is a plot of this space, where the two calibration strategies mentioned above (1 and 2) are represented. Note that, although these two common strategies are in the upper left and lower right regions of this space, it is desired in PAT to develop a strategy that is as close as possible to the upper right portion of this space: the high relevance/high accuracy region. To do this, one must first understand the 'natural' forces that tend to push calibration strategies towards the undesirable lower left part of the space. These include, but are not limited to, chemical instability of synthesized standards, poor reference analysis protocol, raw material variations, nonrepresentative process sampling and instrument drift.

Despite these forces, it is often the case that strategies, such as 1 and 2 discussed above, can be modified in various manners to migrate them towards the more desirable upper right region of the space. For example, strategy 1, which suffers from poor accuracy, could be modified so that reference analyses are obtained from an accompanying (perhaps temporary) on-line analyzer with carefully optimized sampling protocol. Likewise, strategy 2, which suffers from poor relevance, could be changed so that the synthesized standards are injected directly into the field-installed process analyzer. There are many different ways in which the calibration strategy can be moved towards the favorable upper right corner of the calibration strategy space. These depend on the conditions and constraints of the application, and rely heavily on prior knowledge of the

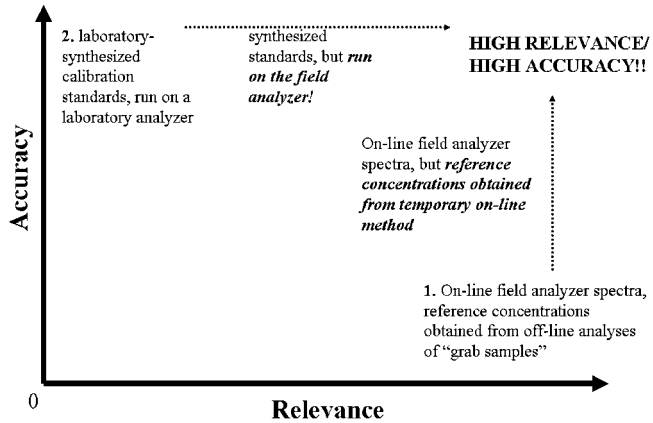


Figure 12.30 *Illustration of the calibration strategy space, where any calibration strategy can be mapped by its relevance (x-axis) and it accuracy (y-axis). In this figure, several calibration strategies are mapped onto this space.*

sample chemistry and physics, process dynamics, and the analyzer system involved, as well as the specific analytical objectives of the problem. However, a common thread in many such strategy improvements lies in the specification and accurate execution of optimal *sampling protocols*, for both the on-line analyzer and any grab samples that are required by the strategy.

12.6.2 Strategies for direct versus inverse modeling methods

In Section 12.3.2, the fundamental differences between direct and inverse modeling methods were discussed. As will be discussed here, this distinction is *not* just a convenient means for classifying quantitative regression methods, but has profound implications regarding calibration strategy and supporting infrastructure.

Recall that inverse methods (MLR, PCR, PLS) only attempt to explain the property values (**C**) in terms of the analyzer responses (**X**). These methods are particularly convenient and flexible for model building, because **C** can refer to any property or properties (not just a concentration) and not all possible contributions to **X** need to be known and characterized. However, it tends to be rather demanding with respect to the calibration data collection strategy. In order to support inverse methods, this strategy must generate matching **X** and **C** data that is not only sufficiently accurate and relevant, but also covers all sample states for which the on-line analyzer is expected to operate effectively. In practical applications, this often leads to the need to patiently collect and process a plethora of data over long periods of process operation, or even the execution of designed experiments on the process to achieve evasive yet relevant process sample states. Strategies that include such elements can be rather time-consuming and costly, and the subsequent data operations often require the handling of very large data sets, which often contain many outliers.

In contrast, direct methods (CLS, ELS and variations thereof) attempt to explicitly model the complete analyzer (**X**) response of a given sample. This can be rather demanding for model building, as one must explicitly compile a *basis set* of vectors (or 'basis spectra') that are sufficient for explaining the analyzer response of any possible process sample. However, there is a considerably wider range of calibration strategies that can effectively support direct methods, as there are many different ways in which these basis spectra could be obtained. A particularly simple and cost-effective strategy, which has long been used for open-path FTIR spectroscopy, is to compile this set from a previously collected library of spectra of the components

that are known or suspected to be present in the process samples. On the other extreme, one could use a more expensive and elaborate extension of the strategy mentioned above for inverse modeling methods: namely the collection of a matching set of **X** and **C** data of standards that cover the expected process variability, but where **C** contains the values of *all* known analytes in the standards (not just the analyte of interest). If physical and economic conditions allowed for this elaborate strategy, it would enable least-squares 'unmixing' of the spectra to yield estimates of the pure component spectra to use in the basis set (Equation 12.36). There are also many other strategies that fall 'in between' these two in terms of complexity and cost: for example, some component's contributions could be based on library spectra, some based on lab-collected spectra of pure components or mixtures, and some based on 'unmixing' using a matching set of **X** and **C** data from mixture samples.

Extended-CLS (ELS)-type modeling, can be supported with the same calibration strategies as those mentioned above for direct methods, with an additional consideration: one must also collect data that can be used to estimate unknown contributions to the spectrum. Such characterization of the unknown contributors often requires the collection of sufficiently relevant process data, or at least reasonable *a priori* estimates of expected interference effects, such as baseline effects and water vapor interferences, in the process spectra.

12.6.3 Hybrid strategies

Although there are many different calibration strategies that might be appropriate for a given problem, there is no 'rule' that prohibits one from attempting to apply more than one of these strategies to the same problem. In fact, combining calibration data collected from two different strategies can generate a 'hybrid' approach that is closer to the preferable 'high accuracy/high relevance' region of the strategy space.

In one specific example [15,16,19], all of the calibration data for an on-line analyzer had been obtained from on-line spectra, with reference concentration values obtained from inferential process models. Although the spectral data was highly relevant, the low accuracy of the reference data was ultimately limiting the effectiveness of many of the deployed calibrations. In response to this situation, it was decided to engineer and construct a special calibration apparatus that enabled the direct injection of carefully-prepared calibration mixtures into the field analyzer as-installed. The thought was that the high-accuracy calibration data generated by this apparatus could be 'mixed' with the existing high-relevance calibration data to improve the accuracy of the subsequent analyzer calibrations generated from the data. Of course, the success of this approach is contingent upon the new calibration data being sufficiently *relevant* to the calibration problem.

Regarding relevance, the spectral 'miscibility' of the data obtained from these two different sources can be readily observed by doing a PCA analysis of the combined spectral data. The scatter plot of the first two PC scores obtained from PCA of such a data set for one of the process analytes is shown in Figure 12.31a. Note that there is considerable common space for the two data sources in the PC1/PC2 space, and there are some regions of this space where only samples from the old calibration strategy lie. A similar pattern is observed in the later PCs of this model. This result indicates that the on-line spectra contain some unique information, but that the on-line and injected-standard spectra are generally quite similar.

With these encouraging results, it was attempted to build a PLS predictive model for the analyte using this combined data set. The first interesting result of this work was that the newer, high-accuracy calibration data enabled more effective detection of y-sample outliers in the older data, thus enabling more extensive outlier screening. Secondly, it was found that a good predictive model could be generated using this data (Figure 12.31b) without a large increase in PLS model complexity (i.e., number of PLS latent variables). Most importantly, though, subsequent deployment of this updated model resulted in more effective use of the analyzer for control purposes.

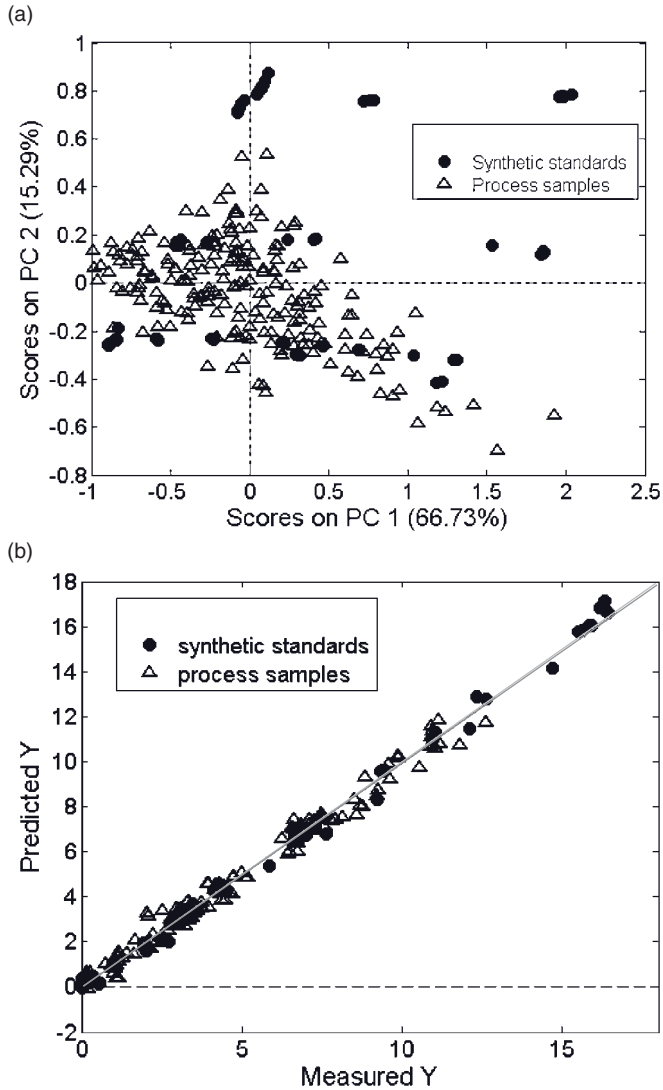


Figure 12.31 *Illustration of a hybrid calibration strategy. (A) Scatter plot of first two PCA scores obtained from a process analytical calibration data set containing both synthesized standards (circles) and actual process samples (triangles). (B) Results of a PLS regression fit to the property of interest, using all of the calibration samples represented in (A).*

12.7 Sample and Variable Selection in Chemometrics

12.7.1 Sample selection

Several PAT calibration strategies, especially those that are intended to support inverse calibration methods, rely heavily on data that is routinely collected from the deployed analyzer, as opposed to data collected from carefully designed experiments. Such data, often called 'happenstance data', can be very inexpensive,

because it requires no special investment or expenditure for its generation. However, such data is notoriously messy, in that it contains a lot of outliers. Furthermore, even if the samples sufficiently cover the required space of sample states, they might be poorly distributed in this space, in that some sample states might be better represented than others. In such a situation, any model built from the data will tend to perform well at the over-represented states and poorly at the under-represented states. The goal of sample selection methods is to determine a subset of these samples that span the same space, but are more evenly distributed in the space.

There are several methods that can be used to select well-distributed calibration samples from a set of poorly-distributed samples. One simple method, called *distance-based selection*, selects a prespecified number of samples that are on the edge of the data space. To do this, the sample farthest from the data mean is first selected, and then subsequent samples are selected based on maximum Euclidean distance from both the mean *and* the previously selected sample(s). It should be noted, though, that this method does not select samples that are in the interior of the data space, which can be problematic if the modeling method must characterize nonlinear effects in the data. An alternative selection method, which is based on the *D*-optimal experimental design [95], selects a subset of a pre-specified number of samples that maximizes the determinant of $\mathbf{X}_{\mathbf{sel}}^{\mathbf{t}}*\mathbf{X}_{\mathbf{sel}}$, where $\mathbf{X}_{\mathbf{sel}}$ contains the spectral data for the selected samples only. Recall that, for a two-dimensional space, the absolute value of the determinant is equal to the area of the parallelogram defined by the vectors in $\mathbf{X}_{\mathbf{sel}}$. With this in mind, the *D*-optimal method strives to select the subset of vectors (spectra) that generate the largest space. Like the leverage-based selection mentioned above, *D*-optimal tends to select samples on the edge of the data space, and consequently it might be necessary to include samples close to the data mean in order to better characterize nonlinearities. Either original *x* variables or PCs can be used as input, and the latter is preferred when there are many *x* variables, so that lengthy computation times can be avoided.

An additional sample selection method uses HCA to determine a prespecified number (*N*) of natural groupings of the samples, with subsequent selection of a single 'representative' from each group [15,16,96]. This method can use original spectral *x* variables or PCA scores (**T**) for a specified number of PCs as input to the algorithm. The selection of each representative is often done based on the maximum distance from the overall data mean, but it can also be done based on distance from the mean of the sample's cluster. This method has the advantage of selecting samples that are distributed throughout the sample space, rather than just on the edges of the space. However, this method can be rather time-consuming, especially if original *x* variables are used as input. Also, the value of *N* that is selected for this method should exceed the number of natural groupings of the samples. In practice, it is better to err on the side of specifying an *N* that is larger than necessary, as the goal is to select samples that are well-distributed in the space. Also, if PCs are to be used as input for this selection method, one must be careful to use sufficient PCs to explain all relevant variability in the spectral data.

An example of the usage of these sample selection methods on actual data is shown in Figure 12.32. In this case, the original data set contained 200 samples, and each of the three selection methods discussed above was used to select 10 samples from this data set. The distance-based selection method used 290 original *x* variables as input, whereas the HCA-based method and the *D*-optimal method used the first four PC scores (which explain 98.97% of the *x*-data variation) as input. Each of the plots in Figure 12.32 shows the samples in the space defined by the first two PCs only, even though there are four significant PCs in this data. In this specific example, note that the distance-based and HCA-based methods tended to select samples that were more distributed in this space, whereas the D-optimal method tended to select the more extreme samples in this space.

12.7.2 Variable selection

This subject has already been discussed, in the context of the MLR regression method, which for most PAT applications (where the number of *x* variables greatly exceeds the number of calibration standards) *requires*

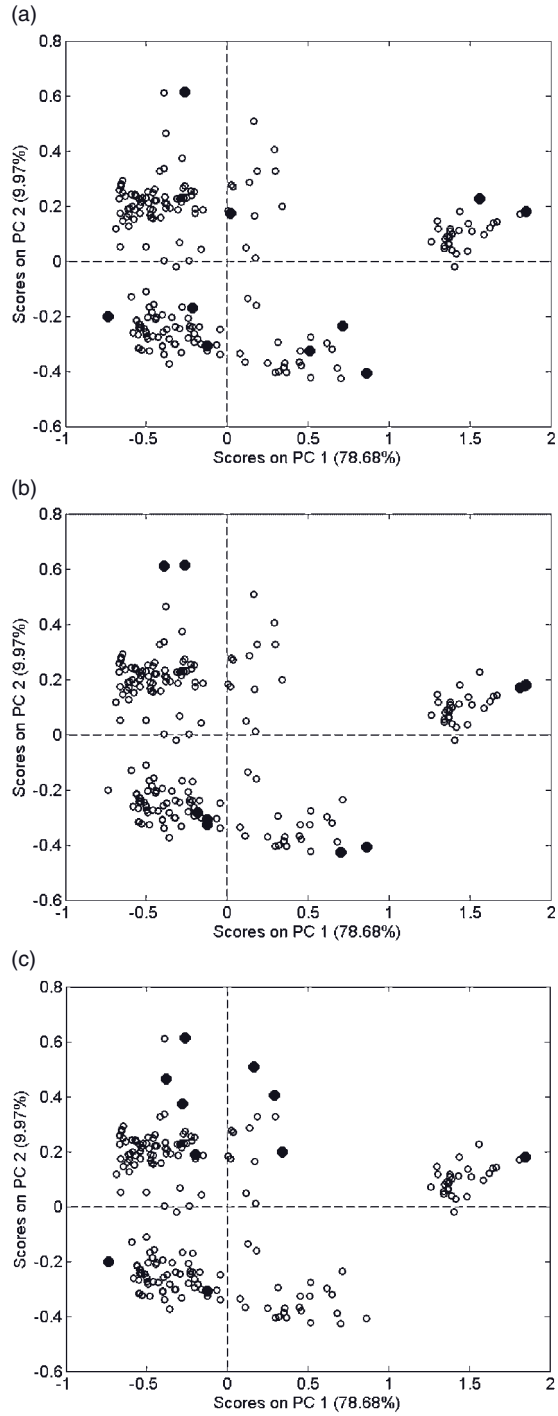


Figure 12.32 *Sample selection: scatter plots of the first two PCs obtained from a process spectroscopy dataset containing 200 samples, indicating the subset of 10 samples selected using the leverage-based method (A), the* D*-optimal method (B), and the HCA-based method (C).*

the selection of a variable subset. However, for those using factor-based inverse calibration methods (PLS, PCR) and direct calibration methods (CLS), can variable selection be useful? After all, with careful planning, these methods should effectively account for covariance between x variables as well as responses that interfere with the analyte/property of interest.

I suggest that that answer to the above question is 'yes', for the following reasons.

- For inverse calibration methods, the fact that reference data (y) is never noise-free in practice allows irrelevant variation in the x variables to find its way into the calibration model.
- Including irrelevant variation in x leads to unnecessary 'burden' in either explaining the x variation (inverse methods) or to specifying sufficient basis vectors (direct methods) during development of the model, which can lead to critical omission/errors during model building.
- Removing irrelevant x variables is an effective means of *simplifying* the calibration model, thus rendering it more stable, less sensitive to unforeseen disturbances, and easier to deploy and maintain.

With these arguments as a backdrop, I will review some empirical variable selection methods in addition to the prior knowledge-based, stepwise and all possible-combinations methods discussed earlier in the MLR section (Section 12.3.2).

A relatively simple variable selection method [97] involves a series of separate linear regressions of each x variable to the property of interest. The relevance of an x variable is then expressed by the ratio of the linear regression slope (b) to the variance of the elements in the linear regression model residual vector (**f**), called the 'signal-to-noise ratio' (S/N). The user can then select variables with the highest S/N values, based on a threshold value. Although this method has been shown to be effective for PAT applications [15,16], its main limitation is that it only considers *univariate* relationships between a single x variable and the property of interest. The fact that chemometrics is being considered for building the model in the first place is due to the suspicion (or knowledge!) that a *multivariate* relationship between several x variables and the property of interest exists.

Some multivariate extensions to the stepwise method discussed earlier in the MLR section are commonly called the 'interval methods', such as 'interval-PLS' (i-PLS) [98]. The i-PLS method involves the sequential selection and retention of intervals of contiguous variables, where selection is based on the PLS model performance improvement that is obtained through the addition of the interval to the variable subset. The model performance is most often specified as the cross-validation error of a model built using the specified interval set. For a given interval set, PLS modeling is done using *all* x variables that are contained in the selected intervals. Interval selection can continue until a prespecified number of intervals are selected, or until the model performance does not significantly improve. Interval methods of selection can be applied to many different modeling methods, including PLS, PCR, and even classification methods such as PLS-DA.

The above paragraph describes the *forward* option of the interval methods, where one starts with no variables selected, and sequentially *adds* intervals of variables until the stop criterion is reached. Alternatively, one could operate the interval methods in *reverse* mode, where one starts using all available x variables, and sequentially *removes* intervals of variables until the stop criterion is reached. Being stepwise selection methods, the interval methods have the potential to select local rather than global optima, and they require careful selection of the interval size (number of variables per interval) based on prior knowledge of the spectroscopy, to balance computation time and performance improvement. However, these methods are rather straightforward, relatively simple to implement, and efficient.

A general alternative to stepwise-type searching methods for variable selection would be methods that attempt to explore as much of the possible solution space as possible. An exhaustive search of all possible combinations of variables is possible only for problems that involve relatively few x variables. However, it

does not take many x variables to make the scope of this problem intractable: for a problem with only 100 variables, there are 9.3×10^{157} possible combinations that would need to be tested. Fortunately, there are 'searching' algorithms that allow one to make an incomplete, yet efficient, search of all of these possibilities. One such algorithm, the genetic algorithm (GA) has seen considerable usage in chemometrics PAT applications for variable selection [99–101]. GA operates in the following manner.

1. A total number of variable subset candidates is specified (along with several other parameters), and an initial set of variable subset candidates is randomly generated.
2. For each variable subset, a regression model is generated.
3. Each of the regression models is evaluated for prediction ability, typically using cross validation.
4. Different pairs of variable subsets are 'mated' together, based on their prediction abilities assessed in (3) above. Those variable subsets that result in the best prediction abilities are most likely to 'mate' with one another, generating a new variable subset that uses some of the selected variables from one mate and some from the other. When this is done for all pairs of variable subsets, this process results in the generation of a new set of variable subset candidates.
5. Steps (2) to (4) are repeated, until a termination condition is met.

Termination of the algorithm can be based on several conditions, such as a prespecified number of iterations, or a prespecified level of agreement between the different selected subsets. Details on the GA method can be found in several references [99–101].

When the GA algorithm is terminated, one is presented with one or more variable subsets, along with the cross-validation error associated with each subset. At this point, there are several ways in which these results could be used to select the final subset of variables. One could select the union or the intersection of the variables that appear in all of these subsets, or simply the subset that generated the lowest cross-validation error. One could also use prior knowledge regarding the stability or reliability of certain variables to make the final selection. However, it should be noted that the GA algorithm starts with a randomly-selected set of variable subsets, and thus will generate different results with the identical x and y data when run more than once. As a result, it is often useful to run the GA several times on the same data, in order to obtain a consensus on the selection of useful variables.

A disadvantage of the GA method of variable selection is that it requires the specification of several parameters, which leads to a significant learning curve. It is also rather computationally intensive, especially when a large number of samples or variables are present in the data. However, significant time savings can be achieved through sample selection [Section 12.7.1] or by 'blocking' adjacent variables (which is appropriate only if the variables represent a continuous physical property, such as spectral wavelength). Furthermore, the GA method generally does a more thorough search of the solution space than the other methods mentioned earlier, lowering the likelihood of selecting a local optimum.

The variable selection methods discussed above certainly do not cover all selection methods that have been proposed, and there are several other methods that could be quite effective for PAT applications. These include a modified version of a PLS algorithm that includes interactive variable selection [102], and a combination of GA selection with wavelet transform data compression [25].

An important aspect of variable selection that is often overlooked is the hazard brought about through the use of cross-validation for two quite different purposes: namely (1) as an optimization criterion for variable selection and other model optimization tasks (including selection of the optimal number of PLS LVs or PCR PCs); and (2) as an assessment of the quality of the final model built using all samples. In this case, one can get highly optimistic estimates of a model's performance, because the same criterion is used to both optimize and evaluate the model. As a result, when doing variable selection, especially with a limited number of calibration samples, it is advisable to do an additional 'outer loop' cross-validation across the entire model

optimization process. This procedure, called *cross-model validation* [103,104], involves the specification of an additional validation sample set (not part of the sample population used to *optimize* the variable selection) to *evaluate* the different variable selections.

12.8 Troubleshooting/Improving an Existing Method

In PAT, chemometric modeling efforts are seldom a static endeavor. Once an initial model is deployed, it is invariably the case that the model needs to be improved, to accommodate changing conditions or requirements, or for continuous improvement purposes.

12.8.1 Method assessment

Any effort to improve a model's performance must start with a good assessment of its current performance. There are several items that can help provide this assessment.

- *Predicted versus actual values of the property of interest (y):* This should include a good assessment of the accuracy and reliability of the reference method used to generate the 'actual' values. However, it should be noted that such a reference method might not exist for a given PAT application once it is deployed in the field, thus requiring one to resort to other assessment methods.
- *Prediction residuals (Q) and Hotelling T2 statistics (Section 12.2.5):* These allow a sample-wise assessment of abnormality of the on-line-collected data. Furthermore, if they are available throughout the usage of the model, this would enable assessment of any temporal effects regarding such abnormalities, which might coincide with known process events.
- *Complaints regarding the method's performance:* These could reflect a wide range of method attributes, including accuracy, stability, repeatability and precision. The most useful observations tend to come from the true 'customers' of the method: for on-line analyzers these are usually the process operators and engineers that rely on the method for control purposes.
- *Performance specifications of the method:* It is important to recall the previously-stated requirements of the method, which could be expressed in terms of accuracy, precision, repeatability, up-time, and/or other attributes. It's also quite possible that these specifications have changed since the method was first put into service.
- *Performance evaluations of the reference method:* Documentation regarding routine QC of reference laboratory methods can be very relevant for model evaluation, if the methods were used to generate reference (y) data for model development.

With the above information in hand, it is then possible to consider different improvement schemes.

12.8.2 Model improvement strategies

Depending on the details obtained from the method assessment above, there are several possible actions that could be considered for improving the performance of an existing calibration model:

- *More calibration data:* Provided that it is sufficiently accurate and relevant, additional calibration data can be useful for improving the characterization of sample states in the required space, thus leading to better model estimates through reduced estimation error [see Figure 12.25].

- *Improved sampling protocol (better calibration data):* As mentioned previously (Section 12.6), one can make adjustments to the sampling or data collection protocol to generate calibration data with improved accuracy, improved relevance, or both. In PAT, such adjustments could include automated or better-controlled sampling, replicate sampling, or instrument hardware tuning/optimization/re-engineering.
- *Improved x-data pretreatment:* Newly discovered effects in the on-line analyzer data could be more effectively filtered out using a different pretreatment method. More effective pretreatment would then reduce the burden on the modeling method to generate an effective model.
- *Improved reference analysis:* If the method assessment reveals that the main source of model error comes from error in the reference analytical method used to generate the (y) calibration data, then efforts to improve the accuracy and precision of this method could prove to be very beneficial.
- *Local modeling:* In some applications, the use of a global model might not be optimal, because of the large number of interferences that must be accounted for (thus reducing the Net Analyte Signal, as discussed above). In such cases, improved model precision can be realized through the use of local models, where separate models are developed for different subsets of expected sample states. However, if this is done, it is critical to communicate the subsequent limitations of each of the local models to the customers of the methods, to avoid inappropriate usage when they are put into service.
- *More appropriate model structure:* This often takes the form of nonlinear models, which in some cases can be more effective at describing inherently nonlinear relationships that cannot be linearized by pre-processing. However, in most cases the marginal improvements in model fits obtained by such methods is outweighed by overfitting concerns and deployment logistics difficulties.
- *Revisit variable selection:* A profound expansion or shift in the relevant analyzer response space, or in the measurement objectives, could result in the need to reconsider the optimal set of variables to use for the calibration model. This action is usually considered when new or additional calibration data are obtained.
- *Slope/bias correction:* This method, which is really a 'postprocessing' of model outputs, is one of the simplest improvement methods, and can be quite effective in cases where temporal shifts in analyzer response are expected [105]. However, when used, it should be accompanied by a well-defined sampling and measurement protocol, in order to generate a sufficiently large population of time-localized standards that can be used to determine stable estimates of slope and bias correction factors.

12.9 Calibration Transfer and Instrument Standardization

Consider the following scenario, which is very common in PAT deployments.

- one would like to deploy the same type of analyzer at each of several geographical locations that run similar or identical processes
- a multivariate calibration method must be used to develop an effective method for the property of interest, and
- although the analyzers will be of the same make and model, their responses will not be *identical* to one another, due to small variations in their parts manufacturing and assembly, as well as differences in sampling interfaces and sampling protocols between the sites.

To minimize calibration development costs, it would be ideal to produce a calibration using data obtained on only one of the analyzers (the *master* instrument), and simply *transfer* this calibration to all of the other analyzers (the *slave* instruments).

There are several strategies that can be taken to allow calibrations to be transferred in such a scenario. If the response variations between the instrument systems do not significantly damage the ability to predict the property of interest, then it might be possible to simply use the identical calibration on all instrument systems. Even in the event that there are some response variations that affect prediction ability, it might be possible to reduce their impact by incorporating one of the many preprocessing methods discussed earlier (Section 12.3.1) into the prediction model. However, if neither of these options prove to be effective, then it is necessary to incorporate a calibration transfer or instrument standardization strategy to enable effective calibration deployments on multiple instruments. It should be noted that, as the complexity of the calibration model required to determine the property of interest increases, the likelihood that such a strategy will be required increases as well, as it becomes more likely that even the smallest response variations between instrument systems will affect prediction of the property of interest.

At this point, it is worthwhile to discuss the distinctions between the two strategies that are commonly used in these scenarios:

- *calibration transfer:* adjustment of a *calibration model* to allow its effective usage on a different measurement system
- *instrument standardization:* adjustment of the *instrument response* to make it look like it was obtained on the master instrument (i.e., the instrument used to generate the calibration data).

Although both of these strategies are used under similar circumstances, and have similar goals, they differ in the specific entity that is adjusted: the instrument data versus the calibration.

Before attempting to define an optimal strategy for a given application, it is critical to assess the specific nature of the inter-instrument variability, and attempt to understand the root causes of this variability. For process spectrometers, there are two general types of inter-instrument response variability that are encountered: (1) response magnitude (y axis) variability and (2) registration shift (x axis) variability. In spectroscopy, response effects can have many possible root causes, including source intensity changes, detector response changes, and sampling interface differences. These effects are best interrogated using intensity reference standards, such as neutral density filters of known transmittance for transmission spectroscopy, and reflectance standards for diffuse reflectance spectroscopy [106]. Registration shift effects can also have many root causes, and for spectrometers are often linked to subtle alignment/configuration differences in optical hardware. Assessment of these effects requires a quite different set of standards, namely those with sharp response features that allow interrogation of the absolute x axis registration. Examples of such standards for spectroscopy are etalons, atomic emission lamps, and rare earth oxide standards [107] for NIR spectroscopy.

Once the specific nature of inter-instrument variability is assessed, it is possible to develop a transfer strategy. An important part of this strategy involves the selection of an appropriate set of *transfer standards*. In general, these standards should:

- be chemically-stable: to avoid temporal response changes
- be relatively few: to limit costs of materials and deployment
- be easily machined/configured for usage with the analyzer's sampling interface
- be designed to avoid response variability from sampling artifacts, and
- have responses that allow interrogation of the inter-instrument differences *that affect determination of the property of interest.*

It is worth noting that these standards could be a subset of the same standards used to develop the calibration model for the property of interest. In this case, there are several sample selection strategies available for identifying the transfer standards from the complete set of calibration samples [107–111].

Concurrent with the selection of transfer standards is the selection of the optimal strategy: deciding whether to use calibration transfer or instrument standardization, assigning the 'master' and 'slave' instruments, and selecting a suitable transfer algorithm. Some commonly used algorithms for calibration transfer and instrument standardization are discussed below.

12.9.1 Slope/intercept adjustment

This method can be considered a calibration transfer method that involves a simple instrument-specific postprocessing of the calibration model outputs [108,113]. It requires the analysis of a subset of the calibration standards on the master and all of the slave instruments. A multivariate calibration model built using the data from the complete calibration set obtained from the master instrument is then applied to the data of the subset of samples obtained on the slave instruments. Optimal multiplicative and offset adjustments for each instrument are then calculated using linear regression of the predicted y values obtained from the slave instrument spectra versus the known y values.

During real-time operation, statistical f-tests can be used to determine whether the slope and bias correction calculated for a particular unknown spectrum are within an expected range [108]. If not, then a warning can be issued indicating that the current sample is not appropriate to apply to the model.

The advantage of the slope/intercept method is that it is rather simple to implement, in that it only requires two parameters for each slave instrument. In addition, it does not require the designation of specific transfer standards to define the transfer. However, it requires that the same subset of calibration samples be analyzed by all of the slave instruments, which can take a while if several slave instruments are involved. Furthermore, it assumes that the sources of inter-instrument response variability are completely linear in nature. Therefore, it might not be optimal if significant nonlinear differences, such as wavelength axis shifts between analyzers, are present. A final limitation is that it assumes that the nature of the inter-instrument variability is constant over all wavelength ranges, and over all possible sample compositions.

12.9.2 Piecewise direct standardization (PDS)

The PDS method [108,109,114,115] can be very effective for spectroscopic analyzers, and other analyzers that generate data on a continuous variable axis. In PDS, the responses of a set of transfer standards are obtained on both the master and the slave instrument (thus producing $\mathbf{X}_m$ and $\mathbf{X}_s$, respectively). It is then desired to obtain a transformation matrix $\mathbf{F}$ such that the spectra obtained on the slave instrument can be transformed into spectra that would have been obtained on the master instrument:

$$\mathbf{X_m} = \mathbf{X_s F} \tag{12.56}$$

Equation 12.56 indicates that the response that would have been obtained on the master instrument is simply a linear combination of the responses obtained on the slave instrument. However, in spectroscopy, it is very unlikely that the response at a specific wavelength on the master instrument should depend on any responses on the slave instrument that are at wavelengths far from the specified wavelength. As a result, the PDS method imposes a structure on the $\mathbf{F}$ matrix such that only wavelengths on the slave instrument that are within a specified range of the target wavelength on the master instrument can be used for the transformation.

An important advantage of the PDS method is its ability to address both intensity (y axis) variations and registration shift (x axis) variations, although some might argue that the wavelength shift variations are not addressed explicitly in the PDS method. In addition, it can account for situations where the nature of the inter-instrument response variability is not constant over all wavelength regions. Another advantage that is

not apparent from the above discussion is that PDS can be used effectively even in cases where the master and slave instruments have very different signal-to-noise properties or spectral resolutions. In fact, it has been demonstrated that, in cases where the master instrument is much less noisy than the slave instrument, that PDS standardization outperforms the 'brute force' method of developing a separate calibration using data from the slave analyzer [114].

Some disadvantages of the PDS method are its relative complexity, and the need to analyze the same set of transfer standards on all analyzers. It has also been shown that the PDS method has very limited effectiveness in cases where registration shift (x axis) variations are the dominant source of inter-instrument variation.

12.9.3 Generalized least squares (GLS) weighting

The GLS method was mentioned earlier, as a preprocessing method that down-weights multivariate directions in the data that correspond to known interfering effects. However, it can also be used in a calibration transfer context, where directions in the data that correspond to *instrumental differences* are down-weighted. The use of GLS weighting for calibration transfer is discussed in reference [116].

12.9.4 Shenk–Westerhaus method

One particular standardization method, patented by Shenk and Westerhaus in 1989 [112,117], has seen high utility in NIR food and agricultural applications. This method involves a wavelength axis shift correction, followed by an intensity correction. More details on the operation of this method can be found in the literature [112], but its basic operations involve the following:

- determination of correlations of each master instrument wavelength with a series of nominally adjacent slave instrument wavelengths
- estimation of the slave wavelength corresponding to each master wavelength, using nonlinear interpolation
- estimation of the slave responses at each of the wavelengths determined in the previous step, using interpolation
- calculation of the slope and intercept correction factors for each slave wavelength, via linear regression of the master responses and corresponding estimated slave responses.

Like other standardization methods, the effectiveness of this method depends greatly on the ability of the standardization samples to assess instrumental biases that affect prediction. In many cases, this method uses a set of sealed transfer standards that are relevant to the specific application. In one study [112], it was found that the effectiveness of this method depended greatly on whether the standardization samples sufficiently cover the spectral intensity ranges that are to be experienced in the on-line process data.

The advantages of this method are that it explicitly addresses both intensity shifts and wavelength axis shifts. In addition, it can handle cases where the nature of the inter-instrument variability varies with wavelength. It is also relatively simple to explain. Like the PDS method mentioned above, it requires the analysis of all transfer standards on all analyzers.

12.9.5 Other transfer/standardization methods

There are additional transfer/standardization methods that can be particularly useful for PAT applications. Some of these are discussed below:

- *Registration shifting methods:* These methods involve the use of wavelength standards to characterize global or localized shifts in the x axis (wavelength) registration between instruments, followed by an interpolation to reduce these effects. In cases where x axis registration shifts are the dominant source of response bias between instruments, these types of methods can be very useful by themselves, or in conjunction with other standardization methods. It should also be mentioned that, in such cases, other methods that do not explicitly handle registration shifts (i.e., PDS) can actually generate unwanted distortions in standardized spectra.
- *Orthogonal signal correction (OSC):* This method explicitly uses y (property or analyte) information in calibration data to develop a general 'filter' for removing any y-irrelevant variation in any subsequent x data [118]. As such, if this y-irrelevant variation includes inter-instrument effects, then this method performs some degree of calibration transfer. The OSC model does not explicitly handle x axis shifts, but in principle can handle these to some extent. It has also been shown that the piecewise (wavelength-localized) version of this method (POSC) can be effective in some cases [119].
- *Finite impulse response (FIR):* This method is particularly easy to use, in that it requires the designation of only a single 'standard' spectrum on the master instrument. Correction is done on each individual slave spectrum using a wavelength-localized version of multiplicative signal correction (MSC, section 12.3.1), where the MSC reference spectrum is the designated standard spectrum. This method can be effective for reducing response magnitude variations between instruments [120].
- *Multiway methods:* For analyzer data where a single sample generates a second order array (ex. GC/MS, LC/UV, excitation/emission fluorescence), multiway chemometric modeling methods, such as PARAFAC (parallel factor analysis) [121,122], can be used to exploit the 'second order advantage' to perform effective calibration transfer and instrument standardization.

12.10 Chemometric Model Deployment Issues in PAT

Model deployment logistics might not be academically interesting, but they are absolutely critical for project success. The most effective method in the world, developed using state of the art modeling methods, is worthless unless it can be deployed in an effective, safe and sustainable manner. Unfortunately, though, the 'deployment landscape' of chemometrics in PAT can vary widely between applications, and thus the details of model deployments can vary widely as well. Nonetheless, this section will attempt to provide a brief summary of the more common deployment issues that arise in PAT applications.

12.10.1 Outliers in prediction

It was mentioned in Section 12.5 that outliers encountered during method development can result in the generation of suboptimal calibration models. However, outliers encountered during method deployment can be equally, or more, damaging. This is the case because any predicted results obtained from the application of an empirical model to a sample state that was *not* represented in the calibration data are invalid. In many cases, attempts to apply a model to such inappropriate sample states result in grossly (and obviously) inaccurate prediction results. However, in more damaging cases, such inappropriate predictions generate plausible, while still invalid, results, leading to incorrect control action and very costly outcomes.

As a result, it is critical to evaluate process samples in real-time for their appropriateness of use with the empirical model. For models built using PCA, PLS, PCR and other factor based methods, the mechanism for such a 'model health monitor' is already 'built into' the model. Equations 12.21 and 12.22 can be used to calculate Hotelling T^2 and Q residual statistics for each process sample using the sample's on-line analytical profile ($\mathbf{x_p}$) and the loadings ($\mathbf{P}$) and scores ($\mathbf{T}$) of the model. An abnormally high T^2 value would indicate

that the sample has a response that is within the model space, but in a region far from the responses of the calibration standards. In contrast, an abnormally high Q residual value would indicate that the sample has a significant portion of response that is outside of the model space. In either case, though, any model prediction result(s) generated from that sample are invalid. In practice, the T^2 and Q metrics are often ratioed to a specific confidence level (typically the 95% confidence level) of these metrics, where this level is calculated using either the metrics calculated from the calibration data or from a separate set of test data that is known to be appropriate for application to the model. The resulting metrics are commonly called *reduced* T^2 and *reduced* Q, and values of these greater than one indicate an abnormal analyzer response.

The use of T^2 and Q prediction outlier metrics as described above is an example of a *model-specific* 'health monitor', in that the metrics refer to the specific analyzer response space that was used to develop a PLS, PCR or PCA prediction model. However, many PAT applications involve the deployment of multiple prediction models on a single analyzer. In such cases, one can also develop an *analyzer-specific* health monitor, where the T^2 and Q outlier metrics refer to a wider response space that covers all normal analyzer operation. This would typically be done by building a separate PCA model using a set of data that covers all normal analyzer responses. Of course, one could extend this concept further, and deploy multiple PCA health monitor models that are designed to detect different specific abnormal conditions.

The T^2 and Q metrics are the 'highest-level' metrics for abnormality detection. For any given prediction sample that produces a high or interesting T^2 or Q value, one can also obtain the contributions of each analyzer variable to the T^2 and Q value, aptly named the *t*- and *q-contributions*, which are defined below:

$$\mathbf{t}_{\text{con}} = \hat{\mathbf{t}}_{\mathbf{p}} * \lambda^{-1/2} \times \mathbf{P}^{\mathbf{t}} \tag{12.57}$$

$$\mathbf{q}_{\text{con}} = \hat{\mathbf{e}}_{\mathbf{p}} \tag{12.58}$$

These contributions can be particularly useful for assessing the *nature* or *source* of a specific process sample's abnormality.

Figure 12.33 shows a time-series plot of the reduced T^2 and Q values of an analyzer-specific PCA health monitor for an actual PAT application, before and after a calibration update. There are several interesting

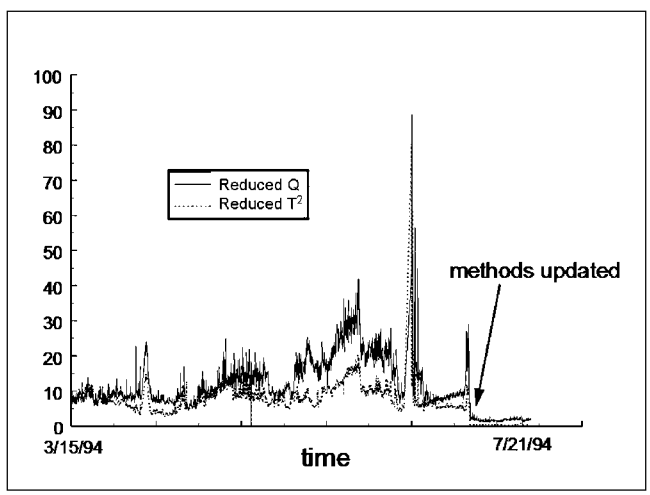


Figure 12.33 *Time series plot of reduced* T^2 *and* Q *statistics associated with the application of an analyzer-specific PCA model to on-line analyzer data – covering a period of approximately 4 months.*

items of note in this plot. First, note that the reduced T^2 and Q values are generally greater than 1 throughout the time before the calibration update, indicating that the existing calibration data was not sufficiently relevant to provide reliable predictions on the process data. Then, note the step change decrease in the metrics near the end of the time frame, when a calibration update was performed. In this specific example, the calibration update involved recalibration using an updated data set that included process samples obtained during the early part of the time frame in this plot. As a result, the decreased metrics are a direct reflection of the increased relevance of the new calibration data as a result of the inclusion of more recent process samples. Finally, note the positive spikes in these metrics immediately before the upgrade, as well as at several other times before the upgrade. These indicate highly abnormal analyzer responses of a short-term or intermittent nature. These could have been caused by intermittent servicing of the analyzer or any part of its sampling system, or by short-term abnormal operation of the process. If one wishes to determine the specific nature or root causes of these abnormal responses, then the t- and q-contributions corresponding to the specific samples where these spikes occurred could provide useful information.

12.10.2 Deployment software

For any application, the software used for model deployment is quite different than the software used for model development. This is because deployment software has a very different function, and thus very different requirements, than development software. Whereas development software requires flexibility and user-friendliness, deployment software requires long-term stability, reliability, and accessibility to critical on-line data. The following are questions that reflect some common deployment software issues in a PAT project:

- Does the deployment software support the model and *preprocessing* types specified for a given method?
- Can the software be run *off-line*, to allow piloting/testing of models before deployment?
- Can the software be accessed remotely, to allow remote troubleshooting?
- Can deployed models be changed easily?
- Does it allow the real-time calculation of 'health monitoring' diagnostic metrics (Q and T^2, and their contributions)?
- Does the software satisfy site-specific and any applicable regulatory information security requirements (e.g., 21 CFR part 11)?
- Who is responsible for long-term support of the deployment software?

Many of these issues are born of the fact that, unlike development software, deployment software must be an integral part of a much larger data handling and process control system.

12.10.3 Data systems, and control system integration

Ultimately, the effectiveness of a model lies in its ability to provide useful and timely information to the customers, which for PAT applications are often process engineers, operators and manufacturing operations personnel. Therefore, one must also consider how best to integrate the chemometric models with the existing data handling and control system. Below are several issues in this area that are commonly experienced in PAT projects:

- How many model outputs can be sent to the control system (DCS) while still avoiding 'sensory overload' of process operators?

- How should the 'health monitor' statistics be used, and will this impact the model outputs that are provided to the customer? (e.g., should high T^2/Q results disable the usage of model outputs for control purposes, or should they only generate alarms?)
- Where/how should on-line data be archived, to enable/facilitate subsequent model updating?
- Remote access to the data system can greatly simplify data acquisition and model updating, but how can this be done without compromising system efficiency and security?

12.10.4 Method updating

Of course, no model deployment is static. Changing conditions in the process, the instrument, the sampling interface, other critical components, as well as changes in the measurement requirements, can render the original model ineffective, thus prompting the need to replace, update, or take out of service. One of the more common dilemmas in practical chemometrics is determining *how* to update an existing model if an update is deemed necessary. For both inverse and direct modeling methods, this boils down to assessing today's relevance of the calibration data that was used to build the existing model. If the relevance is low, then it is advantageous to build a new model using newly-collected calibration data. If not, then a new model can be developed using existing calibration data augmented with new data. There are many items to consider when making this assessment, including the following:

- details regarding changes in analyzer hardware and optics, changes in the process, and changes in the measurement requirements
- were any step changes observed in the health metrics (T^2 and Q) before the model's performance declined?
- the complexity of the existing model could be very high, possibly requiring a more localized modeling approach for the upgrade
- how difficult/costly would it be to collect additional calibration data?

It should also be mentioned that these, and other issues, are not only useful to consider when assessing *how* to update a model, but also when determining *when* or *how frequently* to consider model updates.

12.11 People Issues

It has been said that chemometrics is a highly 'interfacial' discipline, which brings together people from a wide range of different application venues and instrument types. Combine with this the typical PAT venue, involving a wide range of functions (plant operations, plant maintenance, process engineering, plant operations management, technical management, laboratory management, safety and environmental, and laboratory operations) and their corresponding value systems, and the importance of interpersonal skills becomes painfully apparent.

Based on my experiences, and experiences from my colleagues, there are several issues that seem to recur in PAT projects, and some that are unique to those that involve chemometrics:

- The familiarity level of statistics (and, certainly chemometrics) among plant personnel is generally low, resulting in widespread skepticism and apprehension.
- The typical plant laboratory is seldom actively involved in the development and support of process analyzers. This is unfortunate, because laboratory technical personnel are better equipped to understand chemometrics, and thus allay the concerns of others regarding the technology.

- In PAT, the benefits of chemometrics can have a short 'shelf life': they are readily apparent during initial deployment, but can disappear rapidly if changing conditions render initial models irrelevant.
- Plant maintenance organizations are most accustomed to simpler sensors that do not require chemometrics for calibration. Specialized training is required to understand and support more complex chemometrics-enhanced analyzers. To some, such specialization can be considered a career 'dead end'.

Regarding the 'shelf life' issue, it is always prudent to widely and diligently publicize any known limitations to a deployed model, actively monitor the model's performance with respect to both its outputs and its health indicators (T^2 and Q), and to promptly address any model updates that might be required.

12.12 The Final Word

Admittedly, navigating through a reference text on chemometrics can be a rather daunting task. There are many different tools in the toolbox, even though it is likely that you will only need a small subset of these to solve a specific problem. This chapter discussed many of the more commonly used and commonly available tools that are at your disposal. It is hoped that these discussions will improve your comfort level and confidence in chemometrics as an effective technology for solving your problems.

Regardless of your problem, and the specific chemometric tools that you choose to implement, there are three guiding principles that should always be kept in mind.

1. When building a method for on-line use, *keep it simple*! Strive for simplicity, but be wary of complexity.
2. Do your best to *cover the relevant analyzer response space* in your calibration data. If this cannot be achieved, then at least *know the limitations* in your calibration data.
3. Regardless of your background, strive to use both *chemical* and *statistical* thinking
   (a) Use your prior knowledge to guide you (chemical thinking),
   (b) But still 'listen' to the data – it might tell you something new! (statistical thinking)

With these principles in mind, and the tools at your disposal, you are ready to put chemometrics to work for your PAT projects.

References

1. H. Martens and T. Næs, *Multivariate Calibration*, John Wiley & Sons, Ltd, Chichester, 1989.
2. C.E. Miller, Chemometrics and NIR: A Match Made in Heaven? *Am. Pharm. Rev.*, **2**(2), 41–48 (1999).
3. H. Hotelling, Analysis of a complex of statistical variables into principal components, *J. Educ. Psychol*, **24**, 417–441 and 498–520 (1933).
4. K.H. Norris and J.R. Hart, Direct spectrophotometric determination of moisture content of grains and seeds, in *Proceedings 1963 International Symposium on Humidity and Moisture in Liquids and Solids*, A. Wexler (ed.), vol 4; Reinhold, New York, 1965; p. 19.
5. D.R. Massie and K.H. Norris, The spectral reflectance and transmittance properties of grains in the visible and near-infrared, *Trans. Am. Soc. Agric. Eng.*, **8**, 598 (1965).
6. W.R. Hruschka and K. Norris, Least squares curve fitting of near-infrared spectra predicts protein and moisture content in ground wheat, *Appl. Spectrosc.*, **36**, 261–265 (1982).
7. K.H. Norris, R.F. Barnes, J.E. Moore and J.S. Shenk, Predicting forage quality by infrared reflectance spectroscopy, *J. Anim. Sci.* **43**, 889–897 (1976).
8. M. Sjöström, S. Wold, W. Lindberg, J.A. Persson and H. Martens, A multivariate calibration problem in analytical chemistry solved by partial least squares models in latent variables, *Anal. Chim. Acta*, **150**, 61–70 (1983).

9. H. Martens, Factor analysis of chemical mixtures. Non-negative factor solutions for cereal amino acid data, *Anal. Chim. Acta*, **112**, 423–441 (1979).
10. H. Wold, Soft modeling: the basic design and some extensions, in *Systems Under Indirect Observation, Causality–Structure–Prediction, Part 2*, K.G. Joreskog and H. Wold (eds), North-Holland Publishing Co., Amsterdam, 1982, pp 1–54.
11. R.W. Gerlach, B.R. Kowalski and H.O.A. Wold, Partial least squares path modeling with latent variables, *Anal. Chim. Acta*, **112**, 417–21 (1979).
12. E.R. Malinowski and D.G. Howery, *Factor Analysis in Chemistry*, 1st Edition, John Wiley & Sons, New York, 1980.
13. B.R. Kowalski, *Chemometrics: Mathematics and Statistics in Chemistry*, D. Reidel, Dordrecht, Netherlands, 1980.
14. S. Wold, Cross-validatory estimation of the number of components in factor analysis and principal component models, *Technometrics*, **20**, 397–406 (1978).
15. C.E. Miller, The use of chemometric techniques in process analytical method development and operation, *Chemom. Intell. Lab. Syst.*, **30**, 11–22 (1995).
16. C.E. Miller, Chemometrics for on-line spectroscopy applications – theory and practice, *J. Chemom*, **14**, 513–528 (2000).
17. T.F. Kaltenbach, T.B. Colin and N. Redden, Chemometric modeling for process mass spectrometry applications, *AT-PROCESS*, **3**, 43–52 (1997).
18. R. Heikka, P. Minkkinen, and V-M Taavitsainen, Comparison of variable selection and regression methods in multivariate calibration of a process analyzer, *Process Control and Quality*, **6**, 47–54 (1994).
19. C.E. Miller, Chemometrics in process NIR: tales from a 14-year old (and growing…) application, *NIR News*, **18**, 18 (2007)
20. B.J. Prazen, R.E. Synovec and B.R. Kowalski, Standardization of second-order chromatographic/spectroscopy data for optimal chemical analysis, *Anal. Chem.*, **70**, 218–225 (1998).
21. P. Geladi, D. MacDougall and H. Martens, Linearization and scatter correction for near-infrared reflectance spectra of meat, *Appl. Spectrosc.*, **39**, 491 (1985).
22. G.E.P. Box, W.G. Hunter and J.S. Hunter, *Statistics for Experimenters: An Introduction to Design, Data Analysis and Model Building*, John Wiley & Sons, New York, 1978.
23. J.R. Taylor, *An Introduction to Error Analysis*; University Science Books, Mill Valley, CA, USA, 1982.
24. W.F. McClure, A. Hamid, F.G. Giesbrecht, and W.W. Weeks, Fourier analysis enhances NIR diffuse reflectance spectroscopy, *Appl. Spectrosc.*, **38**, 322–329 (1984).
25. M. Cocchi, R. Seeber, and A. Ulrici, Multivariate calibration of analytical signals by WILMA (wavelet interface to linear modeling analysis), *J. Chemom.*, **17**, 512 (2003).
26. Y. Liu and S.D. Brown, Wavelet multiscale regression from the perspective of data fusion, *Anal. Bioanalytical Chemistry*, **380**, 445–52 (2004).
27. S.G. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, New York, 1998.
28. G. Golub and C.F. Van Loan, *Matrix Computations*, Johns Hopkins University Press, Baltimore, MD, USA, 1983.
29. R.A. Fisher, The use of multiple measurements in taxonomic problems, *Ann. Eugenics*, **7**, 179–188 (1936).
30. J.E. Jackson, and G.S. Mudholkar, Control procedures for residuals associated with principal component analysis, *Technometrics*, **21**(3) (1979)
31. J.E. Jackson, *A User's Guide to Principal Components*, John Wiley & Sons, New York, 1991.
32. A. Savitzky and M.J.E. Golay, Smoothing and differentiation of data by simplified least squares procedures, *Anal. Chem.*, **36**, 1627 (1964).
33. P.R. Mobley, B.R. Kowalski, J.J. Workman and R. Bro, Review of chemometrics applied to spectroscopy: 1985–95, Part 2, *Appl. Spec. Rev.*, **31**, 347–368 (1996).
34. C.E. Miller and T. Næs, A pathlength correction method for near-infrared spectroscopy, *Appl. Spectrosc.*, **44**, 895–898 (1990).
35. H. Martens and E. Stark, Extended multiplicative signal correction and spectral interference subtraction: new preprocessing methods for near infrared spectroscopy, *J. Pharm. Biomed. Anal.*, **9**, 625–35 (1991).
36. N.B. Gallagher, T.A. Blake, P.L. Gassman, Application of extended inverse scatter correction to mid-infrared reflectance spectra of soil, *J. Chemom.*, **19**, 271 (2005).
37. H. Martens, J.P. Nielsen and S.B. Engelsen, Light scattering and light absorbance separated by extended multiplicative signal correction. Application to near-infrared transmission analysis of powder mixtures, *Anal. Chem.*, **75**, 394 (2003).
38. H. Martens, M. Hoy, B.W. Wise, R. Bro and P. Brockhoff, Pre-whitening of data by covariance-weighted preprocessing, *J Chemom.*, **17**, 153–165 (2003).

39. R. DiFoggio, Desensitizing models using covariance matrix transforms or counter-balanced distortions, *J. Chemom.*, **19**, 203 (2005).
40. C.E. Miller, Chemical principles of near-infrared technology, in *Near Infrared Technology for the Agricultural and Food Industries*, P. Williams and K. Norris (eds), American Association of Cereal Chemists, St. Paul, MN, USA, 2001, pp 19–36.
41. A.S. Bonanno, J.M. Ollinger and P.R. Griffiths, The origin of band positions and widths in near infrared spectroscopy, in *Near-Infrared Spectroscopy – Bridging the Gap Between Data Analysis and NIR Applications*, K.I. Hildrum, T. Isaksson, T. Næs and A. Tandberg (eds), Ellis-Horwood; New York, 1992.
42. B. Walczak, B. van den Bogaert and D.L. Massart, Application of wavelet packet transform in pattern recognition of near-IR data, *Anal. Chem.*, **68**, 1742–7 (1996).
43. M. Chaoxiong and S. Xueguang, Continuous wavelet transform applied to removing the fluctuating background in near-infrared spectra, *J. Chem. Inf. Comput. Sci.*, **44**, 907–911 (2004).
44. C.R. Mittermayr, H.W. Tan and S.D. Brown, Robust calibration with respect to background variation, *Appl. Spectrosc.*, **55**, 827–833 (2001).
45. T. Fearn and A.M.C. Davies, A comparison of Fourier and wavelet transforms in the processing of near-infrared spectroscopic data: part 1. Data compression, *J. Near Infrared Spectrosc.*, **11**, 3–15 (2003).
46. C.W. Brown and C.W. Obremski, Multicomponent quantitative analysis, *Appl. Spec. Rev.*, **20**, 373–418 (1984).
47. R. Marbach, A new method for multivariate calibration, *J. Near Infrared Spectrosc.*, **13**, 241 (2005).
48. R. Marbach, On Wiener filtering and the physics behind statistical modeling, *J. Biomed. Optics*, **7**, 130 (2002).
49. R.D. Guenard, C.M. Wehlberg, R.J. Pell and D.M. Haaland, Importance of prediction outlier diagnostics in determining a successful inter-vendor multivariate calibration model transfer, *Appl. Spectrosc.*, **61**, 747 (2007).
50. D.M. Haaland and D.K. Melgaard, New prediction-augmented classical least squares (PACLS) methods: application to unmodeled interferents, *Appl. Spectrosc.* **54**, 1303 (2000).
51. D.K. Melgaard and D.M. Haaland, Comparisons of prediction abilities of augmented classical least squares and partial least squares with realistic simulated data: effects of uncorrelated and correlated errors with nonlinearities, *Appl. Spectrosc.*, **58**, 1065–73 (2004).
52. C.E. Miller, B.E. Eichinger, T.W. Gurley and J.G. Hermiller, Determination of microstructure and composition in butadiene and styrene-butadiene polymers by near-infrared spectroscopy, *Anal. Chem.*, **62**, 1778–1785 (1990).
53. M. Blanco, J. Coello, H. Iturriaga, S. Maspoch and J. Pages, NIR calibration in non-linear systems: different PLS approaches and artificial neural networks, *Chemom. Intell. Lab. Syst.*, **50**, 75–82 (2000).
54. T. Næs, Multivariate calibration by covariance adjustment, *Biometrical J.*, **28**, 99–107 (1986).
55. T. Fearn, Misuse of ridge regression in the calibration of near-infrared reflectance instruments, *Appl. Statistics*, **32**, 73–79 (1983).
56. J.H. Friedman and W. Stuetzle, Projection pursuit regression, *J. Am Stat. Assoc.*, **76**, 817–823 (1981).
57. R. Tkachuk, F.D. Kuzina and R.D. Reichert, Analysis of protein in ground and whole field peas by near-infrared reflectance, *Cereal Chemistry*, **64**, 418–22 (1987).
58. B. Tenge and D.E. Honigs, The effect of wavelength searches on NIR calibrations, *Abstracts of Papers of the American Chemical Society*, **193**(APR), 112 (1987).
59. C.L. Mallows, Some comments on Cp, *Technometrics*, **15**, 661–675 (1973).
60. A. Hayden, E. Niple, and B. Boyce, Determination of trace-gas amounts in plumes by the use of orthogonal digital filtering of thermal emission spectra, *Appl. Opt.*, **35**(16), 2802–2809 (1996).
61. A. Lorber, K. Faber and B.R. Kowalski, Net analyte signal calculation in multivariate calibration, *Anal. Chem.*, **69**, 1620–1626 (1997).
62. C.D. Brown, Discordance between net analyte signal theory and practical multivariate calibration, *Anal. Chem*, **76**(15), 4364–73 (2004).
63. J. Mandel, Use of the singular value decomposition in regression analysis, *The American Statistician*, **36**, 15–24 (1982).
64. S. Wold, C. Albano and W.J. Dunn, Pattern recognition: finding and using patterns in multivariate data, *in Food Research and Data Analysis*, H. Martens and H. Russwurm (eds), Applied Science Publishers, London, 1983; pp 147–188.
65. S. de Jong, SIMPLS: An alternative approach to partial least squares regression, *Chemom. Intell. Lab. Syst.*, **18**, 251–63 (1993).
66. L. Elden, Partial least-squares vs. Lanczos Bidiagonalization – I: Analysis of a projection method for multiple regression, *Computational Statistics & Data Analysis*, 46, 11–31, (2004).
67. T. Næs, K. Kvaal, T. Isaksson and C. Miller, Artificial neural networks in multivariate calibration, *J. Near Infrared Spectrosc.*, **1**, 1–11 (1993).

68. K.U. Jagemann, C. Fischbacher, U.A. Muller and B. Mertes, Application of near-infrared spectroscopy for non-invasive determination of blood-tissue glucose using neural networks, *Zeitschrift Phys. Chemie – Int. J. Res. Phys. Chem. Chem. Phys.*, **191**, 179–190 (1995).
69. D.E. Rumelhart, G.E. Hinton and R.J. Williams, Learning internal representations by error propagation, in *Parallel Data Processing*, Rumelhart, D. and McClelland, J., (eds), MIT Press; Cambridge, MA, USA, 1986, pp 318–362.
70. V.N. Vapnik, *The Nature of Statistical Learning Theory*, Springer, New York, 1995.
71. C.J.C. Burges, A tutorial on support vector machines for pattern recognition, *Data Mining and Knowledge Discovery*, **2**, 121–167 (1998).
72. S.J. Dixon and R.G. Brereton, Comparison of performance of five common classifiers represented as boundary methods: Euclidean distance to centroids, linear discriminant analysis, quadratic discriminant analysis, learning vector quantization and support vector machines, as dependent on data structure, *Chemom. Intell. Lab. Syst.*, **95**, 1–17 (2009).
73. U. Thissen, M. Pepers, B. Üstün, W.J. Melssen and L.M.C. Buydens, Comparing support vector machines to PLS for spectral regression applications, *Chemom. Intell. Lab. Syst.*, **73**, 169–179 (2004).
74. U. Thissen, B. Üstün, W.J. Melssen and L.M.C. Buydens, Multivariate calibration with least-squares support vector machines, *Anal. Chem.*, **76**, 3099 (2004).
75. K.V. Mardia, J.T. Kent and J.M. Bibby, *Multivariate Analysis*, Academic Press, London, 1979.
76. C.E. Miller and B.E. Eichinger, Analysis of rigid polyurethane foams by near-infrared diffuse reflectance spectroscopy, *Appl. Spectrosc.*, **44**, 887–894 (1990).
77. M.A. Sharaf, D.L. Illman, and B.R. Kowalski, *Chemometrics*, John Wiley & Sons, New York, 1986.
78. M. Barker and W. Rayens, Partial least squares for discrimination, *J. Chemom.*, **17**(3), 166–173 (2003).
79. S. Chevallier, D. Bertrand, A. Kohler and P. Courcoux, Application of PLS-DA in multivariate image analysis, *J. Chemom.*, **20**(5), 221–229 (2006).
80. K. Fukunaga, *Introduction to Statistical Pattern Recognition*, Academic Press, San Diego, CA, USA, 1990.
81. S. Wold, Pattern recognition by means of disjoint principal component models, *Patt. Recog.*, **8**, 127–139, 1976.
82. F. Rosenblatt, *Principles of Neurodynamics*, Spartan Press, Washington DC, USA, 1961.
83. T. Kohonen, *Self-Organization and Associative Memory*, 2nd edition, Springer-Verlag, Berlin, 1987.
84. N.A. Woody and S.D. Brown, Hybrid Bayesian networks: making the hybrid Bayesian classifier robust to missing training data, *J. Chemom*, **17**, 266–273 (2003).
85. C.E. Miller, Use of near-infrared spectroscopy to determine the composition of high-density/low-density polyethylene blend films, *Appl. Spectrosc.*, **47**, 222 (1993).
86. C.E. Miller and B.E. Eichinger, Analysis of reaction-injection-molded polyurethanes by near-infrared diffuse reflectance spectroscopy, *J. Appl. Polym. Sci.*, **42**, 2169–2190 (1991).
87. W. Windig and H.L.C. Meuzelaar, Numerical extraction of components from mixture spectra by multivariate data analysis, in *Computer Enhanced Analytical Spectroscopy*, H.L.C. Meuzelaar and T.L. Isenhour (eds), Plenum, New York, 1987, pp 81–82.
88. P.J. Gemperline and E. Cash, Advantages of soft versus hard constraints in self-modeling curve resolution problems. Alternating Least squares with penalty functions, *Anal. Chem.*, **75**, 4236 (2003).
89. A. de Juan and R. Tauler, Multivariate curve resolution (MCR) from 2000: progress in concepts and applications, *Crit. Rev. Anal. Chem.*, **36**, 163–176 (2006).
90. W.H. Lawton and E.A. Sylvestre. Self modeling curve resolution, *Technometrics* **13**, 617–633, (1971).
91. R. Tauler, B. Kowalski and S. Fleming, Multivariate curve resolution applied to spectral data from multiple runs of an industrial process, *Anal. Chem.*, **65**, 2040–2047 (1993).
92. P. Gemperline, M. Zhu, E. Cash and D.S. Walker, Chemometric characterization of batch reactions, *ISA Trans.*, **38**, 211–216 (1999).
93. A. Elbergali, J. Nygren, and M. Kubista, An automated procedure to predict the number of components in spectroscopic data, *Anal. Chim. Acta*, **379**, 143–158 (1999).
94. K. Faber and B.R. Kowalski, Critical evaluation of two F-tests for selecting the number of factors in abstract factor analysis, *Anal. Chim. Acta*, **337**, 57–71 (1997).
95. J. Ferré and F.X. Rius, Constructing D-optimal designs from a list of candidate samples, *TrAC Trends Anal. Chem.*, **16**(2), 70–73 (1997).
96. T. Isaksson and T. Næs, Selection of samples for calibration in near-infrared spectroscopy. Part II: selection based on spectral measurements, *Appl. Spectrosc.*, **44**, 1152–1158 (1990).
97. P.J. Brown, Wavelength selection in multicomponent near-infrared calibration, *J. Chemom.*, **6**, 151–161 (1991).

98. L. Nørgaard, A. Saudland, J. Wagner, J.P. Nielsen, L. Munck and S.B. Engelsen, Interval partial least-squares regression (iPLS): a comparative chemometric study with an example from near-infrared spectroscopy, *Appl. Spectrosc.*, **54**, 413 (2000).
99. R. Leardi, Genetic algorithms in chemometrics and chemistry: a review, *J. Chemom*, **15**, 559–569 (2001).
100. D. Jouan-Rimbaud, D.-L. Massart, R. Leardi and O.E. De Noord, Genetic algorithms as a tool for wavelength selection in multivariate calibration, *Anal. Chem.*, **67**, 4295–4301 (1995).
101. R. Leardi and L. Nørgaard, Sequential application of backward interval partial least squares and genetic algorithms for the selection of relevant spectral regions, *J. Chemom.*, **18**, 486 (2004).
102. F. Lindgren, P. Geladi, A. Berglund, M. Sjostrom and S. Wold, Interactive variable selection (IVS) for PLS. Part II: chemical applications, *J. Chemom.*, **9**, 331–342 (1995).
103. E. Anderssen, K. Dyrstad, F. Westad and H. Martens, Reducing over-optimism in variable selection by cross-model validation, *Chemom. Intell. Lab. Syst.*, **84** 69–74 (2006).
104. L. Gidskehaug, E. Anderssen and B.K. Alsberg, Cross model validation and optimisation of bilinear regression models, *Chemom. Intell. Lab. Syst.*, **93**, 1–10 (2008).
105. R. DiFoggio, Examination of some misconceptions about near-infrared analysis, *Appl. Spectrosc.*, **49**, 67 (1995).
106. A. Springsteen, Photometric standards for near-infrared spectrometry, in *Handbook of Vibrational Spectroscopy*, J. Chalmers and P. Griffiths (eds), John Wiley & Sons, Ltd, Chichester, 2002.
107. S.J. Choquette, J.C. Travis, L.E. O' Neal, C. Zhu, and D.L. Duewer, SRM® 2035: a rare earth oxide glass for the wavelength calibration of near infrared dispersive and Fourier transform spectrometers, *Spectroscopy*, **16**(4), 14–19 (2001).
108. E. Bouveresse, C. Hartmann, D.L. Massart, I.R. Last and K.A. Prebble, Standardization of near-infrared spectrometric instruments, *Anal. Chem.*, **68**, 982–990 (1996).
109. Y. Wang, D.J. Veltkamp and B.R. Kowalski, Multivariate instrument standardization, *Anal. Chem.*, **63**, 2750–2756 (1991).
110. R.W. Kennard and L.A. Stone, Computer aided design of experiments, *Technometrics*, **11**, 137–149 (1969).
111. J. Workman and J. Coates, Multivariate calibration transfer: the importance of standardizing instrumentation, *Spectroscopy*, **8**(9), 36–42 (1993)
112. E. Bouveresse, D.L. Massart and P. Dardenne, Calibration transfer across near-infrared spectroscopic instruments using Shenk's algorithm: effects of different standardisation samples, *Anal. Chim. Acta*, **297**, 405–416 (1994).
113. B.G. Osborne and T. Fearn, Collaborative evaluation of universal calibrations for the measurement of protein and moisture in flour by near-infrared reflectance, *J. Food Technol.*, **18**, 453–460 (1983).
114. Y. Wang and B.R. Kowalski, Calibration transfer and measurement stability of near-infrared spectrometers, *Appl. Spectrosc.*, **46**, 764–771 (1992).
115. Y. Wang, M.J. Lysaght and B.R. Kowalski, Improvement of multivariate calibration through instrument standardization, *Anal. Chem.*, **64**, 562–564 (1992).
116. R.P. Cogdill, C.A. Anderson and J.K. Drennen III., Process analytical technology case study, Part III: calibration monitoring and transfer, *AAPS PharmSciTech.*, **6**(2): E284–E297 (2005).
117. J.S. Shenk and M.O. Westerhaus, Optical instrument calibration system, US Pat. 4866644, Sept. 12, 1989.
118. S. Wold, H. Antti, F. Lindgren and J. Öhman, Orthogonal signal correction of near infrared spectra, *Chemom. Intell. Lab. Syst.*, **44**, 175–185, (1998).
119. R.N. Feudale, H. Tan and S.D. Brown, Piecewise orthogonal signal correction, *Chemom. Intell. Lab. Syst.*, **63**(2), 129–138 (2002).
120. T.B. Blank, S.T. Sum, S.D. Brown and S.L. Monfre, Transfer of near-infrared multivariate calibrations without standards, *Anal. Chem.*, **68**, 2987–2995 (1996).
121. R. Bro, *Multi-way Analysis in the Food Industry. Models, Algorithms, and Applications*, 1998, PhD Thesis, University of Amsterdam (NL) and Royal Veterinary and Agricultural University (DK), 1998.
122. A.K. Smilde, R. Bro and P. Geladi, *Multi-way Analysis with Applications in the Chemical Sciences*, John Wiley & Sons, Ltd, Chichester, 2004.

13

On-line PAT Applications of Spectroscopy in the Pharmaceutical Industry

Brandye Smith-Goettler

Merck and Co., Inc., West Point, PA, USA

13.1 Background

The FDA announced in August 2002 a 2-year initiative, titled 'Pharmaceutical cGMPs for the 21st Century: A Risk-Based Approach,' to enhance the regulation of pharmaceutical manufacturing and product quality covering veterinary and human drugs, including biological drug products.[1] In August 2003, the FDA released the science-based risk management goal '5-Part Strategic Action Plan to Protect and Advance America's Health' consisting of the following;[1]

- CFR 21 Part 11, Electronic Records, Electronic Signatures – Scope and Application
- Formal Dispute Resolution: Scientific and Technical Issues Related to Pharmaceutical cGMP
- Sterile Drug Products Produced by Aseptic Processing: Current Good Manufacturing Practices
- Comparability Protocols – Protein Drug Products and Biological Products, Chemistry, Manufacturing, and Controls Information
- Process Analytical Technology (PAT) – A Framework for Innovative Pharmaceutical Manufacturing and Quality Assurance.

Of particular interest is the PAT guidance, finalized in September 2004.[2] Unlike the chemical, petroleum, polymer and food industries, the pharmaceutical industry is under heavy regulatory scrutiny. This is not just regulation by the FDA, but regulatory agencies worldwide. The delivery of safe and effective drug product to improve the quality of life necessitates regulation.

Many in the pharmaceutical industry will recall the *Wall Street Journal* article that declared the manufacture of potato chips more technologically advanced than the manufacture of pharmaceuticals.[3] This may seem a harsh criticism, but it is honest. Dissolution has been used to test drug product quality since the

Process Analytical Technology 2e Edited by Katherine Bakeev
© 2010 John Wiley & Sons, Ltd.

1970s, and not to argue the effectiveness of dissolution testing which obviously works, but think of the advances made in other industries since the 1970s.[4] In the 1970s someone might have listened to their favorite music via an 8-track, in the 1980s by cassette, in the 1990s by CD and in this century by mp3. How much has drug end-product testing changed since then? While we all can appreciate a bag of potato chips, we can also appreciate the FDA initiative to bring pharmaceutical manufacturing into the 21st century.

With that said process analytical chemistry (PAC) has been commonplace in several industries, even the pharmaceutical industry, prior to 2004. Riebe and Eustace differentiate PAC from analytical chemistry with one word; location.[5] PAC is typically spectroscopy, separations or chemometrics, as applied either at-, in- or on-line to manufacturing. Several comprehensive application reviews of process analytical chemistry have been published in *Analytical Chemistry* since 1987.[5–13] In one of these reviews Workman *et al*. comment that PAT has continued to evolve as a more appropriate term than PAC while noting PAC has existed since 1911 but is just now finding common usage.[12] The scope of PAT is much broader than that of PAC as the FDA guidance outlines the following desired state for manufacturing:[2]

- product quality and performance are ensured through the design of effective and efficient manufacturing processes
- product and process specifications are based on a mechanistic understanding of how formulation and process factors affect product performance
- continuous *real time* quality assurance
- relevant regulatory policies and procedures are tailored to accommodate the most current level of scientific knowledge
- risk-based regulatory approaches recognize the following:
  - the level of scientific understanding of how formulation and manufacturing process factors affect product quality and performance
  - the capability of process control strategies to prevent or mitigate the risk of producing a poor quality product.

So PAC has indeed evolved to PAT, which encompasses not only process measurement, real-time quality assurance and control but process understanding to enable product quality by process design. The tools available to us for quality by design as defined by the FDA include:[2]

- multivariate tools for design, data acquisition and analysis
- process analyzers
- process control tools
- continuous improvement and knowledge management tools.

The focus of this chapter will be the application of on-line spectroscopy within the pharmaceutical industry. All of the aforementioned tools are constituents of the on-line spectroscopic methods described herein. It is noteworthy that although this chapter is an extensive review of the literature, it is not an exhaustive review. Representative studies from several unit operations, but not all unit operations, will be outlined. The chapter will discuss on-line spectroscopic methods, and in some instances at-line methods for a historical perspective, as applied to active pharmaceutical ingredient (API) reaction monitoring, API crystallization, API drying, nanomilling, hot melt extrusion, wet and dry granulation, drug product blending, compression, spray coating, fermentation, freeze-drying and cleaning validation. On-line spectroscopy as used within this chapter includes both on-line and in-line sample geometries as defined by the FDA's PAT guidance[2] where the sample is diverted and where the sample is not removed from the process stream (invasive or noninvasive), respectively.

13.2 Reaction Monitoring

Process monitoring and control of API production, sans the regulatory environment, is analogous to that within the chemical industry. Since the early 1990s, numerous papers have been published noting on-line spectroscopic techniques as applied to API reaction monitoring. A representation of some of these on-line spectroscopic reaction monitoring techniques will be provided herein with additional information discussed in Chapter 15.

In 1995, an attenuated total reflection Fourier-transform mid-IR (ATR FTIR) spectroscopic method was developed by scientists at Merck & Co., Inc. to monitor the epoxidation of indene to indene oxide. The ATR-FTIR method combined with multivariate data analysis yielded reliable predictions of indene and indene oxide levels as referenced to off-line high-performance liquid chromatography (HPLC) data.[14] Later in 2002, 2003 and 2006 scientists at Merck & Co., Inc. reported the successful use of on-line ATR FTIR to control a deprotection reaction, lactol activation reaction and a salt formation process, respectively.[15–17] Scientists at Pfizer Inc. successfully used mid-IR spectroscopy to monitor hydrogenation of several functional groups, to compare the relative reaction rates between batches and to quantitate the amount of unreacted starting material. The authors note that mid-IR spectroscopy was preferred over near-infrared spectroscopy (NIRS) since direct information could be obtained without the need for chemometrics.[18] In contrast, De Braekeleer and associates investigated several chemometric algorithms, such as orthogonal projection approach (OPA), principal component analysis (PCA), and multivariate curve resolution alternating least squares (MCR-ALS) to detect reaction end point via on-line mid-IR spectroscopic monitoring.[19]

Scientists at Pfizer Inc. eliminated an estimated 60 min delay by employing an on-line NIR with partial least-squares (PLS) modeling technique for hydrogenation end-point determination, as opposed to the traditional off-line gas chromatography (GC) assay. It is noteworthy that if the reaction were to exceed the optimal end point, undesired side products would be created.[20] Thus real-time monitoring was value added for this application since it diminished costly down times and batch reprocessing. Scientists at Merck & Co., Inc. used real-time NIRS and temperature variance data with multivariate data analysis to assess the extent of a solvent switch process (methanol to *n*-propanol, to isopropyl acetate) which generated reliable predictions of total alcohol as compared to reference GC data.[21] Another example of on-line NIRS for reaction monitoring is given by Coffey and associates where the formation of product as well as the by-product methanol was monitored. The authors report the NIRS technique coupled with PLS calibration, that was referenced with HPLC results, provided product formation predictions, reaction end-point predictions, yielded a cost saving by indicating the reaction was finished in 50 min as compared to the five hour anticipated run time and reduced the risk of by-product formation.[22] Blanco and Serrano also used on-line NIRS with PLS modeling to monitor a chemical reaction: catalyzed esterification of butan-1-ol by acetic acid. They took the methodology one step further by using their method not only to quantitatively determine yields and end points but also rate constants, equilibrium constants and the influence of water on reaction performance.[23]

Quinn *et al*. utilized on-line UV-visible spectroscopy for reaction monitoring.[24] For more details regarding UV-visible spectroscopy as applied to PAT, refer to Chapter 4. The authors note that the collected UV-visible spectra had low information content and a lack of well resolved, unique absorption bands. However, when self-modeling curve resolution (SMCR) was applied to the data, three species were observed including a reaction intermediate that was not detected via traditional off-line HPLC analysis. PLS was used to predict starting material and product concentrations, but yielded poor prediction results as compared to the SMCR results. The SMCR results were reported as not susceptible to errors caused by run-to-run variations since separate curve resolution models were generated for each run. The authors also note that the on-line UV-visible spectroscopy when combined with the SMCR was able to characterize the given reaction without the use of a reference method.[24]

13.3 Crystallization

Crystallization is obviously a constituent of a chemical reaction, but is considered a separate unit operation within this section and also treated as such during process development. In an excellent review article pertaining to the application of PAT, the FDA comments that crystallizations are among the most critical yet least understood pharmaceutical processes. The authors note that historically crystallization processes have been developed empirically but could benefit from applying PAT to identify critical attributes (i.e. crystal size distribution, crystal shape and polymorphic form), to monitor these attributes, to interpret the results via multivariate data analyses and to ultimately control the process.[25] This improves the crystallization process while benefiting downstream unit operations, that is filtration and drying, and end-product quality.[26] More specifically crystal size and shape can affect overall processability, such as drug substance flow, and crystal size and polymorphic form can influence end-product dissolution.[25] Mid-infrared, NIR and Raman spectroscopic techniques have been applied to pharmaceutical crystallization processes to identify polymorphs and to provide information regarding crystal structure. Some of these case studies will be presented here.

Supersaturation, which is known to have an effect on crystal size distribution, was measured via on-line ATR-FTIR spectroscopy in a series of crystallization processes.[27,28] The study yielded results which outline the advantages of on-line monitoring including: (1) characterizing the solute/solvent system, (2) designing on-line control strategies that improve the crystal size distribution and the reproducibility of final product quality, (3) assessing improved operating strategies and (4) the monitoring of polymorphic transitions during cooling crystallization operations.[28] Noteworthy is that the successful application of ATR-FTIR for monitoring crystallization processes requires cautious and rigorous calibration procedures that account for the temperature dependence of mid-infrared spectra.[27,28] The same can be said of IR, NIR and/or Raman calibrations in general.

Two similar ATR-FTIR studies were conducted that additionally utilized focus beam reflectance measurements (FBRM), based on laser backscattering.[26,29] One study used paracetamol (acetaminophen) in water as a model system due to two challenging features: low solubility and the tendency to agglomerate during crystallization. Agglomeration is a familiar difficulty with crystallization and can lead to solvent inclusions which affect drug product stability. The objective of the given study was to operate in conditions that yielded large paracetamol crystals while precluding agglomeration. The initial crystallization optimization step was determining the metastable zone which requires determining the solubility curve: solute to solvent ratio versus temperature. The metastable zone is where the solution is supersaturated, and while operating in this region undesired nucleation is avoided.[29] As with these two studies and the aforementioned studies, on-line ATR-FTIR spectroscopy was used to determine the solubility curve.[26–29] Fujiwara *et al*. comment that advantages of ATR-FTIR over other on-line measurements (i.e. conductivity and densitometry), include its applicability to organic systems, its ability to measure multiple concentrations within a multicomponent system and eliminating the requirement for an external sampling loop.[27,29] ATR-FTIR was used to predict solution concentration, which provides the solubility curve. Being that the solubility curve is the lower boundary of the metastable zone, ATR-FTIR can be used to monitor and control crystallization. Although the authors report that laser backscattering was more sensitive in regards to detecting the metastable limit, ideal paracetamol crystals were produced by controlling the concentration via *in situ* ATR-FTIR.[29] Similarly, Liotta and Sabesan used laser backscattering to define the metastable zone and ATR-FTIR with PLS for real-time concentration predictions. This contributed to real-time supersaturation feedback control, which involved a primary loop to determine a set point for cooling rate and a secondary loop to adjust the heater/chiller to maintain the specified cooling rate.[26]

Another similar study of crystallization was performed where on-line ATR-FTIR spectroscopy was used to provide concentration information, however on-line diffuse reflectance Fourier-transform infrared

(DRIFT-IR) spectroscopy was also used for polymorphic characterization. The authors detail the application of multivariate techniques, multivariate statistical process control (MSPC), PCA and PLS, to the spectroscopic data for a simple yet powerful, rapid evaluation of the given crystallization process.[30]

In 2002, Starbuck and associates reported the use of on-line Raman spectroscopy for the optimization of a complex pharmaceutical polymorphic system. The authors wanted to understand the polymorphic transitions to a desired polymorph, to generate kinetic data for crystal form transitions to and from the desired polymorph and lastly, to probe process upsets that might lead to the instability of the desired polymorph. The use of on-line Raman spectroscopy enabled the generation of kinetic data to establish process cycle times which contributed to a highly robust process that consistently produces the desired polymorph. The authors note that Raman kinetic studies were most useful for defining the process cycle times and investigating process upsets, but only after the thermodynamic properties of the process were characterized.[31] NIR spectroscopy has also been used to enhance process robustness by detecting and monitoring the desired polymorph in a crystallization process.[32,33]

Recently, on-line FBRM, ATR-FTIR spectroscopy, Raman spectroscopy and PLS were used to monitor a complex crystallization system: a racemic free base of a given compound and a chiral acid. The authors first demonstrate that the diastereomeric composition can be estimated using Raman spectral data, slurry density and temperature using a PLS model. Consequently the issue of on-line slurry density prediction, which is not readily available, arises. An additional PLS model was constructed that used the ATR-FTIR spectral data to infer slurry density. Slurry density as predicted in real-time via ATR-FTIR spectroscopy was fed into the aforementioned Raman, slurry density and temperature PLS model to yield a more accurate estimate of the fractional solid composition of the two diastereomers.[34]

13.4 API Drying

The drying of the API, regardless of form, is a time-consuming and typically hazardous process. Historically, in-process controls are in place that require sample thieving and an off-line assay (i.e. moisture determination via Karl Fischer), loss on drying (LOD) or GC.[35] If toxicity is a concern, then in-process sampling needs to be either reduced or removed. Within this section on-line spectroscopic dryer monitoring will be discussed, including (1) moisture determination via direct contact with the drug substance and (2) moisture/solvent determination via effluent monitoring. Drying of drug product within a fluid bed dryer will be covered in Section 13.7.1.

If a drug substance is hygroscopic, off-line moisture determination will require careful sample handling. Otherwise, the amount of moisture determined for the thieved sample may not be reflective of that for the bulk drug substance within the dryer. The aforementioned safety and sample handling issues combined with the fact that water has overtones and combination bands in the NIR region, have resulted in the application of NIRS for drug substance moisture determination. Early off-line applications of NIRS to determine moisture content are noted within, but are not limited to, X. Zhou *et al.*, Last and Prebble, and MacDonald and Prebble.[35–38]

Merck & Co., Inc. used on-line NIRS to determine moisture content in a cytotoxic compound during a laboratory-scale drying study. A PLS regression was performed using NIR spectral data and Karl Fischer data as a reference. The model resulted in a standard error of prediction equal to 0.99% w/w over a 0–15% w/w water range. Furthermore the given method of moisture determination was robust given numerous process variables were being optimized during this study.[39] That same year Merck & Co., Inc. reported on the use of on-line NIRS to determine moisture content during a drying process while differentiating between surface and bound water. A PLS calibration using data from three drying runs, four principal components and second- derivative spectral data in the wavelength range of 1890–1950 nm was constructed

where reference moisture values were determined via Karl Fischer. An independent drying run was used to test the on-line model. Process monitoring began with a wet cake (~45% w/w H_2O) inside an agitated filter dryer at 35 °C under full vacuum. When 25% w/w H_2O was achieved, a humid nitrogen sweep and intermittent agitation were applied. The process continued until the target moisture content of 5–8% w/w was reached.

On-line NIRS contributed to process control by quantitatively predicting the moisture content. NIRS also provided process understanding within this study where the authors note that PCA was able to discriminate amongst various hydration states. The target end point of 5–8% w/w H_2O was chosen because the trihydrate was the desired form. Using on-line NIRS to monitor this drying process was ideal since end point was determined in real-time prior to the formation of an undesired lower hydrate form.[40]

Walker *et al*. noted that one aspect of drug substance drying that presented an opportunity for process monitoring improvement is quantitating the amount of solvent removed. Typically the amount of removed solvent is inferred from a condenser mass. But since the condenser is often significantly displaced from the dryer, common to several dryers and not solvent specific this inference is not necessarily an accurate one. The authors, seeing a potential for real-time monitoring of effluent, applied NIRS and PLS modeling to determine the mass of solvent removed from the drug substance. The NIR spectrometer was interfaced with a gas cell (0.6 m long and 2.5 cm i.d.) positioned on the vapor stream between the dryer and the vacuum line.[41,42] The experimental change in mass with respect to time predictions were used to calculate 'dynamic' vapor density values, which were in agreement with values calculated from the ideal gas law, thus providing potential to also indirectly measure temperature and pressure. The authors also note that calculated 'dynamic' vapor density values could be indicative of undesired materials in the effluent stream, such as an air leak.[42]

13.5 Nanomilling

Poorly water-soluble APIs are noted to have issues with formulation stability and bioavailability, and nanomilling has proven to be an effective processing option to overcome insolubility to enable high drug loading and delivery.[43–47] A critical nanomilling process parameter includes particle size distribution, and thus an opportunity for on-line spectroscopic monitoring and end-point control. Several off-line methodologies exist for measuring particle size: electro-acoustic spectroscopy, scanning electron microscopy, laser light scattering and photon correlation spectroscopy. These methods are valuable for determining particle size, but eliminating the need for sampling/sample preparation and process start/stops by utilizing an on-line technology yields both a more precise particle size distribution measurement and a more efficient nanomilling process. Early work has demonstrated a correlation between NIR reflectance spectra and particle size,[48–50] thus providing proof-of-concept for the application of on-line NIR for processes in which particle size control is desirable. Higgins *et al*. first reported the application of NIR for on-line monitoring of drug compound nanoparticle size in aqueous colloidal dispersions containing a high concentration of solids. The NIR diffuse reflectance probe was inserted into the process pipe that transports the nanomilled colloidal dispersion to the media mill as illustrated in Figure 13.1. A PLS model was constructed to predict D90 values from the spectral data with laser light scattering as the reference technique. The NIR method had an accuracy of 2.4 nm near the end point of nanoparticle production where the volume-weighted D90 particle size was determined to be 200–220 nm. Utilizing this on-line NIR technique had several advantages; one included better bulk sample representation given the large volume of colloidal dispersion sampled, due to the high flow rate, estimated to be 75 mL. The high concentration of solids within the colloidal dispersion investigated would have required sample dilution if the aforementioned off-line techniques were utilized; thus the on-line method eliminated this sample preparation need.[51]

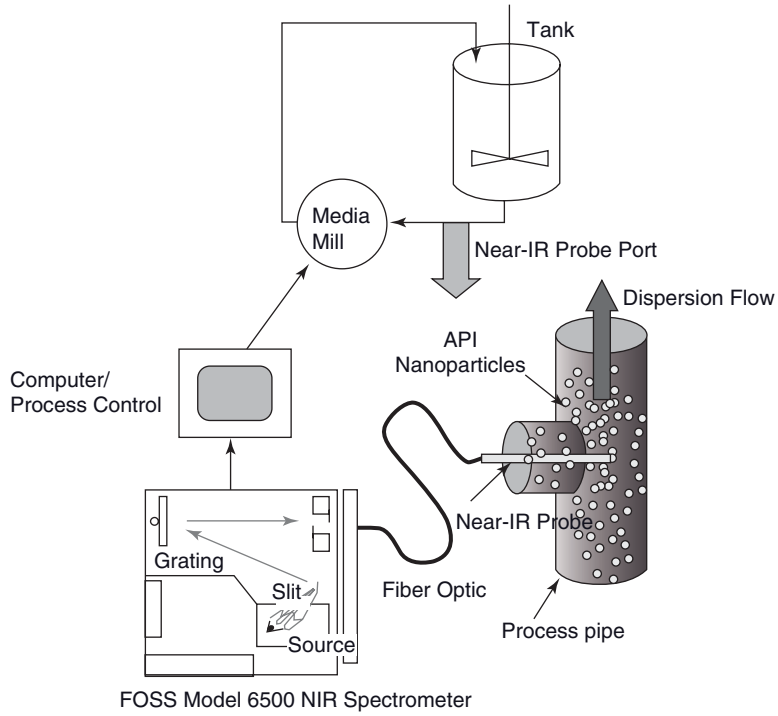


Figure 13.1 *Schematic of the media milling apparatus and the on-line NIR implementation for particle size analysis. Reprinted with permission from Higgins* et al. *(2003)*[51]*. Copyright 2003 American Chemical Society.*

13.6 Hot-melt Extrusion

Hot-melt extrusion (HME) is an increasingly deployed method in the pharmaceutical industry for the preparation of drug delivery systems. The advantages of using HME include, but are not limited to, the following: there is no need to use organic solvent, there are fewer processing steps, and there is the potential of continuous processing.[52] In extruded formulations, a pharmaceutical grade polymer is used. This can be advantageous as earlier noted in regard to commercial-scale manufacturing, but it can also make the extraction of API from the film for analysis complicated.[53] Given this necessity for on-line monitoring, NIRS and Raman have been investigated for API content determination in HME film matrices.[52,53] NIRS was applied off-line by Tumuluri and associates in 2004 where the API clotrimazole was assayed in a polyethylene oxide film. The calibration model ranged from 0 to 20% clotrimazole and yielded predictions within 3.5% error, as compared to the HPLC reference method. The results of this study suggest that on-line implementation of NIRS for API quantitation within HME film matrices is feasible.[52] In 2008, Tumuluri *et al*. did a study utilizing off-line and on-line Raman to quantitate APIs in two extruded formulations. Refer to Figure 13.2 for the experimental set-up used for on-line Raman monitoring. The study yielded errors of prediction equal to 0.97% and 0.94% for the two APIs using the on-line method. The authors also demonstrated model transferability by utilizing two extruders, one lab-scale and the other pilot-scale. Inadvertently during this study, a physical change in a film was detected via a Raman band shift. The root cause of this change is under further investigation, but the fact that this change was detected in real-time further validates on-line Raman for monitoring HME processes.[53]

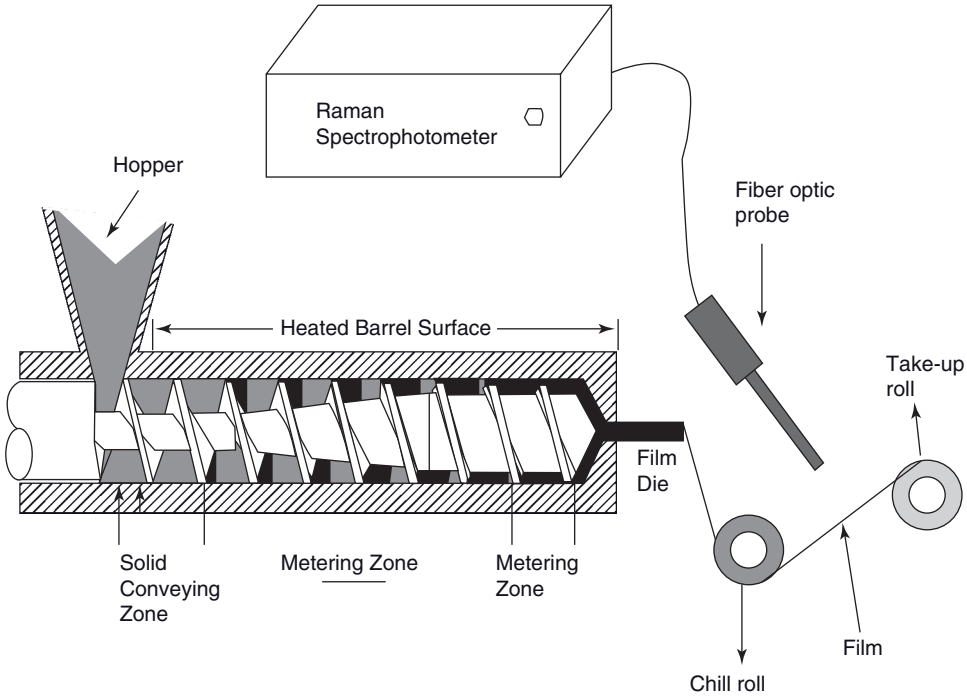


Figure 13.2 *Noncontact optic with the fiber-optic probe clamped over the extruded film formation. Reprinted from Tumuluri* et al. *(2008)[53]. Copyright 2008, with permission from Elsevier Science.*

13.7 Granulation

Within the pharmaceutical industry, granulation is a widely used process to prepare drug substance and excipient particles for compression.[54,55] There are wet and dry granulation procedures; herein high shear wet granulation, fluid bed granulation and roller compaction will be discussed.

13.7.1 Wet granulation

Wet granulation involves the introduction of binder and solvent to pharmaceutical powders while mixing, followed by drying and milling steps, for the purpose of improving processability. There are several contributing process parameters that affect wet granulation processes, but the most notable for fluid bed granulation and high shear wet granulation are spraying conditions and process time, respectively.[54] Several on-line spectroscopic applications have been deployed to monitor and control granulation processes.

Several articles were published in the 1990s describing the use of on-line NIRS for moisture determination during top spray granulation. These include but are not limited to those authored by Watano *et al*., Rantanen *et al*. and Frake *et al*.[56–58] Frake and associates applied on-line NIRS to obtain granule moisture content and particle size change data during fluid bed granulation. Due to the complexity of modeling particle size, the authors did not report the capability to predict particle size via NIRS but rather reported a qualitative assessment to support the primary moisture content predictions. The target granule moisture content is 9.0–10.5% w/w for the end of the spraying phase. The standard error of calibration (SEC) achieved by the NIRS method described was 0.5% w/w and thus the authors report that this metric demonstrated the level

of control required for the given moisture content specification.[57] Rantanen *et al*. reported standard error of predictions (SEPs) equal to 0.11 and 0.20% w/w moisture via NIRS for the spraying and drying phases, respectively.[58]

In 2000, Whitaker *et al*. used on-line passive acoustic emission spectroscopy to monitor a high shear wet granulation process. The authors report that this technique is capable of monitoring changes in powder properties, such as particle size, powder flow and compressibility. Piezoelectric acoustic emission sensors covering frequencies from 20 to 350 kHz were adhered to the outside, bottom of the high shear mixer bowl. Acoustic emission data collected during nine pilot scale granulations, with varying amounts of binder solution to control granule growth, were used to build a PLS calibration with the following reference data: median particle size, Carr's compressibility index and compression properties. The adequate correlations found between the acoustic emission data and the median particle size and Carr's compressibility index, prior to dry screening, suggest that this technique has potential for granulation end-point control.[59]

Although an indirect measurement of particle properties, the advantage of using acoustic emission spectroscopy to monitor high shear wet granulation is that the sensor(s) are located on the outside of the granulator bowl. Granulation involves spraying particles with binder such that they adhere to one another to form granules, and consequently, the potential for probe fouling if within the bowl in this processing environment is high if probe location and angle are not optimized. Self-cleaning probes are another option for such 'sticky' processing conditions.

Watano used an on-line particle imaging probe to directly measure particle growth and shape during both high shear wet granulation and fluid bed granulation. The probe consisted of a CCD camera, optical fibers for lighting and a telephoto lens. To counter the aforementioned risk of probe fouling during wet granulation, the probe utilized a heated air purge. Refer to Figures 13.3 and 13.4 for illustrations of the image probe to unit operation interfaces. When the feasibility of measuring particle growth was established, Watano then applied fuzzy logic to the real-time image data to yield precise granulation end-point control.[60]

In 2005, Rantanen *et al.* used on-line NIRS to monitor a laboratory-scale high shear wet granulation process. The study was a central-composite circumscribed design consisting of 17 runs and 3 variables:

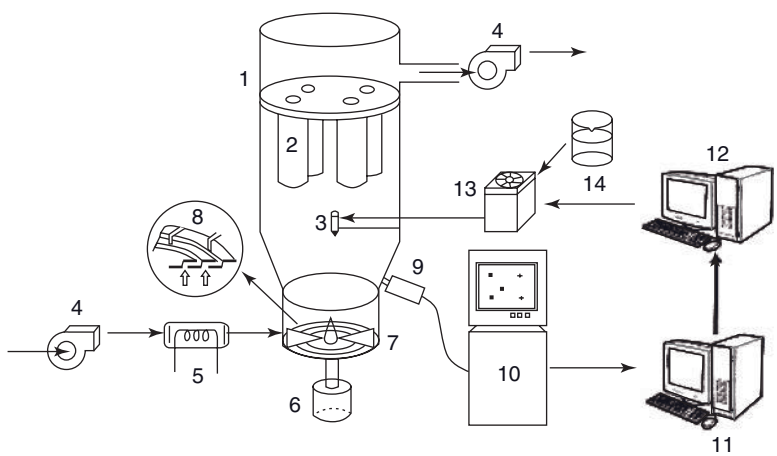


Figure 13.3 *A schematic diagram of agitation fluidized bed granulator (NQ-160, Fuji Paudal). 1: Agitation fluidized bed. 2: Bag filter. 3: Spray nozzle. 4: Blower. 5: Heater. 6: Motor. 7: Agitator blade. 8: Slit plate. 9: Image probe. 10: Image processing system. 11: Host computer. 12: Controller. 13: Pump. 14: Binder. Reprinted from Watano* et al. *(2001)[60] with permission from Elsevier Science.*

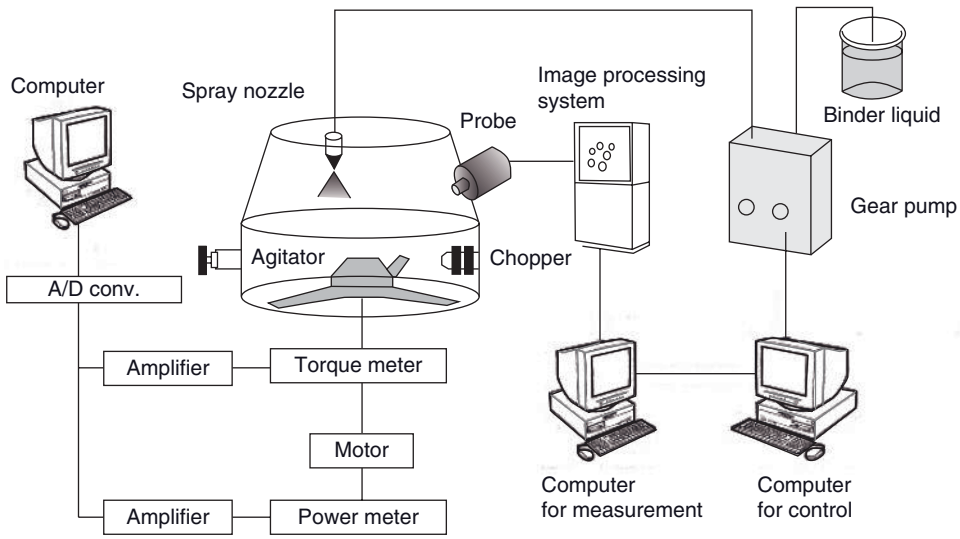


Figure 13.4 *Experimental set-up for high shear granulation control system. Reprinted from Watano* et al. *(2001)*[60] *with permission from Elsevier Science.*

impeller speed (rpm), water amount (g) and wet massing time (min). The on-line NIR measurement was performed through an aperture in the granulator lid. The aperture was covered in glass that was rotated. Both efforts were undertaken to avoid probe fouling. On-line NIRS combined with multivariate data analysis was successfully used to determine the end points of the three phases of granulation (mixing, spraying and wet massing), develop a model for homogeneity monitoring and predict mean granule size.[61] Luukkonen *et al*. also used NIRS to monitor high shear wet granulation and were able to measure the state of the granules, more specifically the particle size, density and the amount of water on the surface. The authors note that NIR spectral data collected during granulation provided a better correlation to granule bulk density, granule median particle size, granule porosity (intra) and tablet hardness as compared to the more traditional granulation monitoring metric, power consumption.[62]

The above-mentioned on-line spectroscopic techniques have involved granulation monitoring for particle size assessments and end-point control; however there are other concerns involved with wet granulation which include polymorphic transformations or solvate formation that can affect end product quality.[55,63,64] As such, Wikström and associates used on-line Raman spectroscopy to monitor hydrate formation during a high shear wet granulation process. The solvent-mediated phase transformation of theophylline anhydrate to monohydrate was clearly observed by the disappearance of anhydrate peaks at 1664 and 1707 cm^{-1} and the appearance of the monohydrate peak at 1686 cm^{-1}. On-line Raman monitoring was used to track theophylline monohydrate formation while varying the granulation process variables, thus facilitating process understanding.[64]

Given that the introduction of solvent obviously produces wet granules, the wet granulation process includes a drying step. Drying typically occurs in a fluid bed dryer. But there are other options, such as microwave vacuum dryers. In 1994, White applied on-line NIRS to monitor moisture content and predict drying end point in two TK Fielder microwave dryers. The NIR spectral data were correlated to off-line Karl Fischer measurements which resulted in a standard error of prediction equal to 0.6% when the samples

were below 6% moisture.[65] The use of on-line NIRS for moisture determination, and thus end-point control, during fluid bed drying have been described by Watano *et al*. and Rantanen *et al*.[56,58] Scientists at Merck & Co., Inc. investigated on-line NIRS for moisture content determination during fluid bed drying but took it a step further by detailing sampling effects. They found that (1) air dynamics within the fluid bed dryer affected spectral quality and (2) the method of reference sampling has a profound impact on apparent error, total analytical error and the number of samples required for moisture prediction model development. They then add that on-line fluid bed drying end-point control based on multiple sample volumes, either by the analysis of flowing sample or by examination of multiple single volume stationary samples as allowable by on-line analysis, are inherently more representative of drying as compared to sparsely thieved in-process control Karl Fischer or LOD samples.[66]

As noted earlier, high shear wet granulation can generate stress in the material which could potentially yield a process-induced transformation. The same is true for fluid bed drying and as a result on-line spectroscopy has been used to track such transformations. As Raman was used by Wikström and associates to monitor theophylline monohydrate formation during a high shear wet granulation process, Aaltonen and associates used on-line NIRS and Raman spectroscopy to monitor the conversion of theophylline monohydrate to theophylline anhydrate during fluid bed drying.[67] Raman spectroscopy was also used as an on-line method for monitoring risedronate sodium hydration state during fluid bed drying which yielded process understanding via (1) knowledge of the relationship between the risedronate hydration state and tablet stability and (2) identifying final granule moisture as a critical process parameter.[68]

In summary, acoustic emission and imaging techniques have been used to monitor high shear wet granulation, NIR and imaging have been used to monitor fluid bed granulation and NIR has been used to monitor fluid bed drying. These on-line techniques have focused on monitoring physical properties such as particle size growth but have also been used to monitor chemical changes, such as moisture content and process induced transformations. The simultaneous assessment of physical and chemical properties via NIR clearly gives this technique an advantage over traditional means of controlling wet granulation, i.e. impeller load, power consumption, exhaust air temperature or a fixed time. It is noteworthy that other techniques such as FBRM technology and microwave resonance are other means to monitor wet granulations in real-time, although not detailed herein.[69–71]

13.7.2 Roller compaction

Roller compaction utilizes an auger-feed system to uniformly deliver powder for compaction between two pressure rollers where the powders exit as a ribbon which is then milled.[72] In a series of papers, Gupta *et al*. apply on-line spectroscopy to monitor roller compaction. The NIR sensor was focused on the ribbons as they exit the rolls for the on-line measurements. In the first paper, the focus was to use on-line NIRS to estimate post-milled particle size distribution. The results suggest that the slope values of the best-fit line through the NIR spectrum offers a robust quality control tool to monitor, control and scale-up roller compaction processes.[73] Noteworthy is that Kirsch and Drennen first reported this calibration approach based upon the change in slope of the best-fit line through a NIR spectrum as a means to predict tablet hardness in 1996.[74] In a second paper the authors correlate on-line NIR data collected during a uniaxial compression and roller compaction via PLS to moisture content, relative density, tensile strength and the Young's modulus with SEPs of 0.8%w/w, 0.08, 1.5 MPa and 0.4 GPa, respectively.[75] The focus of the last paper in this series was the prediction of content uniformity via on-line NIRS and PLS where experimental design was used to generate calibration samples. The variables of the experimental design were the active concentration, the relative humidity and the compaction pressure. The SEP for the prediction of active acetaminophen concentration was approximately 1% w/w.[76]

13.8 Powder Blending

The objective of powder blending is to achieve uniformity at an appropriate mass scale of the API(s) and excipients(s) prior to compression or encapsulation for assurance of dosage form-to-dosage form reproducibility. Powder blending is routine in solid pharmaceutical dosage manufacturing, but this should not imply that it is simple. Powder mixing is dependent upon several parameters including: blender type and design, scale, and physical properties of the constituents (shape, size, density, cohesion and electrostatic charge).[77,78] Thus blending at high speeds and for long periods of time does not guarantee blend homogeneity. For the assurance of uniformity and homogeneity, there are the following regulatory expectations: sample size equal to one to three times the product's run weight and a sampling technique representative of all portions and concentrations of the blend.[79]

Typically during product development, a blending process is stopped periodically for the consistent retrieval of samples of appropriate size from defined locations within the powder blend. Afterwards, there is sample preparation and off-line analysis via HPLC or UV-vis spectroscopy. Two major issues with this scenario are sampling error[77,80,81] and time. For a more detailed discussion of sampling issues, refer to Chapter 3 and to the work of Muzzio *et al*.[77]. In regard to time, ironically a 15 min blend can require days before release due to sampling and testing requirements. This prompts the mention of additional issues, such as manufacturing delays and segregation (demixing) during idle times. The content uniformity measurement via HPLC or UV-vis is based upon the API(s); however the uniformity of excipients can also influence end-product quality. All the aforementioned issues indicate room for improvement for blend homogeneity monitoring and blending end-point determination.

Given that NIRS is a rapid, nondestructive technique and that most pharmaceutical blend constituents absorb in the NIR, it has been applied off-line to assess blend homogeneity.[82] In the mid-1990s scientists at Pfizer utilized on-line NIRS to successfully characterize powder blending.[83–85]

The application of NIRS does not resolve all concerns associated with accurate blend uniformity determination. Key questions arise, such as (1) is the NIR beam sampling within a unit dose and (2) are an effective number of unit doses being sampled throughout a blend? These questions are not limited to NIRS, but are applicable to any diffuse reflectance optical technique.

In regard to the former question, Cho and associates suggest that for most pharmaceutical formulations, the effective mass sampled by NIRS meets the FDA requirement of no more than three times the unit dose.[86] For more information regarding the estimation of effective sample size when using diffuse reflectance NIRS, refer to the work of Berntsson *et al*.[87] The availability of spectrometers with various spot sizes allows some flexibility in controlling the effective sample size. Another effective sample size consideration commonly encountered when applying on-line spectroscopic methods during blending is in regard to material sticking to the optical window(s). Lai and associates discuss this issue and comment that powder blends that were to significantly coat the optical windows within a blending operation would more than likely be considered a poor formulation.[88]

In regard to the question, 'Are an effective number of unit doses being sampled throughout a blend?' it greatly depends on unit dose and blend time. A potential predicament for a product with a high drug load is that on-line NIRS suggests blend uniformity is achieved within three minutes yet it takes greater than three minutes of blending to achieve the effective number of samples. Muzzio *et al.* recommended that for poorly mixed powder blends it is better to focus on the number of samples rather than the sample size. The authors elaborate that sample size is irrelevant unless it is representative of the scale of segregation for the given blend.[77] El-Hagrasy and associates investigated multiple NIR sampling ports and suggest that one sampling port is insufficient for determining blending end point. [89–92] Multiple optical ports on blenders being used in routine, commercial-scale manufacture may not be practical, but as the above-mentioned authors stress, optical port number and location need to be considered when developing an on-line blend

uniformity technique. At the very least, the number of samples should be maximized. Regardless of the number of optical ports utilized, an overall improvement in sampling is realized using on-line spectroscopy as opposed to traditional sample thief methods where the number of samples is typically 10 per time point. For example if one optical port is utilized on a blender operating at a speed of 17 rpm, with one spectrum per rotation and one spectrum sampling half a unit dose, then effectively 42.5 unit doses are sampled in five minutes compared to ten via traditional sampling.

In the series of publications on NIR for blend monitoring by El-Hagrasy and associates, humidity, API concentration and blender speed were studied utilizing a D-optimal experimental design; all three factors were deemed to have a significant effect on blending end-point determination. An evaluation of (1) spectral preprocessing options (standard normal variant transformation, multiplicative scatter correction and second derivative), (2) qualitative pattern recognition models (soft independent modeling of class analogs, SIMCA and principal component modified bootstrap error-adjusted single-sample technique, PC-MBEST), and (3) quantitative models (linear regression, MLR, principal component regression and PLS) was performed. Furthermore, on-line NIR imaging was utilized to assess blend uniformity. Within this study; off-line UV, on-line NIR and on-line NIR imaging yielded comparable blending end-point predictions.[90–93] The advantage of on-line NIR chemical imaging during blending is that it provides simultaneous insight to the concentration and distribution of blend constituents.[93]

NIR methods are not the only on-line applications for blend monitoring; FT-Raman[94,95] and laser induced fluorescence (LIF)[88,96] have been utilized. Refer to Chapter 11 for a comprehensive review of LIF. As stated herein, NIRS is well established as an effective and advantageous means to deem blend homogeneity and blending end point, however there are circumstances in which NIR is insufficient. For example, LIF can be more suitable for blends with low drug load. Lai and Cooney illustrated in a lab-scale experiment that LIF yielded a limit of detection below 0.02% w/w for a given API.[88]

Whether NIR, NIR chemical imaging, Raman or LIF methods are utilized, most will observe when monitoring in real-time that blending end points can vary and hence the necessity to shift from time-defined end points to on-line spectroscopic monitoring of powder blending. In addition, on-line spectroscopic monitoring for blend end-point has the potential for increased process flexibility (i.e. not having to define specifications for blender type and design, scale, and physical properties of the constituents).

13.8.1 Lubrication

Blend uniformity measurements as mandated by the FDA are based solely on API, yet the uniformity of excipients can be essential. A good example of excipient criticality is the distribution of the lubricant (i.e. magnesium stearate), within a blend. If under distributed within the formulation, sticking during compression can result and if overly distributed it can lead to longer dissolution times and decreased tablet strength. Quantitation of magnesium stearate via NIRS has been demonstrated in the literature.[97,98] In an example of real-time magnesium stearate content (%) prediction in placebo and API blends, magnesium stearate was spiked into the blend as a means to emulate unexpected concentration changes of the analyte(s). This experiment was done to illustrate the sensitivity, determined at ±0.1 wt%, of the given NIR method.[98]

13.8.2 Powder flow

Upon completion of the blending operation, the powder blend can be immediately discharged into a hopper for compression or encapsulation. However, the blends may be stored and, depending upon the physical properties of the blend, there is the risk that particles could segregate during storage or post blend processing. Muzzio *et al*. discuss this issue in detail and ask the question, to what extent do such processing conditions (i.e. discharge and transport) affect blend homogeneity.[77] In circumstances where post blend segregation

is a high risk, on-line NIR can be used to monitor powder flow.[99,100] Benedetti and associates developed an on-line, real-time NIR method to characterize the homogeneity of flowing powder mixtures while being discharged from a hopper prior to compression or encapsulation. A PLS model was used to predict API concentration from flowing blend powder spectra. The authors comment that the predicted API values varied over short time scales but these variations were decreased when the data were averaged to be representative of one unit dose.[100]

13.9 Compression

In the 1990s there were several papers that discussed NIRS for the determination of tablet hardness.[74,101–103] Since then there have been a plethora of at-line spectroscopic applications for tablet analysis such as: laser induced breakdown spectroscopy (LIBS) for coating thickness and uniformity,[104] terahertz spectroscopy for the determination of active and excipient concentrations within tablets,[105] NIRS for content uniformity and dissolution prediction of intact tablets,[106] and Raman imaging and multivariate data analysis for characterizing the spatial distribution of constituents within a tablet.[107] In regard to at-line spectroscopic application for encapsulation processes, the use of at-line NIRS and transmission Raman spectroscopy have been reported in the literature.[108–110] Given the speed at which tablets and capsules are produced (i.e. several hundred thousand per hour), on-line spectroscopic applications to characterize or control these processes will be difficult. CK Lai and associates suggested the use of LIF as a potential on-line monitoring technique for tabletting. In their study the LIF technique yielded a penetration depth into the tablet of 3 mm and was capable of predicting total drug content as referenced to UV values. The reported method was used on-line and demonstrated consistent LIF profiles from simulated production rates up to 3000 tablets per minute; hence the authors' recommendation that LIF be implemented as a continuous, on-line and qualitative method for compression monitoring.[111] Ometrics and Uhlmann VisioTec have issued white papers, accessible via their respective web sites, where they have used spectroscopy for 100% real-time, on-line testing during the compression operation.

13.10 Coating

Coating, whether cosmetic or functional, is another process that benefits from on-line spectroscopic monitoring for understanding and control. NIRS, LIBS, and terahertz pulsed image mapping have been used at-line to provide valuable information regarding the coating process.[112–118] In regards to on-line spectroscopic techniques for monitoring the coating process, NIRS has been employed in a fluidized bed process and a pan coater.[118,119] Andersson *et al*. were the first to report NIRS for on-line monitoring of pellets during coating within a fluidized bed in the production of controlled release drug product. The objective of their study was to estimate coating thickness and process end point. A root-mean-squared error of prediction (RMSEP) as good as 2.1 μm for coating thickness was achieved within a calibration range of 0–50 μm. This coating process calibration was justified by good correlation between the NIR estimates and reference image analysis on dissected pellets and a theoretical nonlinear coating thickness growth model. The authors note that this method can enable coating end-point determination with high precision and suggest that the on-line NIR technique may be further developed to provide information on the underlying mechanisms of the coating process, the process dynamics and the particle flow distribution within the fluidized bed. Additionally the authors discuss a shortcut for multivariate batch calibration and suggest that further studies are needed to develop this promising methodology.[118] Pérez-Ramos and associates at Purdue University positioned a NIR probe within a pan coater to collect data for modeling film growth.[119] They were the first to report on-line

monitoring within a pan coater for end-point prediction. To protect the NIR probe from coating droplets, the top and sides of the sensor were enclosed in a Plexiglass box with a nitrogen purge. In their article the authors outline twelve steps to monitor a 0–11% weight gain coating process for biconvex tablets, where a mean coating thickness of up to 180 mm and 135 mm could be achieved on the tablet face and tablet band, respectively. They report being able to predict film growth in the tablet face and band regions as a function of time and to estimate an accurate film density from a new batch based solely on the on-line NIR measurement.[119] As with the finding by Andersson *et al*., Pérez-Ramos *et al*. suggest that the on-line NIR technique could be developed to better understand the extent of coating variability within individual pellets/tablets.[120–122]

13.11 Biologics

As the pharmaceutical industry shifts from product pipelines consisting entirely of small molecules to those with both small and large molecules, PAT as applied to biopharmaceutical manufacturing will hopefully increase in both the pharmaceutical and, of course, biotechnology industries. The operative word is increase, given that several on-line spectroscopic technologies have been deployed in various bioprocesses. This section will focus on two areas of application within biological processing: fermentation and freeze-drying.

13.11.1 Fermentation

The need for real-time sensors for fermentation monitoring and control has long been noted in the literature.[120,121] In 1990, Cavinato *et al.* applied on-line NIRS to a fermentation process for ethanol concentration prediction where GC was used as the reference method. The NIR fiber-optic probe was positioned outside of a glass fermentation vessel, thus eliminating sampling and sterility issues. This study illustrated the potential of NIRS for fermentation monitoring, noting that the developed calibration was robust yet cautioning that the investigated technology was not capable of monitoring ethanol at trace levels as would be expected in aerobic fermentations.[122] In a similar experiment the authors were able to determine cell density of yeast fermentations with an average SEP of 1.6 g/L. Noteworthy is that the cell density predictions were only valid when the stirring and bubbling rates remained constant.[122] Arnold *et al*. further challenged the application of NIRS to fermentation processes by measuring *in situ*, under aerobic conditions and mimicking the more vigorous aeration and agitation conditions expected when operating on an industrial scale. A steam-sterilizable NIR probe was used and, despite the challenges of *in situ* monitoring under harsh processing conditions, the study illustrated that biomass monitoring is feasible. Due to the reduction in spectral quality when collecting NIRS data *in situ*, the authors recommend signal optimization studies.[123] Two later studies also applied *in situ* NIRS to aerobic fermentations, each reporting success at estimating biomass, concentration[124] and viscosity.[125] The latter study performed *in situ* fermentation monitoring at both the pilot scale and industrial scale.[125]

Mid-infrared spectroscopy has also been demonstrated as a technique for monitoring fermentation.[126–128] One such study details the effects of temperature variation on substrate and metabolite concentration predictions, and used an artificial neural network creating a nonlinear multivariate model to improve concentration predictions.[126] Another study notes the effects of temperature on the mid-infrared spectral data as well, but also noted that the sensor was not affected by reactor operating conditions such as agitation, airflow and backpressure.[127]

For further details regarding chemometric techniques as applied to on-line spectroscopic data collected during fermentation, refer to publications by Blanco *et al*. and Triadaphillou *et al*.[128,129] The former study

notes the use of NIRS and multivariate curve resolution – alternating least squares in combination with hard modeling information as spectral equality constraints for the prediction of glucose, ethanol and biomass in fermentation process where the authors note the potential of the given method for other bioprocesses.[128] In the latter study, a new calibration approach for monitoring fermentation is presented that adds a wavelength selection procedure.[129]

13.11.2 Freeze-drying

Freeze-drying, also known as lyophilization, is typically performed on heat sensitive drug products, such as biologics, to remove water and thus enhance stability. This lengthy process has three stages: (1) freezing, (2) primary drying where the pressure is lowered to instigate sublimation thus removing unbound water and then (3) secondary drying where the temperature is elevated and typically the pressure is lowered further to initiate desorption thus removing bound water. Temperature and time are typically used for freezing, primary drying and secondary drying end-point control; however, monitoring the chemical phase changes via on-line spectroscopy is another means for process control. As with drug substance drying procedures, on-line spectroscopy has been deployed to directly measure the physical changes within a sample and to indirectly determine moisture removed via monitoring the vapor stream. Although possible, directly measuring drug product can be complicated in a lyophilization chamber since the bulk material is partitioned into numerous vials which are placed on multiple shelves. NIRS[130,131] and Raman spectroscopy[131,132] have been used to qualitatively and quantitatively monitor a single vial of material in real-time within a lyophilization chamber. The authors note a disadvantage of the setup is that only one sample vial can be monitored[130,131] and thus stress the importance of sample selection and presentation.[130]

Gieseler *et al*. utilized tunable diode laser absorption spectroscopy to detect water vapor concentrations, gas velocities and mass flow during freeze-drying of pure water at different pressure and shelf temperature settings and of a 5% w/w mannitol solution. The analyzer was interfaced to the spool that connected the dryer chamber to the condenser. The reported method was advantageous in that primary and secondary drying end-point control based upon mass flow rate was independent of freeze-dryer size and configuration.[133]

13.12 Cleaning Validation

Pharmaceutical manufacturing facilities typically run multiple batch processes where the equipment is common to several drug products. A recall of drug product in 1988 and an import alert for a foreign bulk pharmaceutical manufacturer in 1992 increased awareness within the FDA that cross-contamination could pose a serious health risk to the public.[134] Consequently prior to switching production of one drug substance/drug product to another drug substance/drug product, equipment cleaning and cleaning validation are required. The cleaning validation should not only detect the absence of the former API, but also detect the absence of residues of reaction byproducts and degradants from the previous process and residues from the cleaning process itself (i.e. detergents, solvents).[134]

Given that cleaning and cleaning validation can lead to long equipment downtimes, on-line spectroscopic techniques have been explored as a rapid means to confirm manufacturing equipment cleanliness. Such examples have included *in situ* Fourier transform infrared reflection–absorption spectroscopy (IRRAS) by Mehta *et al*. who noted that this method is solvent-free and more accurate relative to traditional methods.[135] Scientists at the University of Canterbury and Remspec Corporation have also used *in situ* IRRAS, with their latter study simultaneously quantitating two chemically similar APIs on a glass surface. In all they have applied this technique to detect and quantitate contaminants on glass, aluminum and stainless steel

surfaces.[136,137] Hellings and Vanbaelen used an on-line UV analyzer during the cleaning process of a fully automatic chromatographic system to quantitate the amount of residual product, but also used this analyzer in combination with process data to optimize the cleaning process.[138]

13.13 Conclusions

The FDA's PAT guidance lists the following PAT tools to enable process understanding, continuous improvement and risk-mitigation strategies: (1) multivariate tools for data acquisition and analysis, (2) process analyzers, (3) process control tools and (4) continuous improvement and knowledge management tools.[2] Within the pharmaceutical industry, NIRS, Raman spectroscopy, FTIR spectroscopy, IRRAS, UV spectroscopy, LIF, LIBS, terahertz spectroscopy, tunable diode laser absorption spectroscopy, acoustic emission and chemical imaging are all examples of process analyzers that have been used at-line or on-line with success. As noted earlier, representative on-line spectroscopic methods detailed herein include all four components as given by the FDA. Most of the given on-line PAT applications of spectroscopy included multivariate tools; additionally some have included design of experiments and have integrated process data into the method. Several of the noted spectroscopic methods have been used for process control but to state that on-line spectroscopic methods are routinely controlling pharmaceutical manufacturing processes is overly ambitious. If continuous improvement and knowledge management tools are viewed as a means to collate various data streams and to convert this collection of data to process knowledge, then the given on-line applications are and will be utilizing this as well. Simply put, a spectrometer is a process analyzer but an on-line spectroscopic method is much more. Lastly, there is room for improvement regarding PAT in pharmaceutical manufacturing, but as evident by the spectroscopic methods given herein, great progress has been made.

References

1. US Food and Drug Administration, Pharmaceutical cGMPS for the 21st Century – A Risk-Based Approach: Second Progress Report and Implementation Plan, 2003, available http://www.fda.gov/cder/gmp/2ndProgressRept_Plan.htm accessed 27 November 2009.
2. US Food and Drug Administration, PAT – A Framework for Innovative Pharmaceutical Manufacturing and Quality Assurance, 2004, available http://www.fda.gov/CDER/GUIDANCE/6419fnl.htm accessed 27 November 2009.
3. L. Dunlap, FDA backs PAT to help solve manufacturing woes: process analytical technology is a key part of Agency's plan for the 21st century, *Genetic Engineering & Biotechnology News*, **28** (16), (2008), available http://www.genengnews.com/articles/chitem.aspx?aid=2596&chid=3 accessed 27 November 2009.
4. R.G. Manning, S.Z. Wahab, W.E. Brown, W.W. Hauck, S. Schuber and R.L. Williams, Dissolution testing and metrological measurement of quality for solid oral dosage forms: USP applies metrological principles to the dissolution procedure alone and in collaborative studies to understand and minimize potential sources of variability, 2007, available http://pharmtech.findpharma.com/pharmtech/InsideUSP/article/detail/423552 accessed 27 November 2009.
5. M.T. Riebe, D.J. Eustace, Process analytical-chemistry – an industrial perspective, *Anal. Chem.*, **62** (2), A65–A71 (1990).
6. J.B. Callis, D.L. Illman and B.R. Kowalski, Process analytical chemistry, *Anal. Chem.*, **59**, 624A–635A (1987).
7. K.R. Beebe, W.W. Blaser, R.A. Bredeweg, *et al.*, Process analytical chemistry, *Anal. Chem.*, **65** (12), R199–R216 (1993).
8. W.W. Blaser, R.A. Bredeweg, R.S. Harner, *et al.*, Process analytical chemistry, *Anal. Chem.*, **67** (12), R47–R70 (1995).
9. J. Workman, D.J. Veltkamp, S. Doherty, *et al.*, Process analytical chemistry, *Anal. Chem.*, **71** (12), 121R–180R (1999)

10. J. Workman, K.E. Creasy, S. Doherty, L. Bond, M. Koch, A. Ullman and D.J. Veltkamp, Process analytical chemistry, *Anal. Chem.*, **73** (12), 2705–2718 (2001).
11. J. Workman, M. Koch and D. Veltkamp, Process analytical chemistry, *Anal. Chem.*, **75** (12), 2859–2876 (2003).
12. J. Workman, Jr., M. Koch and D. Veltkamp, Process analytical chemistry, *Anal. Chem.*, **77** (9), 3789–3806 (2005).
13. J. Workman, M. Koch and D. Veltkamp, Process analytical chemistry, *Anal. Chem.*, **79** (12), 4345–4363 (2007).
14. Z.H. Ge, R. Thompson, S. Cooper, D. Ellison and P. Tway, Quantitative monitoring of an epoxidation process by Fourier-Transform infrared spectroscopy, *Process Contr. Qual.*, **7**(1), 3–12 (1995).
15. M. Cameron, G.X. Zhou, M.B. Hicks, *et al.*, Employment of on-line FT-IR spectroscopy to monitor the deprotection of a 9-fluorenylmethyl protected carboxylic acid peptide conjugate of doxorubicin, *J. Pharm. Biomed. Anal.*, **28**, 137–144 (2002).
16. Y. Chen, G. Zhou, N. Brown, T. Wang and Z. Ge, Study of lactol activation by trifluoroacetic anhydride via *in situ* Fourier transform infrared spectroscopy, *Anal. Chim. Acta*, **497**(1–2), 155–164 (2003).
17. D. Marziano, C.A. Sharp, P.J. Dunn, and P.A. Hailey, On-line mid-ir spectroscopy as a real-time approach in monitoring hydrogenation reactions, *Org. Process Res. Dev.*, **4**, 357–361 (2000).
18. K. De Braekeleer, A. de Juan, F.C. Sanchez, P.A. Hailey, D.C.A. Sharp, P. Dunn and D.L. Massart, Determination of the end point of a chemical synthesis process using on-line measured mid-infrared spectra, *Appl. Spectrosc.*, **54**(4), 601–607 (2000).
19. W.W. Howard, S. Sekulic, M.J. Wheeler, *et al.*, On-line determination of reaction completion in a closed-loop hydrogenator using NIR spectroscopy, *Appl. Spectrosc.*, **52**(1), 17–21 (1998).
20. Z.H. Ge, B. Buchanan, J. Timmermans, D. De Tora, D. Ellison and J. Wyvratt, On-line monitoring of the distillates of a solvent switch process by near-infrared spectroscopy, *Process Contr. Qual.*, **11**(4), 277–287 (1999).
21. C. Coffey, B.E. Cooley and D.S. Walker, Real time quantitation of a chemical reaction by fiber optic near-infrared spectroscopy, *Anal. Chim. Acta*, **395** (3), 335–341 (1999).
22. M. Blanco and D. Serrano, On-line monitoring and quantification of a process reaction by near-infrared spectroscopy. Catalysed esterification of butan-1-ol by acetic acid, *Analyst*, **125**, 2059–2064 (2000).
23. Z. Lin, L. Zhou, A. Mahajan, S. Song, T. Wang, Z. Ge and D. Ellison, Real-time end point monitoring and determination for a pharmaceutical salt formation process within in-line FT-IR spectroscopy, *J. Pharm. Biomed. Anal.*, **41**, 99–104 (2006).
24. A.C. Quinn, P.J. Gemperline, B. Baker, M. Zhu and D.S. Walker, Fiber-optic UV/visible composition monitoring for process control of batch reactions, *Chemom. Intell. Lab. Syst.*, **45**, 199–214 (1999)
25. L.X. Yu, R.A. Lionberger, A.S. Raw, R. D'Costa, H. Wu and A.S. Hussain, Applications of process analytical technology to crystallization processes, *Adv. Drug Delivery Rev.*, **56**, 349–369 (2004).
26. V. Liotta and V. Sabesan, Monitoring and feedback control of supersaturation using ATR-FTIR to produce an active pharmaceutical ingredient of a desired crystal size, *Org. Process Res. Dev.*, **8**, 488–494 (2004).
27. F. Lewiner, J.P. Klein, F. Puel and G. Févotte, On-line ATR-FTIR measurement of supersaturation during solution crystallization processes. Calibrations and applications on three solute/solvent systems, *Chem. Eng. Sci.*, **56**, 2069–2084 (2001).
28. G. Févotte, New perspectives for the on-line monitoring of pharmaceutical crystallization processes using *in situ* infrared spectroscopy, *Int. J. Pharm.*, **241**, 263–278 (2002).
29. M. Fujiwara, P.S. Chow, D.L. Ma and R.D. Braatz, Paracetamol crystallization using laser backscattering and ATR-FTIR spectroscopy: metastability, agglomeration, and control, *Cryst. Growth Des.*, **2**(5), 363–370 (2002).
30. K. Pöllänen, A. Häkkinen, S-P. Reinikinen, J. Rantanen, M. Karjalainen, M. Louhi-Kultanen and L. Nyström, IR spectroscopy together with multivariate data analysis as a process analytical tool for in-line monitoring of crystallization process and solid-state analysis of crystalline product, *J. Pharm. Biomed. Anal.*, **38**, 275–284 (2005).
31. C. Starbuck, A. Spartalis, L. Wai, *et al.*, Process optimization of a complex pharmaceutical polymorphic system via *in situ* Raman spectroscopy, *Cryst. Growth Des.*, **2**(6), 515–522 (2002).
32. T. Norris, P.K. Aldridge and S.S. Sekulic, Determination of end-points for polymorph conversions of crystalline organic compounds using on-line near-infrared, *Analyst*, **122** (6), 549–552 (1997).
33. G.X. Zhou, L. Crocker, J. Xu, J. Tabora and Z. Ge, In-line measurement of a drug substance via near infrared spectroscopy to ensure a robust crystallization, *J. Pharm. Sci.*, **95**(11), 2337–2347 (2006).
34. S.-W. Wong, C. Georgakis, G.D. Botsaris, K. Saranteas and R. Bakale, Online estimation and monitoring of diastereomeric resolution using FBRM, ATR-FTIR, and Raman spectroscopy, *Ind. Eng. Chem. Res.*, **47**, 5576–5584 (2008).
35. X. Zhou, P.A. Hines, K.C. White and M.W. Borer, Gas chromatography as a reference method for moisture determination by near-infrared spectroscopy, *Anal. Chem.*, **70**, 390–394 (1998).

36. X. Zhou, P.A. Hines, and M.W. Borer, Moisture determination in hygroscopic drug substances by near infrared spectroscopy, *J. Pharm. Biomed. Anal.*, **17**, 219–225 (1998).
37. I.R. Last and K.A. Prebble, Suitability of near-infrared methods for the determination of moisture in a freeze-dried injection product containing different amounts of the active ingredient, *J. Pharm. Biomed. Anal.*, **11**(11/12), 1071–1076 (1993).
38. B.F. MacDonald and K.A. Prebble, Some applications of near-infrared reflectance analysis in the pharmaceutical industry, *J. Pharm. Biomed. Anal.*, **11**(11/12), 1077–1085 (1993).
39. M.B. Hicks, G.X. Zhou, D.R. Lieberman, *et al.*, *In situ* moisture determination of a cytotoxic compound during process optimization, *J. Pharm. Sci.*, **92**(3), 495–501 (2003).
40. G.X. Zhou, Z. Ge, J. Dorwart, B. Izzo, J. Kukura, G. Bicker and J. Wyvratt, Determination and differentiation of surface and bound water in drug substances by near infrared spectroscopy, *J. Pharm. Sci.*, **92**(5), 1058–1065 (2003).
41. C. Coffey, A. Predoehl, D.S. Walker, Dryer effluent monitoring in a chemical pilot plant via fiber-optic near-infrared spectroscopy, *Appl. Spectrosc.*, **52**(5), 717–724 (1998).
42. S.C. Harris and D.S. Walker, Quantitative real-time monitoring of dryer effluent using fiber optic near-infrared spectroscopy, *J. Pharm. Sci.*, **89**(9), 1180–1186 (2000).
43. J.E. Kipp, The role of solid nanoparticle technology in the parenteral delivery of poorly water-soluble drugs, *Int. J. Pharm.*, **284**, 109–122 (2004).
44. R.H. Muller, R. Becker, B. Kruss and K. Peters, Pharmaceutical nanosuspensions for medicament administration as systems with increased saturation solubility and rate of solution, US Patent 5858410 (1999).
45. R.H. Muller and K. Peters, Nanosuspensions for the formulation of poorly soluble drugs I. Preparation by a size-reduction technique, *Int. J. Pharm.*, **160**, 229–237 (1998).
46. E. Merisko-Liversidge, G.G. Liversidge and E.R. Cooper, Nanosizing: a formulation approach for poorly-water-soluble compounds, *Eur. J. Pharm. Sci.*, **18**, 113–120 (2003).
47. J. Hu, K.P. Johnston and R.O. Williams III, Nanoparticle engineering processes for enhancing the dissolution rates of poorly water soluble drugs, *Drug Dev. Ind. Pharm.*, **30**(3), 233–245 (2004).
48. E.W. Ciurczak, R.P. Tourlini and M.P. Demkowicz, Determination of particle size of pharmaceutical raw materials using near infrared reflectance spectroscopy, *Spectroscopy*, **1**, 36–39 (1986).
49. J.L. Ilari, H. Martens and T. Isaksson, Determination of particle size in powders by scatter correction in diffuse near infrared reflectance, *Appl. Spectrosc.*, **42**, 722–728 (1988).
50. A.J.O Neil, R. Jee and A.C. Moffat, The application of multiple linear regression to the measurement of the median particle size of drugs and pharmaceutical excipients by near-infrared spectroscopy, *Analyst*, **123**, 2297–2302 (1998).
51. J.P. Higgins, S.M. Arrivo, G. Thurau, *et al.*, Spectroscopic approach for on-line monitoring of particle size during the processing of pharmaceutical nanoparticles, *Anal. Chem.*, **75**, 1777–1785 (2003).
52. V.S. Tumuluri, S. Prodduturi, M.C. Crowley, S.P. Stodghill, J.W. McGinity, M.A. Repka and B.A. Avery, The use of near-infrared spectroscopy for the quantitation of a drug in hot-melt extruded films, *Drug Dev. Ind. Pharm.*, **30**(5), 505–511 (2004).
53. V.S. Tumuluri, M.S. Kemper, I.R. Lewis, S. Prodduturi, S. Majumdar, B.A. Avery and M. Repka, Off-line and on-line measurements of drug-loaded hot-melt extruded films using Raman spectroscopy, *Int. J. Pharm.*, **357**, 77–84 (2008).
54. A. Faure, P. York and R.C. Rowe, Process control and scale-up of pharmaceutical wet granulation processes: a review, *Eur. J. Pharm. Biopharm.*, **52**(3), 269–277 (2001).
55. R.P. Patel and A.M. Suthar, Ensuring better control of granulation, available http://www.pharmamanufacturing.com/articles/2008/096.html?page=1 accessed 27 November 2009.
56. S. Watano, K. Terashita and K. Miyanami, Determination of end-point with a complex granulation applying infrared moisture sensor, *Chem. Pharm. Bull.*, **39**(4), 1013–1017 (1991).
57. P. Frake, D. Greenhalgh, S.M. Grierson, J.M. Hempenstall and D.R. Rudd, Process control and end-point determination of a fluid bed granulation by application of near infra-red spectroscopy, *Int. J. Pharm.*, **151**, 75–80 (1997).
58. J. Rantanen, S. Lehtola, P. Ramet, J. Mannermaa, J. Yliruusi, On-line monitoring of moisture content in an instrumented fluidized bed granulator with a multi-channel NIR moisture sensor, *Powder Technol.*, **99**(2), 163–170 (1998).
59. M. Whitaker, G.R. Baker, J. Westrup, P.A. Goulding, D.R. Rudd, R.M. Belchamber and M.P. Collins, Application of acoustic emission to the monitoring and end point determination of a high shear granulation process, *Int. J. Pharm.*, **205**, 79–91 (2000).

60. S. Watano, Direct control of wet granulation processes by image processing system, *Powder Technol.*, **117**(1–2), 163–172 (2001).
61. J. Rantanen, H. Wikström, R. Turner, and L.S. Taylor, Use of in-line near-infrared spectroscopy in combination with chemometrics for improved understanding of pharmaceutical processes, *Anal. Chem.*, **77**(2), 556–563 (2005).
62. P. Luukkonen, M. Fransson, I.N. Björn, J. Hautala, B. Lagerholm and S. Folestad, Real-time assessment of granule and tablet properties using in-line data from a high-shear granulation process, *J. Pharm. Sci.*, **97**(2), 950–959 (2008).
63. K. Morris, U. Griesser, C. Eckhardt and J. Stowell, Theoretical approaches to physical transformations of active pharmaceutical ingredients during manufacturing processes, *Adv. Drug Delivery Rev.*, **48**(1), 91–114 (2001).
64. H. Wikström, P.J. Marsac and L.S. Taylor, In-line monitoring of hydrate formation during wet granulation using Raman spectroscopy, *J. Pharm. Sci.*, **94**(1), 209–219 (2005).
65. J.G. White, Online moisture detection for a microwave vacuum dryer, *Pharm. Res.*, **11**(5), 728–732 (1994).
66. R.L. Green, G. Thurau, N.C. Pixley, A. Mateos, R.A. Reed and J.P. Higgins, In-line monitoring of moisture content in fluid bed dryers using near-IR spectroscopy with consideration of sampling effects on method accuracy, *Anal. Chem.*, **77**(40), 4515–4522 (2005).
67. J. Aaltonen, K. Kogermann, C.J. Strachan and J. Rantanen, In-line monitoring of solid-state transitions during fluidization, *Chem. Eng. Sci.*, **62**, 408–415 (2007).
68. D.S. Hausman, R.T. Cambron and A. Sakr, Application of on-line Raman spectroscopy for characterizing relationships between drug hydration state and tablet physical stability, *Int. J. Pharm.*, **299**, 19–33 (2005).
69. C. Buschmüller, W. Wiedey, C. Döscher, J. Dressler and J. Breitkreutz, In-line monitoring of granule moisture in fluidized-bed dryers using microwave resonance technology, *Eur. J. Pharm. Biopharm.*, **69**, 380–387 (2008).
70. X. Hu, J.C. Cunningham and D. Winstead, Study growth kinetics in fluidized bed granulation with at-line FBRM, *Int. J. Pharm.*, **347**, 54–61 (2008).
71. High Shear Wet Granulation, Mettler Toledo website, available http://us.mt.com/mt/filters/applications_autochem-applications_formulation-autochem_formulation-granulation/High_Shear_Granulation_FilterHierarchy-ApplicationFamily_1198795528152.jsp accessed 27 November 2009.
72. M.D. Tousey, The Granulation Process 101: Basic Technologies for Tablet Making, available http://pharmtech.findpharma.com/pharmtech/data/articlestandard//pharmtech/432002/36208/article.pdf accessed 27 November 2009.
73. A. Gupta, G.E. Peck, R.W. Miller and K.R. Morris, Nondestructive measurements of the compact strength and the particle-size distribution after milling of roller compacted powders by near-infrared spectroscopy, *J. Pharm. Sci.*, **93**(4), 1047–1053 (2004).
74. J.D. Kirsch and J.K. Drennen, Nondestructive tablet hardness testing by near-infrared spectroscopy: a new and robust spectral best-fit algorithm, *J. Pharm. Biomed. Anal.*, **19**(3–4), 351–362 (1999).
75. A. Gupta, G.E. Peck, R.W. Miller and K.R. Morris, Influence of ambient moisture on the compaction behavior of microcrystalline cellulose powder undergoing uni-axial compression and roller compaction: a comparative study using near-infrared spectroscopy, *J. Pharm. Sci.*, **94**(10), 2301–2313 (2005).
76. A. Gupta, G.E. Peck, R.W. Miller and K.R. Morris, Real-time near-infrared monitoring of content uniformity, moisture content, compact density, tensile strength, and Young s modulus of roller compacted powder blends, *J. Pharm. Sci.*, **94**(7), 1589–1597 (2005).
77. F.J. Muzzio, P. Robinson, C. Wightman and D. Brone, Sampling practices in powder blending, *Int. J. Pharm.*, **155**, 153–178 (1997).
78. T.P. Garcia, S.J. Wilkinson, and J.F. Scott, The development of a blend-sampling technique to assess the uniformity of a powder mixture, *Drug Dev. Ind. Pharm.*, **27**(4), 297–307 (2001).
79. USA vs. Barr Laboratories, 812 F. Supp. 458, 485 (D.N. J. 1993).
80. J. Berman and J.A. Planchard, Blend uniformity and unit dose sampling, *Drug Dev. Ind. Pharm.*, **21**(11), 1257–1283 (1995).
81. J. Berman, A. Schoeneman, J.T. Shelton, Unit dose sampling: a tale of two thieves, *Drug Dev. Ind. Pharm.*, **22**(11), 1121–1132 (1996).
82. D.J. Wargo and J.K. Drennen, Near-infrared spectroscopic characterization of pharmaceutical powder blends, *J. Pharm. Biomed. Anal.*, **14**, 1415–1423 (1996).
83. S.S. Sekulic, H.W. Ward, II, D.R. Brannegan, *et al.*, On-line monitoring of powder blend homogeneity by near-infrared spectroscopy, *Anal. Chem.*, **68**, 509–513 (1996).
84. P.A. Hailey, P. Doherty, P. Tapsell, T. Oliver and P.K. Aldridge, Automated system for the on-line monitoring of powder blending processes using near-infrared spectroscopy Part I. System development and control, *J. Pharm. Biomed. Anal.*, **14**, 551–559 (1996).

85. S.S. Sekulic, J. Wakeman, P. Doherty and P.A. Hailey, Automated system for the on-line monitoring of powder blending processes using near-infrared spectroscopy Part II. Qualitative approaches to blend evaluation, *J. Pharm. Biomed. Anal.*, **17**, 1285–1309 (1998).
86. J.H. Cho, P.J. Gemperline, P.K. Aldridge, S.S. Sekulic, Effective mass sampled by NIR fiber-optic reflectance probes in blending processes, *Anal. Chim. Acta*, **348**, 303–310 (1997).
87. O. Berntsson, L-G. Danielsson and S. Folestad, Estimation of effective sample size when analyzing powders with diffuse reflectance near-infrared spectrometry, *Anal. Chim. Acta*, **364**, 243–251 (1998).
88. C-K. Lai and C.L. Cooney, Application of a fluorescence sensor for miniscale on-line monitoring of powder mixing kinetics, *J. Pharm. Sci.*, **93**(1), 60–70 (2004).
89. A.S. El-Hagrasy, H.R. Morris, F.D Amico, R.A. Lodder and J.K. Drennen III, Near-infrared spectroscopy and imaging for the monitoring of powder blend homogeneity, *J. Pharm. Sci.*, **90**(9), 1298–1307 (2001).
90. A.S. El-Hagrasy, F.D Amico and J.K. Drennen III, A process analytical technology approach to near-infrared process control of pharmaceutical powder blending. Part I: D-optimal design for characterization of powder mixing and preliminary spectral data evaluation, *J. Pharm. Sci.*, **95**(2), 392–406 (2006).
91. A.S. El-Hagrasy, F.D Amico and J.K. Drennen III, A process analytical technology approach to near-infrared process control of pharmaceutical powder blending. Part II: Qualitiative near-infrared models for prediction of blend homogeneity, *J. Pharm. Sci.*, **95**(2), 407–421 (2006).
92. A.S. El-Hagrasy, F.D Amico and J.K. Drennen III, A process analytical technology approach to near-infrared process control of pharmaceutical powder blending. Part III: Quantitative near-infrared calibration for prediction of blend homogeneity and characterization of powder mixing kinetics, *J. Pharm. Sci.*, **95**(2), 422–434 (2006).
93. H. Ma and C. Anderson, Characterization of pharmaceutical powder blends by NIR chemical imaging, *J. Pharm. Sci.*, **97**(8), 3305–3320 (2008).
94. G.J. Vergote, T.R.M. De Beer, C. Vervaet, J.P. Remon, W.R.G. Baeyens, N. Diericx and F. Verpoort, In-line monitoring of a pharmaceutical blending process using FT-Raman spectroscopy, *Eur. J. Pharm. Sci.*, **21**, 479–485 (2004).
95. D.S. Hausman, R.T. Cambron and A. Sakr, Application of Raman spectroscopy for on-line monitoring of low dose blend uniformity, *Int. J. Pharm.*, **298**, 80–90 (2005).
96. C-K. Lai, D. Holt, J.C. Leung and C.L. Cooney, Real time and noninvasive monitoring of dry powder blend homogeneity, *AIChE J.*, **47**(11), 2618–2622 (2001).
97. N-H. Doung, P. Arratia, F. Muzzio, A. Lange, J. Timmermans and S. Reynolds, A homogeneity study using NIR spectroscopy: tracking magnesium stearate in Bohle Bin-Blender, *Drug Dev. Ind. Pharm.*, **29**(6), 679–687 (2003).
98. A.S. El-Hagrasy, S-Y. Chang and S. Kiang, Evaluation of risk and benefit in the implementation of near-infrared spectroscopy for monitoring of lubricant mixing, *Pharm. Dev. Technol.*, **11**, 303–312 (2006).
99. M.J. Barajas, A.R. Cassiana, W. Vargas, C. Conde, J. Ropero, J. Figueroa and R.J. Romanach, Near-infrared spectroscopic method for real-time monitoring of pharmaceutical powders during voiding, *Appl. Spectrosc.*, **61**(5), 490–496 (2007).
100. C. Benedetti, N. Abatzoglou, J-S. Simard, L. McDermott, G. Leonard and L. Cartilier, Cohesive, multicomponent, dense powder flow characterization by NIR, *Int. J. Pharm.*, **336**, 292–301 (2007).
101. J.D. Kirsch and J.K. Drennen, Near-infrared spectroscopy: applications in the analysis of tablets and solid pharmaceutical dosage forms, *Appl. Spectrosc. Rev.*, **30**(3), 139–174 (1995).
102. J.D. Kirsch and J.K. Drennen, Determination of film-coated tablet parameters by near-infrared spectroscopy, *J. Pharm. Biomed. Anal.*, **13** 1273–1281 (1995).
103. K.M. Morisseau and C.T. Rhodes, Near-infrared spectroscopy as a nondestructive alternative to conventional tablet hardness testing, *Pharm. Res.*, **14**(1), 108–111 (1997).
104. M.D. Mowery, R. Sing, J. Kirsch, A. Razaghi, S. Béchard and R.A. Reed, Rapid at-line analysis of coating thickness and uniformity on tablets using laser induced breakdown spectroscopy, *J. Pharm. Biomed. Anal.*, **28**, 935–943 (2002).
105. H. Wu, E.J. Heilweil, A.S. Hussain and M.A. Khan, Process analytical technology (PAT): quantification approaches in terahertz spectroscopy for pharmaceutical application, *J. Pharm. Sci.*, **97**, 970–984 (2008).
106. M. Blanco, M. Alcalá, J.M. González and E. Torras, A process analytical technology approach based on near infrared spectroscopy: tablet hardness, content uniformity, and dissolution test measurements of intact tablets, *J. Pharm. Sci.*, **95**, 2137–2144 (2006).
107. L. Zhang, M. Henson and S. Sekulic, Multivariate data analysis for Raman imaging of a model pharmaceutical tablet, *Anal. Chim. Acta*, **545**(2), 262–278 (2005).

108. O. Berntsson, G. Zackrisson and G. Östling, Determination of moisture in hard gelatin capsules using near-infrared spectroscopy: applications to at-line process control of pharmaceutics, *J. Pharm. Biomed. Anal.*, **15**, 895–900 (1997).
109. P. Matousek and A.W. Parker, Non-invasive probing of pharmaceutical capsules using transmission Raman spectroscopy, *J. Raman Spectrosc.*, **38**(5), 563–567 (2007).
110. C. Eliasson, N.A. Macleod, L.C. Jayes, F.C. Clarke, S.V. Hammond, M.R. Smith and P. Matousek, Non-invasive quantitative assessment of the content of pharmaceutical capsules using transmission Raman spectroscopy, *J. Pharm. Biomed. Anal.*, **47**, 221–229 (2008).
111. C.K. Lai, A. Zahari, B. Miller, W.E. Katstra, M.J. Cimna and C.L. Cooney, Nondestructive and on-line monitoring of tablets using light-induced fluorescence technology, *AAPS PharmSciTech*, **5** (1), Article 3 (http://www.aapspharmscitech.org/) (2004).
112. J.D. Kirsch and J.K. Drennen, Determination of film-coated tablet parameters by near-infrared spectroscopy, *J. Pharm. Biomed. Anal.*, **13**, 1273–1281 (1995).
113. M.D. Mowery, R. Sing, J. Kirsch, A. Razaghi, S. Béchard and R.A. Reed, Rapid at-line analysis of coating thickness and uniformity on tablets using laser induced breakdown spectroscopy, *J. Pharm. Biomed. Anal.*, **28**, 935–943 (2002).
114. J.A. Spencer, Z. Gao, T. Moore, *et al.*, Delayed release tablet dissolution related to coating thickness by terahertz pulsed image mapping, *J. Pharm. Sci.*, **97**(4), 1543–1550 (2008).
115. M.C. Madamba, W.M. Mullett, S. Debnath and E. Kwong, Characterization of tablet film coatings using a laser-induced breakdown spectroscopic technique, *AAPS PharmSciTech*, **8** (4) Article 103 (http://www.aapspharmscitech.org/) (2007).
116. S.H. Tabasi, R. Fahmy, D. Bensley, C.O Brien and S.W. Hoag, Quality by design, Part II: Application of NIR Spectroscopy to monitor the coating process from a pharmaceutical sustained release product, *J. Pharm. Sci.*, **97** (9), 4052–4066 (2008).
117. S.H. Tabasi, R. Fahmy, D. Bensley, C.O Brien and S.W. Hoag, Quality by design, Part III: Study of curing process of sustained release coated products using NIR spectroscopy, *J. Pharm. Sci.*, **97** (9), 4052–4066 (2008).
118. M. Andersson, S. Folestad, J. Gottfries, M.O. Johansson, M. Josefson and K-G Wahlund, Quantitative analysis of film coating in a fluidized bed process by in-line NIR spectrometry and multivariate batch calibration, *Anal. Chem.*, **72**, 2099–2108 (2000).
119. J.D. Pérez-Ramos, W.P. Findlay, G. Peck and K.R. Morris, Quantitative analysis of film coating in a pan coater based on in-line sensor measurements, *AAPS PharmSciTech*, **6** (1) Article 20 (http://www.aapspharmscitech.org) (2005).
120. D.J. Clarke, M.R. Calder, R.J. Carr, B.C. Blake-Coleman, S.C. Moody and T.A. Collinge, The development and application of biosensing devices for bioreactor monitoring and control, *Biosensors*, **1**(3), 213–320 (1985).
121. A.G. Cavinato, D.M. Mayes, Z.H. Ge and J.B. Callis, Noninvasive method for monitoring in fermentation processes using fiberoptic near-infrared spectroscopy, *Anal. Chem.*, **62**(18), 1977–1982 (1990).
122. Z.H. Ge, A.G. Cavinato and J.B. Callis, Noninvasive spectroscopy for monitoring cell-density in a fermentation process, *Anal. Chem.*, **66**(8), 1354–1362 (1994).
123. S.A. Arnold, R. Gaensakoo, L.M. Harvey and B. McNeil, Use of at-line and *in-situ* near-infrared spectroscopy to monitor biomass in an industrial fed-batch *Escherichia coli* process, *Biotechnol Bioeng*, **80**, 405–413 (2002).
124. S. Tosi, M. Rossi, E. Tamburini, G. Vaccari, A. Amaretti and D. Matteuzzi, Assessment of in-line near-infrared spectroscopy for continuous monitoring of fermentation processes, *Biotechnol. Prog.*, **19**, 1816–1821 (2003).
125. L.O. Rodrigues, L. Vieira, J.P. Cardoso and J.C. Menezes, The use of NIR as a multi-parametric *in situ* monitoring technique in filamentous fermentation systems, *Talanta*, **75**, 1356–1361 (2008).
126. P. Fayolle, D. Picque, B. Perret, E. Latrille, G. Corrieu, Determination of major compounds of alcoholic fermentation by middle-infrared spectroscopy: Study of temperature effects and calibration methods, *Appl. Spectrosc.*, **50**(10), 1325–1330 (1996).
127. D.J. Pollard, R. Buccino, N.C. Connors, T.F. Kirschner, R.C. Olewinski, K. Saini and P.M. Salmon, Real-time analyte monitoring of a fungal fermentation, at pilot scale, using *in situ* mid-infrared spectroscopy, *Bioprocess Biosyst. Eng.*, **24**(1), 13–24 (2001).
128. M. Blanco, A.C. Peinado and J. Mas, Monitoring alcoholic fermentation by joint use of soft and hard modeling methods, *Anal. Chim. Acta*, **556**, 364–373 (2006).
129. S. Triadaphillou, E. Martin, G. Montague, A. Norden, P. Jeffkins and S. Stimpson, Fermentation process tracking through enhanced spectral modeling, *Biotechnol. Eng.*, **97**(3), 554–567 (2007).
130. M. Brülls, S. Folestad, A. Sparén, and A. Rasmuson, In-situ near-infrared spectroscopy monitoring of the lyophilization process, *Pharm. Res.*, **20**(3), 494–499 (2003).

131. T.R.M. De Beer, M. Allesø, F. Goethals, *et al.*, Implementation of a process analytical technology system in a freeze-drying process using Raman spectroscopy for in-line process monitoring, *Anal. Chem.*, **79**, 7992–8003 (2007).
132. S. Romero-Torres, H. Wikstrom, E.R. Grant and L.S. Taylor, Monitoring of mannitol phase behavior during freeze-drying using non-invasive Raman spectroscopy, *PDA J. Pharm. Sci. Technol.*, **61**(2), 131–145 (2007).
133. H. Gieseler, W.J. Kessler, M. Finson, *et al.*, Evaluation of tunable diode laser absorption spectroscopy for in-process water vapor mass flux measurements during freeze drying, *J. Pharm. Sci.*, **96**(7), 1776–1793 (2007).
134. US Food and Drug Administration, Guide to Inspections Validation of Cleaning Processes (1993).
135. N.K. Mehta, J. Goenaga-Polo, S.P. Hernandez-Rivera, D. Hernandez, M.A. Thomson and P.J. Melling, Development of an *in situ* spectroscopic method for cleaning validation using mid-IR fiber optics, *Spectroscopy*, **18**(4), 14 (2003).
136. M.L. Hamilton, B.B. Perston, P.W. Harland, B.E. Williamson, M.A. Thomson and P.J. Melling, Grazing-angle fiber-optic IRRAS for *in situ* cleaning validation, *Org. Process Res. Dev.*, **9**, 337–343 (2005).
137. B.B. Perston, M.L. Hamilton, B.E. Williamson, P.W. Harland, M.A. Thomson and P.J. Melling, Grazing-angle fiber-optic Fourier transform infrared reflection-absorption spectroscopy for *in situ* detection and quantification of two active pharmaceutical ingredients on glass, *Anal. Chem.*, **79**, 1231–1236 (2007).
138. M. Hellings and H. Vanbaelen, The application of PAT for cleaning and cleaning validation, *Am. Pharm. Rev.*, **11**(5), 12–18 (2008).

14

NIR Spectroscopy in Pharmaceutical Analysis: Off-line and At-line PAT Applications

Marcelo Blanco Romía and Manel Alcalá Bernárdez

Department of Chemistry, Faculty of Sciences, Universitat Autònoma de Barcelona (UAB), Barcelona, Spain

14.1 Introduction

Drug manufacturing is an industrial activity comprehensively regulated by both national institutions (pharmacopoeias, Food and Drug Administration (FDA), drug regulatory agencies) and international bodies (International Conference on Harmonisation (ICH), European Medicines Agency (EMA)). Applicable regulations require strict control of not only raw materials, but also production processes. Such control is currently performed mainly by using chromatographic techniques. While chromatography is suitable for most of the analytical processes involved, it has some disadvantages including sluggishness, the need for careful sample preparation and the production of waste with a potential environmental impact, all of which have promoted a search for alternative techniques avoiding or minimizing these problems. Spectroscopic techniques are highly suitable candidates in this respect.

Molecular spectroscopic techniques have been widely used in pharmaceutical analysis for both qualitative (identification of chemical species) and quantitative purposes (determination of concentration of species in pharmaceutical preparations). In many cases, they constitute effective alternatives to chromatographic techniques as they provide results of comparable quality in a more simple and expeditious manner. The differential sensitivity and selectivity of spectroscopic techniques have so far dictated their specific uses. While UV-vis spectroscopy has typically been used for quantitative analysis by virtue of its high sensitivity, infrared (IR) spectrometry has been employed mainly for the identification of chemical compounds on account of its high selectivity. The development and consolidation of spectroscopic techniques have been strongly influenced by additional factors such as the ease of sample preparation and the reproducibility of measurements, which have often dictated their use in quality control analyses of both raw materials and finished products.

Process Analytical Technology 2e Edited by Katherine Bakeev
© 2010 John Wiley & Sons, Ltd.

Spectra differ markedly in appearance depending on the particular technique. UV-vis and near-infrared (NIR) spectra typically consist of broad bands that make unambiguous identification difficult. By contrast, IR and Raman spectra contain a number of sharp bands at well-defined positions that facilitate the identification of specific compounds. Recent improvements in sensitivity and reproducibility of vibrational spectroscopic equipment have substantially expanded the scope of these techniques. Both dispersive and Fourier transform NIR spectrometers allow the development of spectral libraries of high value for the identification of active pharmaceutical ingredients (APIs) and excipients. The ability of some techniques (e.g. IR, NIR and Raman spectroscopies) to provide spectra for undissolved solid samples has considerably expedited analyses and facilitated the determination of additional physical properties such as average particle size, polymorphism, etc. IR and Raman spectra also provide structural information that allows not only the identification of characteristic functional groups facilitating the establishment of chemical structures, but also the unequivocal identification of some compounds. NIR spectra can be obtained without sample preparation and due to this fact NIR has become one of the most promising spectroscopic tools for process monitoring and manufacturing control in the pharmaceutical industry. Also, NIR chemical imaging (NIR-CI) has enabled the identification of critical properties of both raw materials and finished products. (See Chapter 8 for more information on NIR-CI.)

For the above-described reasons, molecular spectroscopic techniques have become the most common choices for pharmaceutical analysis in addition to chromatography. The latter, however, are being gradually superseded by the former in some industrial pharmaceutical processes. Recent technological advances have led to the development of more reliable and robust equipment. The ubiquity of computers has enabled the implementation of highly powerful chemometric methods. All this has brought about radical, widespread changes in the way pharmaceutical analyses are conducted.

Chemometric techniques have gained enormous significance in the treatment of spectral information by virtue of their ability to process the vast amount of data produced by modern instruments over short periods of time with a view to extracting the information of interest they contain and improving the quality of the results. In some cases, the operator is unacquainted with the chemometric techniques (spectral smoothing, baseline correction, etc.) embedded in the software used by the instrument; in others, the chemometric tools involved are inherent in the application of the spectroscopic technique concerned (e.g. in NIR spectroscopy) and thus indispensable to obtaining meaningful results. (See Chapter 12 for more information on chemometrics.)

Because NIR spectroscopy has played the most prominent role in this context, the discussion that follows mainly revolves around it.

14.1.1 Operational procedures

The procedures used to record NIR spectra for samples are much less labor-intensive than those involved in other spectroscopic analytical techniques. NIR spectral information can be obtained from transmission, reflectance and transflectance measurements (Figure 14.1); this allows the measurement process to be adapted to the physical characteristics of the sample and expedites analyses by avoiding the need for sample preparation.

One major difference between NIR spectroscopy and other spectroscopic techniques is its ability to provide spectra for untreated or minimally treated samples. As a result, the instrumentation is adjusted to the characteristics of the samples rather than the opposite.

The absorbance spectrum of a liquid or solution can be readily measured by using quartz or sapphire cuvettes and fiber-optic probes of variable path lengths. The most suitable solvents for this purpose are those not containing O–H, N–H and C–H groups which exhibit little or no absorption in this spectral region, such as $CHCl_3$, CCl_4 or CS_2.

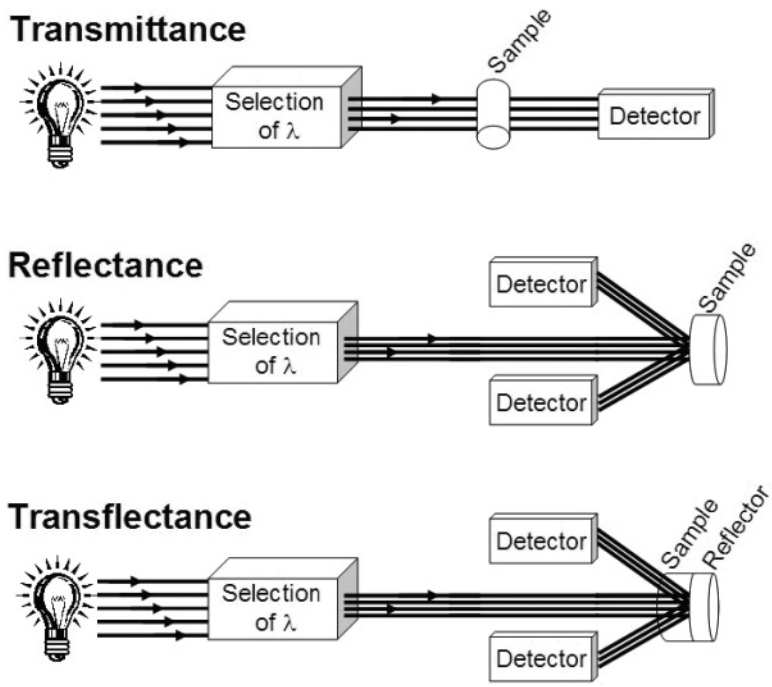


Figure 14.1 *Configuration for spectral acquisition (transmittance, reflectance and transflectance).*

The NIR spectrum for a solid sample can be obtained by using various types of devices. The most frequent choices when the sample is a powder or granules are reflectance cuvettes with a transparent window material (e.g. quartz or glass) and fiber-optic probes. The latter considerably facilitate recording of spectra; however, light losses resulting from transport along the fiber result in increased signal-to-noise ratio at higher wavelengths, when long fiber lengths are used. The spectra for samples in tablet form require no sample preparation such as powdering, sieving or homogenization and can be recorded by using three different types of equipment, namely:

1. Reflectance cuvettes for tablets, which provide spectra subject to strong scattering arising from dead spaces between tablets placed in the cuvette.
2. Devices for recording reflectance spectra for individual tablets.
3. Instruments that allow transmission spectra for individual tablets to be recorded.

It should be noted that the type of standard to be used for reflectance measurements remains a subject of debate. According to ASTM International, the perfect standard for this purpose is a material absorbing no light at any wavelength and reflecting light at an angle identical with the incidence angle. Teflon is the material that fulfills these conditions best.[1] Because no single material meets these requirements, the standards used in this context should be stable, homogeneous, nontransparent, nonfluorescent materials of high, fairly constant relative reflectance. Springsteen and Ricker[2] discussed the merits and pitfalls of materials such as barium sulfate, magnesium oxide, Teflon and ceramic plates as standards for reflectance measurements. Some manufacturers supply National Institute of Standards and Technology (NIST)-certified reference standards for standardizing spectroscopic equipment.[3]

14.1.2 Instrument qualification

Ensuring that the instrument to be used performs correctly is the first step in developing an instrumental methodology. The European Pharmacopoeia[4] and the guidelines of the Pharmaceutical Analytical Sciences Group (PASG)[5] recommend that NIR instruments be qualified as per the manufacturer's instructions, which should include at least the following:

1. Checking for wavelength accuracy by using one or more suitable wavelength standards exhibiting characteristic maxima at the specific wavelength regions of interest.
2. Checking for wavelength repeatability by using one or more suitable standards (e.g. polystyrene or rare-earth oxides). The repeatability of measurements should be consistent with the spectrophotometer specification.
3. Checking for repeatability in the response by using one or more suitable standards (e.g. reflective thermoplastic resins doped with carbon black). The standard deviation of the maximum response should be consistent with the spectrophotometer specification.
4. Checking for photometric linearity by the use of a set of transmittance or reflectance standards (e.g. Spectralon, carbon black mixtures).
5. Checking for the absence of photometric noise by using a suitable reflectance standard (e.g. white reflective ceramic tiles or reflective thermoplastic resins, such as Teflon). The reflectance standard should be scanned in accordance with the recommendations of the spectrophotometer's manufacturer and peak-to-peak photometric noise calculated over a specified wavelength region. The amount of photometric noise should be consistent with the spectrophotometer specification.[6]

Two different instruments will inevitably exhibit small spectral differences that can be reduced by applying standardization corrections based on appropriate references or standards. Reference standardization uses NIST-traceable photometric standards (80% reflectance) to measure the absolute reflectance of the ceramic tile reference at each wavelength on a given instrument. When a sample spectrum is then compared to the ceramic reference, because the absolute absorbance of the reference at each wavelength is known, the absolute absorbance of the sample can be accurately measured – no negative absorbance values are possible. Master standardization uses a sample of similar spectral pattern to that of the products being analyzed to be scanned first on the master instrument and then on the host instrument(s). A standardization file is then created for the host instrument. When a sample is scanned on the host instrument, the spectrum is adjusted using the standardization file to cause it to more closely resemble that measured on the master. This is a 'cloning' algorithm. The above-described corrections allow calibration models to be readily transferred between instruments[7].

14.2 Foundation of Qualitative Method Development

Identifying pharmaceuticals, whether APIs or excipients used to manufacture products, and the end products themselves is among the routine tests needed to control pharmaceutical manufacturing processes. Pharmacopoeias have compiled a wide range of analytical methods for the identification of pharmaceutical APIs and usually several tests for a product are recommended. The process can be labor-intensive and time-consuming with these conventional methods. This has raised the need for alternative, faster methods also ensuring reliable identification. Of the seven spectroscopic techniques reviewed in this book, IR and Raman spectroscopy are suitable for the unequivocal identification of pharmaceuticals as their spectra are compound-specific: no two compounds other than pairs of enantiomers or oligomers possess the same IR

spectrum. However, IR spectrometry is confronted with some practical constraints such as the need to pretreat the sample. The introduction of substantial instrumental improvements and the spread of ATR and IR microscopy techniques have considerably expanded the scope of IR spectroscopy in the pharmaceutical field. Raman spectroscopy, which has been incorporated by the pharmaceutical industry, somewhat more recently, is similar to IR spectroscopy. Some of its more interesting and recent applications are the identification of impurities and adulterants.[8] On the other hand, UV-vis spectra are too uncharacteristic to allow the identification of specific compounds.

NIR spectroscopy, which is subject to none of the previous shortcomings, has been used to monitor all the steps in the drug manufacturing process including chemical synthesis, analysis of raw materials, process intermediates (following blending, grinding, drying, tableting) and preparations, and finally in quality control. More information on this topic can be found in Chapter 13. The quality attributes studied using this technique include identity (authentication), characterization of polymorphs and measurement of mixture uniformity in finished products. Also, chemical composition can be determined without losing any physical information (e.g. particle size, density, hardness), which can be very useful in characterizing some products. The construction of spectral libraries avoids the need to develop a specific method for each individual product and enables sample identification in a simple, direct, expeditious, and economic manner. The whole process can be automated, computer-controlled and conducted by unskilled operators.

14.2.1 Pattern recognition methods

NIR spectral bands are usually broad and strongly overlapping. This precludes the direct identification of compounds by matching spectra to an individual spectrum in a library, which is the usual procedure in IR spectroscopy. Rather, NIR identification relies on pattern recognition methods (PRMs). Essentially, the identification process involves two steps, namely: (1) recording a series of spectra for the product in order to construct a 'spectral library'; and (2) recording the sample spectrum and comparing it to those in the previously compiled library on the basis of mathematical criteria for parameterizing spectral similarity. If the similarity level exceeds a preset threshold, then the spectra are considered to be identical and the sample is identified as the corresponding product in the library. A wide variety of PRM techniques exist for use in specific fields. Figure 14.2 classifies the most commonly used PRMs.

Many PRMs are based on similarity measurements. Similarity here is taken to be the extent to which an object (a spectrum) is identical with another or to a mean of spectra; very frequently, similarity is expressed in terms of correlation[9] or distance.[10] Depending on whether the objects are known to belong to specific classes, PRMs are called 'supervised' or 'unsupervised.' Unsupervised methods search for clustering in an *N*-dimensional space without knowing the class to which a sample belongs. Cluster analysis,[11] minimal spanning tree (MST)[12] and unsupervised (Kohonen) neural networks[13] are among the most common unsupervised PRMs.

Supervised methods rely on some prior training of the system with objects known to belong to the class they define. Such methods can be of the discriminant or modeling types.[14] Discriminant methods split the pattern space into as many regions as the classes encompassed by the training set and establish bounds that are shared by the spaces. These methods always classify an unknown sample as a specific class. The most common discriminant methods include discriminant analysis (DA),[15] the *K*-nearest neighbor (KNN)[16,17] and potential function methods (PFMs).[18,19] Modeling methods establish volumes in the pattern space with different bounds for each class. The bounds can be based on correlation coefficients, distances (e.g. the Euclidian distance in the pattern recognition by independent multicategory analysis methods [PRIMA][20] or the Mahalanobis distance in the unequal [UNEQ] method[21]), the residual variance[22,23] or supervised artificial neural networks (e.g. in the multilayer perceptron[24]).

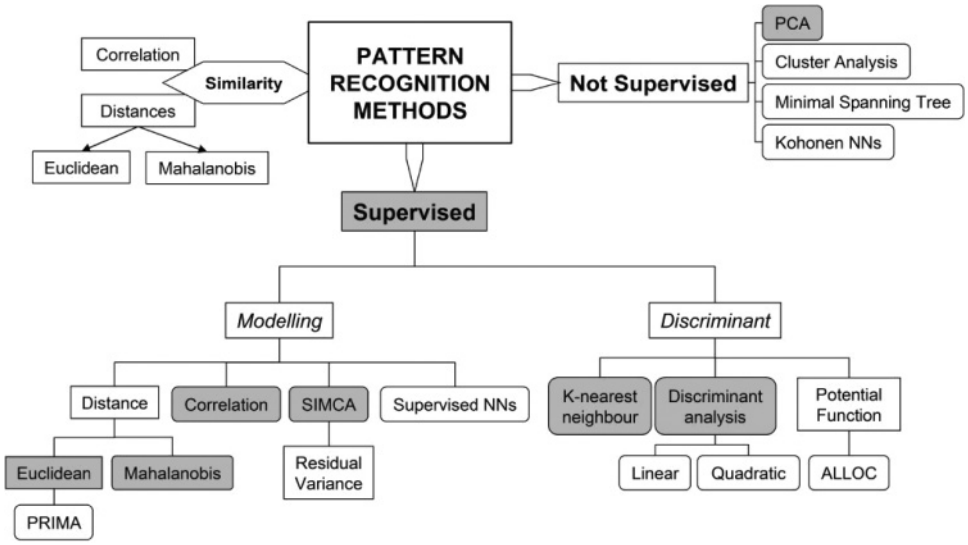


Figure 14.2 *Pattern recognition methods classification used in pharmaceutical identification. Shaded blocks refer to the more commonly used techniques.*

Not every PRM is suitable for constructing spectral identification libraries. These are usually compiled by using supervised modeling methods, and unknown samples are identified with those classes they resemble most.

14.2.2 Construction of spectral libraries

A NIR identification library should encompass all the raw materials used by the manufacturer in order to be able to identify all possible substances and avoid or minimize errors. The method to be used should allow the unequivocal identification of each compound present in the library and the exclusion of those not present. It should also be able to distinguish between very similar compounds used in different applications (e.g. products with different particle sizes, polymorphs, products in different grades or from different suppliers).

Figure 14.3 schematizes the use of spectral information for constructing multivariate calibration models. The pretreatment of the sample spectrum and variable selection processes involve constructing both identification libraries and calibration models for quantifying APIs; however, the samples and spectra should suit the specific aim in each case.

The correlation coefficient (in wavelength space) is especially suitable for constructing a general library as it has the advantage that it is independent of library size, uses only a few spectra to define each product and is not sensitive to slight instrumental oscillations. This parameter allows the library to be developed and validated more rapidly than others. Correlation libraries can also be expanded with new products or additional spectra for an existing product in order to incorporate new sources of variability in an expeditious manner.

One of the crucial factors to ensure adequate selectivity in constructing a spectral library is the choice of an appropriate threshold value or acceptance criteria. Such a threshold is the lowest value (for correlation) required to unequivocally assign a given spectrum to a specific class. Too low a threshold can lead to ambiguities between substances with similar spectra, whereas too high a threshold can result in spectra belonging

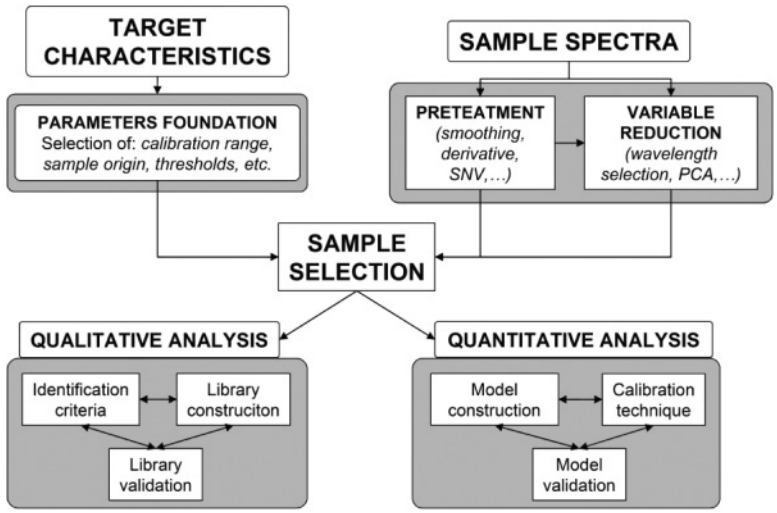


Figure 14.3 *Schema for constructing multivariate models.*

to the same class being incorrectly classified. Choosing an appropriate threshold entails examining the spectra included in the library in an iterative manner. The threshold is successively changed until that resulting in the smallest number of identification errors and ambiguities is reached. In some cases, the threshold thus selected may not distinguish some compounds whose spectra are too similar. This problem can be overcome by identifying the compounds concerned in two steps, using a general library in the first and a smaller one (a subset of the general set with a higher discriminating power) in the second. This methodology has been termed 'cascading identification' as it involves identifying the unknown sample against the general library and, if the result is inconclusive, using a sublibrary for qualification.

This procedure can consist of different stages with one sublibrary being constructed at each stage. The first one enables the identification of the product within an ample library, a second defines a characteristic that produces an important variability in the spectra (e.g. particle size) and a third defines a smaller spectral variability (difference in the content of impurities, origin of manufacture, etc.).

Once the general library has been constructed, those products requiring the second identification step are pinpointed and the most suitable method for constructing each sublibrary required is chosen. Sublibraries can be constructed using various algorithms including the Mahalanobis distance or the residual variance. The two are complementary,[25] so which is better for the intended purpose should be determined on an individual basis.

The procedure to be followed to construct the library (Figure 14.2) involves five steps, namely:

1. *Recording the NIR spectra.* For a set of samples of known identity, either certified by the supplier or confirmed using alternative identification tests, the NIR spectra are recorded in a defined manner.
2. *Choosing spectra.* For each substance, the spectra used to construct the calibration set should be from various batches so that physicochemical variability can be effectively incorporated. Not every spectrum recorded should be used to construct the library as some should be set aside for its validation.
3. *Constructing the library.* First, one must choose the PRM to be used (e.g. correlation, wavelength distance, Mahalanobis distance, residual variance, etc.), the choice being dictated by the specific purpose of the library. Then, one must choose construction parameters such as the math pretreatment (standard

normal variate (SNV), derivatives, etc.), wavelength range and threshold to be used. When a method involving a preliminary variable reduction procedure, such as principal component analysis (PCA) is used, one must also decide upon the values for other parameters such as the number of latent variables, the proportion of variance accounted for, etc. The next step involves validating the library internally in order to check for incorrectly labelled spectra – spectra yielding identification errors or unidentified (or ambiguously identified) substances. Based on the validation results, one must determine if some change in the threshold, spectral range, pretreatment, etc., must be introduced. This is an iterative process that must be repeated until the developer obtains the desired specificity level. Ambiguous identifications should be resolved by developing appropriate sublibraries.

4. *Constructing cascading sublibraries.* Each sublibrary should include all mutually related substances, which will have very similar spectra and may result in ambiguous identification with the general library. An adequate number of representative spectra should be included in the sublibraries in order to ensure accurate description of each class. The number of spectra used should exceed that of spectra required to define an individual class in the general library. Once spectra have been chosen, the process is similar to that followed to construct the general library: an appropriate PRM is selected – the residual variance and the Mahalanobis distance are especially suitable here – the characteristics of the library (namely, threshold, spectral range and pretreatment, proportion of variance accounted for or number of PCs, etc.) are established and calculations performed. Based on the results, the procedure is repeated until all compounds can be unequivocally distinguished.
5. *External validation.* The general library and its sublibraries must be validated by checking that external spectra (for the validation step) are correctly, unambiguously identified. Likewise, samples not present in the library should not be identified with any of the compounds it contains.

14.2.3 Identification and qualification

Control analyses rely on the use of appropriate procedures or measurements assuring the identity of the materials involved in each step of the manufacturing process from receipt of raw materials to delivery of the finished products. NIR spectroscopy is an advantageous alternative to wet chemical methods and instrumental techniques such as IR, Raman and nuclear magnetic resonance (NMR) spectroscopies for positive identification.

The earliest reported studies on the identification of substances by NIR focused on structural characterization, in the vein of many studies in the mid-IR region. However, it is usually difficult to identify a substance by mere visual inspection of its NIR spectrum; rather, a PRM must be applied to a body of spectra (a library) for this purpose. The reliability of the results obtained in the qualitative analysis of a product depends on whether the library is constructed from appropriate spectra. For each product encompassed by the library, one should ensure that the spectra include all potential sources of variability associated with their recording and the sample manufacturing process. The former should be incorporated by including spectra for the same sample recorded by different operators on different days, and the latter by including spectra of samples from different production batches. How many spectra to include in order to construct a 'proper' library is difficult to tell *a priori*. For a product manufactured in a highly reproducible manner, the manufacturing variability can usually be encompassed by samples from 5–10 batches and 20–40 spectra. On the other hand, if the manufacturing procedure is scarcely reproducible, the number of batches and spectra can easily double.

One other important consideration in constructing a library is checking that all spectra are correct. This can be done by visual inspection of the spectra and outlier detection tools. For various reasons (e.g. an inadequately filled cuvette, large temperature fluctuations), spectra may differ from others due to factors other than natural variability. Any such spectra should be excluded from the library. Spectral libraries are

used for two different purposes, namely: identification (of chemical structure) and qualification (of chemical and physical attributes).

Identification involves the confirmation of a certain chemical entity from its spectrum by matching against the components of a spectral library using an appropriate measure of similarity such as the correlation coefficient,[9,26] also known as the spectral match value (SMV). SMV is the cosine of the angle formed by the vectors of the spectrum for the sample and the average spectrum for each product included in the library.

In theory, if the product spectrum coincides with one in the library, then its correlation coefficient should be unity. However, the random noise associated with all spectral measurements precludes an exact match. The SMV has the advantage that it is independent of library size, which facilitates the building of libraries containing large numbers of raw materials as well as correct identifications with libraries consisting of a few spectra for a single product.

However, SMV has a relatively low discriminating power, which may require the use of near-unity thresholds in order to distinguish very similar products. It can also result in ambiguous identification of batches of the same substance with different physical characteristics (e.g. samples of the same product with different grain sizes can have very similar correlation coefficients and may be confused) or different substances with high chemical similarity. In these cases, cascading libraries, as previously discussed, can be used.

NIR spectroscopy not only advantageously replaces other techniques for identification purposes, but also goes one step further in qualitative analysis with its qualification capabilities. The pharmaceutical industry must assure correct dosing and manufacturing of stable products. This entails strictly controlling the raw materials and each of the steps involved in their processing by determining parameters such as potency, moisture, density, viscosity or particle size in order to identify and correct any deviations in the manufacturing process in a timely fashion. Such controls can be achieved by quantifying the different parameters required using appropriate analytical methods or by comparing the NIR spectrum of the sample with a body of spectra of the samples meeting the required specifications and encompassing all possible sources of manufacturing variability. The latter choice is known as 'qualification,' which can be defined as the fulfilment of the product specifications. If the parameter values for a given sample do not fall within the normal variability ranges, then the sample is subjected to thorough analyses. The qualification process can be considered the second stage of an identification process in cascade. Similarity in this context is expressed as a distance. Distance-based methods are subject to two main shortcomings. They use a large number of correlated spectral wavelengths, which increases the distance without providing additional information. They also use more samples than do correlation methods. The Euclidian distance in wavelength space and the Mahalanobis distance – which corrects correlation between spectra – in principal component space are the two most widely used choices.

Distance-based methods possess a superior discriminating power and allow highly similar compounds (e.g. substances with different particle sizes or purity grades, products from different manufacturers) to be distinguished. One other choice for classification purposes is the residual variance, which is a variant of soft independent modeling of class analogy (SIMCA).

14.3 Foundation of Quantitative Method Development

The characteristics of the NIR spectrum require the use of multivariate calibration techniques in order to establish a quantitative relationship between the spectrum and the chemical (composition) and/or physical parameters (particle size, viscosity and density) of the sample.

Calibration is the process by which a mathematical model relating the response of the analytical instrument (a spectrophotometer in this case) to specific quantities of the samples is constructed. This can be done

by using algorithms (usually based on least-squares regression) capable of establishing an appropriate mathematical relation such as single absorbance versus concentration (univariate calibration) or spectra versus concentration (multivariate calibration).

The calibration methods most frequently used to relate the property to be measured to the analytical signals acquired in NIR spectroscopy are multiple linear regression (MLR),[27,28] principal component regression (PCR) and partial least-squares regression (PLSR).[29] Most of the earliest quantitative applications of NIR spectroscopy were based on MLR because spectra were then recorded on filter instruments, which afforded measurements at a relatively small number of discrete wavelengths only. However, applications involving PCR and PLSR have become more widely used since the availability of commercial instruments allowing the whole NIR region to be scanned. More information on the topic of chemometrics can be found in Chapter 12.

Essentially, the procedure for quantitation using multivariate calibration involves the following steps: (a) selecting a representative sample set; (b) acquiring the spectra; (c) obtaining the reference values; (d) mathematically processing the spectra; (e) selecting the model that relates the property to be determined and the signals; and (f) validating the model. Each step is described in detail below,[30] with special emphasis on the problems specific to NIR analysis of pharmaceuticals.

14.3.1 Selection and preparation of samples

The starting point for any calibration is collecting a set of representative samples encompassing all possible sources of variability for the manufacturing process and spanning a wide enough range of values of the target parameter. These requirements are easily met for liquid samples, but not quite so for solid samples (e.g. grain, tablet or pill forms). More information on the topic of sampling can be found in Chapter 3.

The NIR spectrum depends not only on the chemical composition of the sample, but also on some physical properties such as the size, form, particle distribution and degree of compression of the sample. This is useful in some cases as it enables the spectroscopic determination of some physical parameters of the sample. However, physical differences can lead to multiplicative effects in the spectrum, which, together with other additive effects such as baseline shift or chemical absorption, can complicate calibration models and detract from the quality of the results of quantitative analyses if not properly understood and accounted for.

These effects can be modeled by ensuring that the calibration set encompasses all possible sources of variability in the samples, which in turn can be ensured by using samples from an adequate number of production batches.[31]

Different production batches of the same pharmaceutical preparation can have very similar concentrations of API (±5% around the nominal value) and excipients close to the nominal value. It is difficult to obtain a set of production samples of the target product spanning the concentration range required to develop a calibration model that will allow the pharmacopoeial recommendations to be met and different enough values (±20% around the nominal one) of potency, content uniformity, etc., accommodated.

Incorporating all sources of variability in the production process with a known or presumed influence on the spectra for the samples entails expanding the calibration set with new samples obtained from the pilot plant or prepared in the laboratory. Preparing these additional samples can be more or less complicated depending on their physical state; because no unique methodology exists for this purpose, this section describes the most widely used and efficient choices.

The samples used to construct the models should be similar to the production samples; also, their spectra should be recorded in the same mode (reflectance, transmission or transflectance) as those for the samples to be subsequently predicted, and include all potential sources of variability. Although such sources are relatively limited – pharmaceutical samples usually have a well-defined qualitative and quantitative composition from raw material to end product, and production processes are solidly established and reproducible – their

effects on the sample spectra cannot always be accurately predicted. This may prevent assuring the product quality levels required to meet the stringent rules governing drug production processes.

Appropriate external samples can be of various types, namely:

- *Pilot plant samples (A).* A pilot plant mimicking the actual production plant is used to obtain a set of samples with physical properties similar to those of the production process and containing API at concentrations spanning the required range. Similar samples can be laboratory-made by following the same operating procedure (granulation, milling, pressing) as in the industrial production process (granulation, milling, compaction). This is usually rather labor-intensive and not always feasible. However, it is the only way of ensuring that all sources of production variability will be incorporated into the calibration set.
- *Overdosed and underdosed production samples (B).* Properly powdered production samples are supplied with known amounts of the API (overdosing) or a mixture of the excipients (underdosing) in order to obtain the desired range of API concentrations. The amounts of API or excipients added are usually small, so the ensuing changes in the physical properties of the samples are also small and their matrix can be assumed to be essentially identical with that of the production samples. This is an effective, less labor-intensive method than the previous one and allows the required API concentration range to be uniformly spanned.
- *Synthetic laboratory samples (C).* These are prepared by weighing known amounts of API and each excipient, and blending them to a homogeneous mixture. This procedure has the advantage that it is less labor-intensive and expensive than the previous two; also, the required sample concentrations can be easily obtained. In addition, it affords the use of experimental sample selection designs in order to reduce correlation between component concentrations and provides more robust models as a result. However, the synthetic samples may differ from actual production samples in some physical characteristics (particle size, polymorphs formed during the production process) as a result of the differences between the two preparation procedures and thus may lead to substantial errors in predicting production samples. This shortcoming, however, can be circumvented by including an adequate number of production samples in the calibration set. This procedure has proved effective in work conducted by the authors' group with a view to determining the active ingredient content of some pharmaceutical preparations.[32,33] In some cases, however, spectral differences between laboratory-made and production samples are so large that the ensuing models are rather complex and scarcely robust.

14.3.2 Preparation and selection of samples

The procedure of choice for each type of sample depends on its physical state and also on the particular purpose of the measurement. Thus, if the target chemometric models are to be applied to several steps of the process, then the simplest possible model (namely that using a single type of sample) can be expanded with new samples in order to facilitate prediction of subsequent production steps. This section discusses some major aspects of the preparation procedure for each type of sample.

1. *Solid samples'* preparation procedures vary with their specific form of presentation, namely:
   (a) *Powdered samples* are possibly best prepared by using method (situation C above), which involves weighing the different individual components. An experimental design can be used to reduce correlation between samples. The ensuing model can usually predict production samples with acceptable accuracy without the need to incorporate external samples.
   (b) *Granulated samples* can be prepared by using method (B) provided the amounts of API or excipients to be added are small enough to avoid substantially altering the preparation matrix while ensuring accurate prediction of process samples. Method (A) which involves granulating, milling

and sieving samples similarly as in the industrial process, can be effective for this purpose. It is somewhat more time-consuming but reduces correlation between components, which is highly desirable when the API spectrum is very similar to that of some excipient.

(c) *Tablets* are best prepared in intact form by using method (A), whether at the pilot plant or in the laboratory. This procedure is effective for both transmission and reflectance measurements of intact tablets.[34]

2. *Liquid samples* can be efficiently prepared with methods (B) and (C), which provide samples similar to their production counterparts that span the desired API concentration range. These samples are measured in the transmission mode.
3. *Semisolid samples*. As with liquid samples, methods (B) and (C) are the best choices for this type of sample. The specific choice will depend on the rheological properties (viscosity, density, air retention) of the particular preparation. These samples are best measured in the transflectance mode. Liquid and semisolid samples may contain a mixture of solvents of disparate volatility which may evaporate separately during the measurement process. Differences in solvent volatility can alter the sample matrix and lead to errors in the determination which are best avoided by using a set of calibration samples spanning an expanded range of solvent proportions.[35,36]

The body of samples selected is split into two subsets, namely the calibration set and the validation set. The former is used to construct the calibration model and the latter to assess its predictive capacity. A number of procedures for selecting the samples to be included in each subset have been reported. Most have been applied to situations of uncontrolled variability spanning much wider ranges than those typically encountered in the pharmaceutical field. One especially effective procedure is that involving the selection of as many samples as required to span the desired calibration range and encompassing the whole possible spectral variability (i.e. the contribution of physical properties). The choice can be made based on a plot of PCA scores obtained from all the samples.

Too small a calibration set may result in some source of variability in the product being excluded and hence cause spurious results in analyzing new samples. According to some authors, the optimum number of samples to be included depends on their complexity, the concentration range to be spanned and the particular calibration methods used.[37,38] Thus, when the aim is to quantify using one to four PLS factors and the samples exhibit relatively small differences in their physical and chemical properties, a calibration set comprising a minimum of 15–20 samples will be more than adequate.

14.3.3 Determination of reference values

In order to construct a calibration model, the values of the parameters to be determined must be obtained by using a reference method. The optimum choice of reference method will be that providing the highest possible accuracy and precision. The quality of the results obtained with a multivariate calibration model can never exceed that of the method used to obtain the reference values, so the choice should be carefully made as the quality of the model will affect every subsequent prediction. The averaging of random errors inherent in regression methods can help construct models with a higher precision than the reference method.

14.3.4 Acquisition of spectra

The spectra for the samples in the above-described set are typically recorded on the same equipment that will subsequently be used for the measurement of the unknown samples. Such spectra will include the variability in the instrument. The essential prerequisites for constructing the calibration model are that the spectrum should contain the information required to predict the property of interest and that the contribution

of such a property to the spectrum should be much greater than that of the noise it may contain. It is recommended to acquire the spectra for the calibration samples in a more careful manner than those for the other samples in order to minimize their noise (e.g. by strictly controlling the environmental conditions, averaging an increased number of scans, recording spectra on different days) on the assumption that a better model will result in better predictions – even if the latter spectra are noisier.

The spectra will contain information about the physical properties of the samples that do not contribute to determining ingredient concentrations. Models can be simplified by subjecting the spectra to effective mathematical pretreatments in order to cancel or reduce such contributions. The spectral pretreatments most frequently used for this purpose are smoothing derivatives, multiplicative scatter correction (MSC) and SNV, which can be implemented via algorithms included in the software bundled with many instruments. Choosing an appropriate pretreatment and the way it is applied are important as not all are equally effective in removing or reducing unwanted contributions to the spectrum. For example, spectral derivatives increase signal-to-noise ratio, so derivative spectra are frequently smoothed with a suitable algorithm such as that of Savitsky and Golay.[39]

14.3.5 Construction of the calibration model

Although many calibration models use the whole spectrum, using a selected range of wavelengths can facilitate the construction of more simple, accurate and hence robust models. The selection can be made after careful spectroscopic data visualization. The most straightforward and intuitive way of selection is probably that based on the range where the analyte exhibits its strongest response and hence its greatest contribution to the spectral signal for the sample. There are various algorithms that can be used to aid in identifying such regions.

The choice of the calibration method is dictated by the nature of the sample, the number of components to be simultaneously determined, the *a priori* knowledge about the system studied and available data on it. The most salient features of the different types of calibration methods are described below.

MLR is the usual choice with filter instruments and is also used with those that record whole spectra. It is an effective calibration approach when the analytical signal is linearly related to the concentration, spectral noise is low and the analyte does not interact with other sample components. In addition, the MLR technique affords modeling linear relations as it assumes modeling errors to arise from concentrations that follow Beer's law. However, MLR can only be used at a small number of wavelengths, which, if incorrectly selected, may result in overfitting (i.e. in the modeling of noise or random errors). Also, since spectral data are highly collinear, then the precision of the results will suffer appreciably with the use of multiple collinear wavelengths. A detailed description of available procedures for determining how many and which wavelengths should be used can be found elsewhere.[29,40,41]

Whole-spectrum methodologies (namely PCR and PLSR) have the advantage that they can use every single wavelength in a recorded spectrum with no prior selection. Also, they allow the simultaneous determination of several components in the same sample and avoid the problems arising from colinearity among spectral data and noise-related variability. One important decision regarding such calibration is the choice of the number of factors to be used in order to avoid overfitting (modeling noise) and the ensuing sizeable errors in the prediction of unknown samples.

Nonlinearity in NIR signals is ascribed to nonlinear detector responses that result in curved signal–concentration plots, as well as to physical and/or chemical factors giving rise to shifts and width changes in spectral bands.[42,43] In some cases, slight nonlinearity can be corrected by including an additional factor in the PLS model; strong nonlinearity, however, requires the use of a nonlinear calibration methodology such as neural networks,[44–48] locally weighted regression,[49,50] projection pursuit regression[51,52] or quadratic versions of the PCR and PLSR algorithms.[53,54]

14.3.6 Model validation

NIR models are validated in order to ensure quality in the analytical results obtained in applying the method developed to samples independent of those used in the calibration process. Although constructing the model involves the use of validation techniques that allow some basic characteristics of the model to be established, a set of samples not employed in the calibration process is required for prediction in order to confirm the goodness of the model. Such samples can be selected from the initial set, and should possess the same properties as those in the calibration set. The quality of the results is assessed in terms of parameters such as the relative standard error of prediction (RSEP) or the root mean square error of prediction (RMSEP).

14.3.7 Prediction of new samples

Once its goodness has been checked, the model can be used to predict production samples. This entails applying it to new samples that are assumed to be similar in nature to those employed to construct the model. This, however, is not always the case as uncontrolled changes in the process can alter the spectra of the samples and lead to spurious values for the target parameter. Instrument-control software typically includes statistical methods for identifying outliers and excluding them prior to quantitation; the use of sample identification and qualification techniques prior to the quantitation solves this problem. One other important consideration is that the model retains its quality characteristics when interpolating: that is when the values to be predicted fall within the concentration range spanned by the calibration curve. In fact, extrapolation is unacceptable in routine analyses.

14.4 Method Validation

Every analytical method should be validated in order to ensure that it is clearly and completely understood. This is accomplished by applying appropriate methodologies to determine properly selected parameters. The methodology usually employed to validate analytical methods has been designed for univariate techniques (e.g. HPLC, UV-vis spectrophotometry). Its adaptation to NIR spectroscopy is made difficult by the multivariate nature of calibration. The EMA,[55] the Pharmaceutical Analytical Sciences Group,[5] ICH[56] and Pharmacopoeias have issued guidelines for the validation of NIR methods. The quality parameters recommended for the determination of major compounds include specificity, linearity, range, accuracy, repeatability, intermediate precision and robustness. Other parameters such as the detection limit, quantitation limit and reproducibility are unnecessary. In a review, Moffat[57] described and critically compared the characteristics of different validation parameters. The special features of NIR warrant some comment. The greatest difference lies in specificity. As noted earlier, NIR spectrometry is unselective, so it does not necessarily allow the presence of a given analyte in a pharmaceutical preparation to be uniquely identified. The application of a quantitative method to a spectrum can yield spurious results if the samples examined are different from those used to construct the model. Probably the only reliable approach to the demonstration of specificity is the use of both identification and quantitation tests; a pharmaceutical preparation passing the identification test will be of the same type as those used for calibration and hence fit for application of the model. This makes a substantial difference to the methodology used to validate methods based on other techniques.

14.5 Calibration Transfer

Constructing a multivariate model is a time-consuming task that involves various operations including the preparation of a sizeable number of samples, recording their spectra and developing and validating the cali-

bration equation. Obtaining an appropriate calibration model therefore involves a substantial investment of time – and of money as a result. In the process, the usual variability in laboratory work (e.g. in the recording of spectra by different operators and/or on different days, in the samples to be analyzed) is incorporated into the model. The incorporation of variability sources makes the model more robust and avoids adverse effects of typical laboratory variations on the determination of the target parameter.

The model can lose its predictive capacity for reasons such as a dramatic change in the nature of the samples. A new calibration model better adjusted to the new situation must therefore be constructed. In other cases, however, the model can be re-adjusted (restandardized) in order to ensure a predictive capacity not significantly different from the original one.

The calibration model may exhibit a decrease in the ability to predict new samples due to changes in the instrumentation. The differences in instrumentation may introduce variability in the spectra. The potential sources of variability include the following:

1. Replacing the analytical equipment with one of similar, better or rather different characteristics. If both systems are available simultaneously for an adequate length of time, then spectra can be recorded on both in order to derive as much information as possible in order to understand the differences and make appropriate corrections to the data to account for them. The original equipment is known as the *master* and that replacing it as the *slave*.
2. Transferring a calibration to another spectrometer. Because the ability to use the master and slave simultaneously may not exist, the amount of useful spectral information may be inadequate.
3. Breakdown of the instrument or end of its service lifetime. Repairing or replacing an instrument may alter spectral responses if the instruments are not perfectly matched. Because this is an unexpected event, spectra recorded prior to the incident are rarely available.

One plausible solution involves recalibration with the new equipment. This drastic choice entails ignoring the whole work previously done for this purpose, which is only acceptable in extreme cases where no alternative solution exists.

A host of mathematical techniques for standardizing calibration models to facilitate their transfer is available. These generally focus on the coefficients of the model, the spectral response or the predicted values.[58]

Standardizing the coefficients of the model entails modifying the calibration equation. This procedure is applicable when the original equipment is replaced (situation 1 above). Forina *et al.*[59] developed a two-step calibration procedure by which a calibration model is constructed for the master (Y–X), its spectral response correlated with that of the slave (X–X') and, finally, a global model correlating variable Y with both X and X' is obtained. The process is optimized in terms of SEP and SEC for both instruments as it allows the number of PLS factors used to be changed. Smith *et al.*[60] propose a very simple procedure to match two different spectral responses.

Standardizing the spectral response is mathematically more complex than standardizing the calibration models but provides better results[61] as it allows slight spectral differences – the most common between very similar instruments – to be corrected via simple calculations. More marked differences can be accommodated with more complex and specific algorithms. This approach compares spectra recorded on different instruments, which are used to derive a mathematical equation, allowing their spectral response to be mutually correlated. The equation is then used to correct the new spectra recorded on the slave, which are thus made more similar to those obtained with the master. The simplest methods used in this context are of the univariate type, which correlate each wavelength in two spectra in a direct, simple manner. These methods, however, are only effective with very simple spectral differences. On the other hand, multivariate methods allow the construction of matrices correlating bodies of spectra recorded on different instruments for the above-described purpose. The most frequent choice in this context is piecewise direct standardization

(PDS),[62] which is based on the use of moving windows that allow corrections to be applied in specific spectral regions. Other methods[58] used for this purpose include simple wavelength standardization (SWS), multivariate wavelength standardization (MWS) and direct standardization (DS).

Standardizing the predicted values is a simple, useful choice that ensures smooth calibration transfer in situations (a) and (b) above. The procedure involves predicting samples for which spectra have been recorded on the slave using the calibration model constructed for the master. The predicted values, which may be subject to gross errors, are usually highly correlated with the reference values. The ensuing mathematical relation, which is almost always linear, is used to correct the values subsequently obtained with the slave.

Alternative mathematical methods such as artificial neural networks (ANN), maximum likelihood PCA and positive matrix factorization have also proved effective for calibration transfer, but are much more complex than the previous ones and are beyond the scope of this chapter. For more information about this topic see Chapter 12.

14.6 Pharmaceutical Applications

14.6.1 Identification of raw materials

The current widespread use of NIR spectroscopy for identifying pharmaceutical raw materials has been promoted by its ability to identify each individual product in a simple, unequivocal manner. Identification is a requirement for cGMP and also to ensure quality in the end product (e.g. in the US FDA's PAT Initiative). Blanco *et al.*[63] demonstrated this concept by developing and validating a self-contained spectral library affording the identification of specific raw materials and their characterization based on differences in physical properties such as polymorphism, particle size or chirality. The library initially consisted of the spectra for 125 products that were distinguished in terms of the correlation coefficients for second-derivative spectra between the unknown sample and the centroid of the body of spectra constituting each class. If the coefficient was greater than 0.97, the identification was deemed positive. The use of cascading sublibraries allowed spectral ambiguity, which occurs when one or more spectra possess a correlation coefficient exceeding a preset threshold, to be avoided. Pattern recognition methods, which afford improved spectral discrimination, use PCA to identify highly similar spectra for materials differing in some physical property (particle size, polymorphs present). The original library has been expanded to 300 products and continues to be routinely used to identify all input raw materials. Table 14.1[63–73] lists selected references on the use of NIR for identification of pharmaceutical raw materials.

14.6.2 Homogeneity

Ensuring homogeneous mixing of the components (APIs and excipients) of a pharmaceutical preparation is a crucial prerequisite for obtaining proper solid dosages (tablets and capsules). Pharmaceutical manufacturers invest much time, labor and material resources in this process. Blending is intended to ensure uniform distribution of all components in the end products,[74] so that each dose will contain the correct amount of the active ingredient.

As a rule, controlling the blending process entails stopping the blender, collecting samples at preset points within it – usually by hand or by using a sample thief – and analysing them by high performance liquid chromatography (HPLC) or UV-vis spectroscopy.[75] NIR spectroscopy enables the analysis of complex matrices without the need to manipulate samples. This results in substantially decreased analysis time relative to wet chemical methods. The large number of NIR applications, qualitative and quantitative, reported in recent years testify to its high efficiency. Many such uses have focused on the determination of APIs in

Table 14.1 *Applications of NIR spectroscopy to API and excipient identification*

Analyte	Pattern recognition method	Remarks	References
Raw materials	Mahalanobis distance and residual variance	Classification models development	24
Raw materials	Correlation coefficient, Mahalanobis distance, residual variance	Libraries and sublibraries	63
Oxytetracycline	PCA and SIMCA	Sample identity determination before multivariate quantitation	64
Pharmaceuticals	PCA and visual classification	Drug authenticity assurance, counterfeiting control	65
Excipients	PCA and cluster analysis	Transferability between different spectrometers	66
Metronidazole	Hierarchical and stepwise classification, stepwise discriminant analysis and PCA	Assure the quality of the drug	67
Excipients	Polar qualification system	Sample identification and qualification	68
Corn starch	Correlation coefficient and distance	Spectral library as quality control system	69
Herbal medicines	Wavelength distance, residual variance, Mahalanobis distance and SIMCA	PRMs comparison	70
Acetaminophen	Maximum distance in wavelength space	Identification of tablets brand	71
Excipients	SIMCA	Classification models development	72
Pharmaceuticals	Correlation coefficient, wavelength distance, polar coordinates, gravity center, scanning probability windows	PRMs comparison	73

pharmaceutical preparations[76,77] or physical parameters of samples.[78,79] Others have enabled the identification and characterization of raw materials[80,81] or even the monitoring of product degradation.[82] There are, however, comparatively few references to the validation of blending processes,[83–85] all of which have involved a search for objective criteria to determine how long blending should be continued. In fact, the visual inspection of spectra recorded during the blending process usually helps, but may not allow one to precisely identify the time when the mixture becomes uniform. The need to define these objective criteria continues to be reflected in current studies in this context.

Some of the procedures used to assess mixture uniformity (e.g. the visual inspection of spectra) are subjective. Those based on mathematical parameters are objective. However, most require spectral information from homogeneous samples, so criteria can be established only after the process is finished.

The qualification concept was used by Wargo and Drennen[84] to verify the homogeneity of solid mixtures. Qualitative analytical algorithms based on the Bootstrap Error-adjusted Single-sample Technique (BEST) proved sensitive to variations in sample homogeneity. Plugge and van der Vlies converted NIR spectra into polar coordinates and used them to calculate the Mahalanobis distance[86] in order to identify nonhomogeneous samples.

Hailey *et al.*[87] and Sekulic *et al.*[88] developed systems for monitoring the homogenization of solid mixtures based on measurements made with a fiber-optic probe fitted to the mixer. The most salient advantage of these systems is that they allow one to identify the end point of the homogenization process noninvasively in real time. In both cases, mixture homogeneity was determined by plotting the standard deviation for several replicates against the homogenization time. Cuesta[85] conducted a similar study using samples withdrawn from the blender. Sekulic[89] explored various parameters of the blending process and applied various spectral pretreatments including detrending, SNV and the second spectral derivative in conjunction with various criteria (running block standard deviation, dissimilarity and PCA) to identify the end point of the process.

Blanco[90] proposed the use of the mean square difference between two consecutive spectra plotted against the blending time in order to identify the time that mixture homogeneity was reached.

One chemometric method used to monitor mixing involves comparing the spectrum for the unknown sample with that for one assumed to be homogeneous via the so-called 'conformity index',[91] which is calculated by projecting the spectrum for the unknown sample onto the wavelength space of the spectrum or mean of spectra for the 'homogeneous' sample. This procedure is similar to that involving the calculation of distances in a principal component space.

14.6.3 Moisture

Water is one of the most strongly absorbing compounds in the NIR spectral region. This makes it an unsuitable solvent and considerably restricts the determination of species in water. Due to its strong absorbance in the NIR region, water can be determined as an analyte at very low concentrations. Direct inspection of a NIR spectrum allows one to distinguish differences in moisture content among batches of the same product. The presence of crystallization or adsorbed water in pure substances and pharmaceutical preparations, whether during treatment or storage of the sample, causes significant changes in those properties that influence chemical decay rates, crystal dimensions, solubility and compaction power, among others. NIR spectroscopy is an effective alternative to traditional analytical methods such as the Karl Fischer (KF) titration for quantifying moisture. The NIR spectrum for water exhibits five absorption maxima at 760, 970, 1190, 1450 and 1940 nm. The positions of these bands can be slightly shifted by temperature changes[92–94] or hydrogen bonding between the analyte and its matrix.[95,96] The bands at 760, 970 and 1450 nm correspond to the first three overtones of O–H stretching vibrations, whereas the other two are due to combinations of O–H oscillations and stretching. The specific band to be used to determine water will depend on the desired sensitivity and selectivity levels.[97] As a rule, the overtone bands are appropriate for this purpose when using solutions in solvents that do not contain O–H groups; on the other hand, the band at 1940 nm provides increased sensitivity.

The earliest applications for quantitative analysis of liquid samples[98] and solid preparations entailed sample dissolution in an appropriate solvent. A number of moisture determinations in APIs and pharmaceutical preparations based on both reflectance and transmission measurements have been reported.[99–103] Their results are comparable to those of the KF method. The high sensitivity provided by the NIR technique has fostered its use in the determination of moisture in freeze-dried pharmaceuticals.[104–109] The noninvasive nature of NIR has been exploited in determination of moisture in sealed glass vials.[110–112]

Borer *et al.*[113] developed a NIR method to determine moisture in a highly hygroscopic product. Because the sample absorbed water very rapidly (0.02%/min), use of the Karl Fischer method as reference was very complicated. This led them to the choice of NIR spectroscopy, where the sample can be kept inside a sealed vial instead. Analyses were completed within a few seconds and the samples not exposed to ambient air at any time, which prevented moisture absorption and the ensuing determination errors.

14.6.4 Determination of physical parameters

The NIR spectrum contains information about the physical characteristics of samples, as well as chemical information, which are the basis of the previously discussed applications. NIR is thus suitable for determining physical parameters.

14.6.4.1 Particle Size

Particle size determinations are of paramount significance to the pharmaceutical industry and incorrect grain-size analyses can lead to altered properties such as coating power, hinder subsequent mixing of powders (for tablet formulations) or powder and liquids (suspensions), and result in defective pressing of solid mixtures in manufacturing tablets. Because particle size is one of the physical parameters most strongly influencing NIR spectra, it is an effective alternative to the traditional methods involving sieving, light scattering by suspensions, gas adsorption on solid surfaces or direct inspection under a microscope.

Ciurczak *et al.*[114] used the linear relationship of band intensity at a constant concentration on the average particle size at a preset wavelength to determine the average particle size of pure substances and granules. They found plots of absorbance versus particle size at variable wavelengths to consist of two linear segments, and the effect of particle size on reflectance measurements to be significantly reduced below 80 μm.

Ilari *et al.*[115] explored the feasibility of improving the determination of the average particle size of two highly reflecting inorganic compounds (crystalline and amorphous NaCl) and an NIR-absorbing species (amorphous sorbitol), using the intercept and slope obtained by subjecting spectra to MSC treatment as input parameters for PLSR.

Blanco *et al.*[116] determined the average particle size of Piracetam over the average size range of 175–325 μm with an error of ± 15 μm. NIR was used for quantification of particle size with the aid of MLR and PLSR calibrations.

Frake *et al.*[117] compared various chemometric approaches to the determination of the median particle size in lactose monohydrate with calibration models constructed by MLR, PLS, PCR or ANNs. Overall, the ensuing models allowed mean particle sizes over the range 20–110 μm to be determined with an error less than 5 μm, which is comparable to that of the laser light diffraction method used as reference. Predictive ability was similar for models based on absorbance and second-derivative spectra; this confirms that spectral treatments do not suppress the scattering component arising from differences in particle size.

O'Neil *et al.*[118] demonstrated the ability of NIR to determine median particle size (d_{50}) in various types of materials. By using MLR calibration models, the authors succeeded in determining d_{50} over two different ranges (19.2–183 and 240–406 μm) and found models based on ln d_{50} to provide better results than those based on d_{50}. In another study the same authors[119] studied cumulative particle size distribution in 11 different quantiles (5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 95%). They used the forward angle laser light scattering (FALLS) method as reference and MLR and PCR – of which the latter provided a more robust method – for calibration. The procedure followed to obtain the whole cumulative particle size distribution involved constructing a calibration model for each level to be assessed, which made the chemometric development step rather laborious. However, the ability to obtain the whole curve directly from the NIR spectrum within a few seconds offset this disadvantage. The same authors[119] also determined particle size distribution curves by using PLS2 multivariate calibration models, which allow the simultaneous modeling and determination of several variables. Similarly, Blanco *et al.*[120] compared the use of PLS and ANN for calibration in order to obtain particle size distribution curves for inorganic aggregates. While the models were not applied to pharmaceutical samples, their methodology exhibits the potential of NIR spectroscopy for the rapid, accurate, precise determination of particle size distribution.

14.6.4.2 Compaction Pressure and Tablet Hardness

The compaction pressure used in the tableting press can strongly influence some properties of the resulting tablets (e.g. their dissolution rate). The technical ability of NIR spectroscopy to extract both physical and chemical information from spectra for intact samples affords monitoring of the tablet compaction pressure. Blanco and Alcalá[34] developed a NIRS method to determine compaction pressure and drug content per dosage unit; increasing the tablet compaction pressure was found to increase the spectral baseline similarly to an increase in particle size. The prediction errors of their PLS models based on laboratory samples compacted at pressures over the range 150–800 MPa were all less than 11%.

Determining tablet hardness usually involves the destructive diametral crushing of each individual tablet. However, Kirsch and Drennen[121] developed a nondestructive method based on the slope of NIR spectra for intact tablets. Tablet hardness ranged from 1 to 7 kp and was predicted with errors of 0.2–0.7 kp by their calibration models. Their approach, the robustness of which was confirmed by the results, was claimed to surpass PCR calibration models in many respects as a result of its being unaffected by the absorption of individual bands.

14.6.4.3 Dissolution Testing

The outcome of dissolution tests on pharmaceutical formulations can be influenced by a number of factors including the nature of the excipients and lubricants, blending method and granulation procedure used. When the solubility of the API is not the limiting factor, the compaction pressure used to produce tablets is inversely proportional to their dissolution rate and, as noted in the previous section, directly proportional to the slope of the NIR spectra for intact tablets. Blanco *et al.*[122–124] have developed various approaches for the determination of typical dissolution testing parameters requiring no other input than the NIR spectra for intact tablets. Briefly, the procedure involves preparing tablets spanning a wide range of compaction pressures (74–887 MPa), recording the NIR spectra for intact tablets and determining each typical parameter ($yield_{30min}$, $lag_{80\%}$ and the complete profile) with the usual methods. These authors proposed determining various parameters based on NIR spectra in order to confirm compliance with specification. To this end, they used (a) a qualitative method based on discriminant PLS methodology (b) a quantitative method using PLS1to obtain $yield_{30min}$ and (c) a quantitative method based on PLS2 to obtain the complete profile and $lag_{80\%}$. The ensuing methods afford dissolution testing in a simple, clean manner; they are also advantageous as they use no solvents and produce no waste since all they need is the NIR spectrum for intact tablets.

NIR spectra are sensitive to some of the variables of the tablet coating process. Changes in pharmaceutical formulations during coating, coating quality and lacquering time have been related to the dissolution rate of the resulting tablets. These properties can be determined simply by using NIR spectra for intact tablets in combination with appropriate qualitative and quantitative chemometric methods.[125] While particle size continues to be the physical property most frequently determined by NIR spectroscopy, several other parameters including the dissolution rate and the thickness and hardness of the ethylcellulose coating on theophylline tablets have also been determined using NIR, all with good errors of prediction.[126]

14.6.4.4 Characterization of Polymorphs

The significance of polymorphism in the pharmaceutical industry lies in the fact that polymorphs can exhibit differential solubility, dissolution rate, chemical reactivity, melting point, chemical stability or bioavailability, among others. Such differences can have considerable impact on a drug's effectiveness. Usually, only one polymorph is stable at a given temperature, the others being metastable and evolving to the stable phase

with time. Regulatory authorities have established the need to control the polymorph selected for pharmaceutical use in each case and to ensure its integrity during formulation and storage. Polymorphs are usually identified, characterized and determined using X-ray diffraction (the reference method), IR, solid-state NMR or Raman spectroscopy. NIR spectroscopy[127,128] is an effective alternative as polymorphs usually differ in their NIR spectra because the vibrations of their atoms are influenced by the differential ordering in their crystal lattices (especially by the presence of hydrogen bonds). NIR spectroscopy enables the fast, reproducible recording of spectra without the need to pretreat or alter the sample in any way, thereby preventing polymorph interconversion. This makes NIR spectroscopy especially attractive for this purpose. NIR can be used to rapidly identify the different polymorphs. Applicable references include the discrimination of polymorphs by use of pattern recognition methods,[129] the on-line monitoring of polymorph conversions in reactors,[130] the monitoring of such conversions in active drug ingredients using MLR[131] and the qualitative monitoring of polymorph conversion as a function of ambient moisture.[57] Polymorphs can be characterized in an expeditious, convenient manner by using the above-described spectral libraries; the spectral differences involved are almost always large enough to use the correlation coefficient as the discriminating criterion. By using sublibraries one can define classes to the level of distinguishing pure polymorphs from contaminated forms containing relatively small amounts of other polymorphs.[35]

Some chiral compounds exhibit a different crystal structure for each pure enantiomer and their mixture; the difference can be used to distinguish them. Again, the use of cascading libraries allows pure compounds to be distinguished from compounds contaminated with relatively small amounts of the other chiral forms.

The presence of an unwanted polymorph, even in small proportions, in an API can significantly affect its response by inducing its polymorphic transformation. A number of API polymorphs have been determined using NIR[132–134] or Raman spectroscopy.[135] Blanco *et al.* applied this technique to the determination of the crystalline forms in amorphous antibiotics[136,137] and in tablets.[133]

Some authors have examined polymorphic transformations resulting from hydration of the product[102] or the passage of time.[87,130]

The presence of some excipients in a pharmaceutical formulation can induce a polymorphic transformation in the API after wet granulation. Blanco *et al.*[138] have developed a method based on multivariate curve resolution (MCR) that exposed a polymorphic change in the API during wet granulation of a formulation containing microcrystalline cellulose as the major excipient. By contrast, a formulation containing lactose as the major excipient exhibited no similar changes. Analyses, which involved recording the NIR spectra for intact tablets, were quite expeditious. Also, the chemometric algorithm used allowed the spectrum for the pure API to be obtained without the need to isolate it from the formulation.

14.6.5 Determination of chemical composition

The strict regulations of the pharmaceutical industry have a significant effect on the quality control of final products, demanding the use of reliable and fast analytical methods. The capacity that the technique has for the simultaneous determination of several APIs with no need of, or with minimum, sample preparation has considerably increased its application in pharmaceutical analytical control. The main limitation of NIR is the relatively low sensitivity that limits the determination of APIs in preparations when their concentration is less than 0.1%. Nevertheless, instrumental improvements allow the determination below this limit depending on the nature of the analyte and the matrix, with comparable errors to the ones obtained with other instrumental techniques. The reference list presents an ample variety of analytical methodologies, types of samples, nature of analyte and calibration models. A detailed treatment of each one is beyond the scope of this chapter. Many applications have been gathered in recent reviews.[125,139–144] Table 14.2[32,35,145–150] summarizes some examples where NIR spectroscopy is used in this context.

Table 14.2 *Applications of NIR spectroscopy in pharmaceutical determination*

Analyte	Sample type	Remarks	References
Paracetamol	Tablet	Drug determination of grinded tablets in reflectance mode. Standard error of calibration (0.48% m/m) and prediction (0.71% m/m)	32
Dexketoprofen	Gel	Non-destructive drug quantitation method of a hydrogel in transflectance mode	35
Acetylsalicylic acid	Tablet	Non-destructive quantitation method in three different formulations. Reflectance and transmittance modes compared	146
Otilonium bromide	Powder and tablet	Non-destructive quantitation method for QC of three production steps (blend, cores and coated tablets). Validated method following ICH guidelines. Relative errors lower than 1%	147
Acetylsalicylic acid	Tablet	Drug determination in three different formulations. Intact analysis in transmittance mode and sample grinding in reflectance mode.	148
Analgin	Powder	Non-destructive drug determination by means of artificial neural network. Different calibration strategies compared. Relative predictive error lower than 2.5%	149
Miokamycin	Powder	Drug determination in three different preparations. Validated method following ICH guidelines. Relative errors lower than 1.5%	150
Resorcinol	Water solution	Non-destructive quantitation method. Precision lower than 0.15% in a temperature range of 9–35 °C. Method validated using ICH-adapted guidelines	151

14.6.5.1 Tablets

Tablets account for more than 80% of all pharmaceutical formulations; therefore, the development and implementation of NIR methods for determining APIs in intact tablets is of a high interest with a view to assuring content uniformity and quality in the end product. Blanco *et al.*[122] developed an innovative strategy to prepare calibration samples for NIR analysis by using laboratory-made samples obtained by mixing the API and excipients in appropriate proportions and compacting the mixture at a pressure similar to that used industrially. This way of matching laboratory and production samples affords more simple and robust NIR methods which require the use of neither HPLC nor UV-vis spectroscopy as reference; rather, reference values are obtained by weighing during preparation of the samples. The PLS calibration models thus constructed exhibited a good predictive ability with various production batches.

One other important point in using NIR spectroscopy in pharmaceutical analyses is its limited sensitivity, which complicates the determination of analytes at low concentrations by effect of too small analyte signals

and high correlation with the excipients. However, using an appropriate experimental design to prepare the calibration set can help construct effective models to quantify drugs at very low concentrations (less than 0.02% w/w) in intact tablets from spectra recorded in the reflectance or transmission mode. Alcalá *et al.*[151] reported a procedure for selecting concentration ranges in order to construct PLS calibration models based on multivariate detection and quantitation limits. Such a simple approach allows the most suitable concentration range for determining analytes at very low levels to be easily identified. Based on the results, a PLS model spanning the 0–0.02% w/w range provided a limit of quantitation of 0.005% w/w and a limit of detection of 0.002% w/w.

14.6.5.2 Powders and Granulates

Wet granulation is often used to increase the fluency of solid mixtures via an increased particle size. Although the process does not alter the chemical composition of the mixture, it considerably affects some physical properties which are reflected in a NIR spectrum. In developing NIR-based quality control analytical methods for such mixtures, one should therefore consider the potentially influential physical properties of the mixture via appropriate multivariate calibration models. Blanco *et al.*[152] used a straightforward approach to preparing calibration sets encompassing both chemical variability in the preparation and sources of physical changes caused by granulation. The calibration set consisted solely of laboratory samples that were prepared in accordance with an experimental design intended to avoid cross-correlation between the excipient and API concentrations, and granulated by hand to a similar extent as in the industrial process. The ensuing PLS calibration models exhibited very good predictive ability with both laboratory and production samples. This methodology requires no reference (HPLC, UV-vis) values to construct the models, so it can be deemed an absolute analytical method.

14.6.5.3 Liquids and Semiliquids

Water absorbs strongly in the NIR spectral region. This makes it an excellent target analyte for NIR determination. Its high concentration when used as a solvent, however, poses serious problems including saturation of detector by its strong signals and overlap of their broad absorption bands. Such is the case with the determinations of APIs in liquid or semiliquid formulations, which are frequently prepared in a water/alcohol matrix. Blanco *et al.*[36] showed the joint use of NIR spectroscopy and an experimental design for the calibration set to allow both the API and two preservatives present at very low concentrations in a formulation to be accurately determined. Partial evaporation of the solvents during the measurement process resulted in substantial errors due to exposure of the sample to the air; however, an effective correction was introduced by using samples spanning an expanded concentration range for volatile compounds. The method was validated for analytical control of the preparation and provided a prediction error of $0.5\,mg\,g^{-1}$ for the drug (nominal content $25\,mg\,g^{-1}$) and $0.035\,mg\,g^{-1}$ for the less concentrated preservative (nominal content $0.2\,mg\,g^{-1}$).

14.7 Conclusions

NIR spectroscopy is probably the most successful technique for the development of qualitative and quantitative methods in the pharmaceutical industry. NIR spectra contain both chemical and physical information from samples (solid and liquid). Spectra can be acquired off-line in three different modes: transmittance, reflectance and transflectance. In all cases, the spectra are obtained in a few seconds without or minimum sample pretreatment. Multivariate data analysis techniques are usually needed for the development of the

analytical methods, because of the properties of NIR spectra (low-intensity, wide and overlapped bands). Multivariate chemometric techniques extract the relevant information from the spectra and allow the development of robust multivariate calibration models and pattern recognition methods. After finishing the calibration and validation processes, NIR methods can be implemented as routine analytical techniques for quality monitoring purposes in the pharmaceutical industry.

References

1. M. Blanco and I. Villarroya, NIR spectroscopy: a rapid-response analytical tool, *Trends Anal. Chem.*, **21**(4), 240–250 (2002).
2. A. Springsteen and T. Ricker, Standards for the measurement of diffuse reflectance. Part I, *NIR News* **7**, 6 (1996).
3. P.Y. Barnes, E.A. Early and A.C. Parr, NIST measurement services: Spectral reflectance. *NIST Special Publication*, SP250–48, 1988.
4. *European Pharmacopoeia*, 4th Edition (2002) Section 2.2.40: Near infrared spectrophotometry.
5. N. Broad, P. Graham, P. Hailey, *et al.*, Guidelines for the development and validation of near-infrared spectroscopic methods in the pharmaceutical industry. In *Handbook of Vibrational Spectroscopy*, vol. **5**, J.M. Chalmers and P.R. Griffiths (eds), John Wiley & Sons Ltd, Chichester, 3590–3610, 2002.
6. S.J. Choquette, J.C. Travis, C. Zhu and D.L. Duewer, wavenumber standards for near-infrared spectrometry. In *Handbook of Vibrational Spectroscopy*, vol. **1**, J.M. Chalmers, P.R. Griffiths (eds.), John Wiley & Sons Ltd, Chichester, 899–905, 2002.
7. J.S. Shenk and M.O. Westerhaus, New standardization and calibration procedures for NIRS analytical systems, *Crop Sci.* **31**, 1694–1696 (1991).
8. J.M. Olinger, P.R. Griffiths and T. Burger, Theory of diffuse reflectance in the NIR region, In *Handbook of Near-Infrared Analysis*, 2nd edition, D. Burns and E.W. Ciurczak (eds), Marcel Dekker, New York, 19–52, 2001.
9. M. Blanco, J. Coello, H. Iturriaga, S. Maspoch, C. Pezuela and E. Russo, Control analysis of a pharmaceutical preparation by near-infrared reflectance spectroscopy: a comparative study of a spinning module and fibre optic probe, *Anal. Chim. Acta*, **298**, 183–191 (1994).
10. H. Mark, Qualitative discriminant analysis. In *Handbook of Near-Infrared Analysis*, 2nd edition, D. Burns and E.W. Ciurczak (eds), Marcel Dekker, New York, 351–362, 2001.
11. N. Bratchell, Cluster analysis, *Chemom. Intell. Lab. Syst.*, **6**, 105–125 (1989).
12. O. Strouf, *Chemical Pattern Recognition*, Research Studies Press, Letchworth, 1986.
13. P. Cáceres-Alonso and A. García-Tejedor, Non-supervised neural categorisation of near infrared spectra. Application to pure compounds, *J. Near Infrared Spectrosc.*, **3**, 97–110 (1995).
14. M.P. Derde and D.L. Massart, Supervised pattern recognition: The ideal method? *Anal. Chim. Acta*, **191**, 1–16 (1986).
15. P.A. Lachenbruch, *Discriminant Analysis*, Hafner Press, New York, 1975.
16. D. Coomans and D.L. Massart, Alternative K-nearest neighbour rules in supervised pattern recognition. Part 1: K-nearest neighbour classification by using alternative voting rules, *Anal. Chim. Acta*, **136**, 15–27 (1982).
17. D. Coomans and D.L. Massart, Alternative K-nearest neighbour rules in supervised pattern recognition. Part 2: Probabilistic classification on the basis of the kNN method modified for direct density estimation, *Anal. Chim. Acta*, **138**, 153–165 (1982).
18. D. Coomans and I. Broeckaert, *Potential Pattern Recognition*, John Wiley & Sons, New York, 1986.
19. D. Coomans, M.P. Derde, D.L. Massart and I. Broeckaert, Potential methods in pattern recognition. Part 3: Feature selection with ALLOC, *Anal. Chim. Acta*, **133**, 241–250 (1981).
20. I. Jurickskay and G.E. Veress, PRIMA: A new pattern recognition method, *Anal. Chim. Acta*, **171**, 61–76 (1981).
21. M.P. Derde and D.L. Massart, UNEQ: A disjoint modeling technique for pattern recognition based on normal distribution, *Anal. Chim. Acta*, **184**, 33–51 (1986).
22. VISION User Manual, Foss NIRSystems, Silver Spring, MD, USA, 1998.
23. P.J. Gemperline and L.D. Webber, Raw materials testing using soft independent modeling of class analogy analysis of near infrared reflectance spectra, *Anal. Chem.*, **61**, 138–144 (1989).
24. E. Bertran, M. Blanco, J. Coello, H. Iturriaga, S. Maspoch and I. Montoliu, Near infrared spectrometry and pattern recognition as screening methods for the authentication of virgin olive oils of very close geographical origins, *J. Near Infrared Spectrosc.*, **8**, 45–52 (2000).

25. N.K. Shah and P.J. Gemperline, Combination of the Mahalanobis distance and residual variance pattern recognition techniques for classification of near-infrared reflectance spectra, *Anal. Chem.*, **62**, 465–470 (1990).
26. J.J. Workman Jr, P.R. Mobley, B.R. Kowalski and R. Bro, Review of chemometrics applied to spectroscopy: 1985–1995, Part I, *Appl. Spectrosc. Rev.*, **31**, 73–124 (1996).
27. E.V. Thomas and D.M. Haaland, Comparison of multivariate calibration methods for quantitative spectral analysis, *Anal. Chem.*, **62**, 1091–1099 (1990).
28. N. Draper and H. Smith, *Applied Regression Analysis*, 2nd edition, John Wiley & Sons, *New* York, 1981.
29. H. Martens and T. Næs, *Multivariate Calibration*, John Wiley & Sons, New York, 1991.
30. ASTM E1655-05. Standard practices for infrared, multivariate, quantitative analysis, In: Annual Book of Standards, vol. 3.06, ASTM International, West Conshohocken, PA, 2005.
31. M. Blanco, J. Coello, H. Iturriaga, S. Maspoch and C. Pezuela, Strategies for constructing the calibration set in the determination of active principles in pharmaceuticals by near infrared diffuse reflectance spectrometry, *Analyst*, **122**, 761–765 (1997).
32. D. Trafford, R.D. Jee, A.C. Moffat and P. Graham, A rapid quantitative assay of intact paracetamol tablets by reflectance near-infrared spectroscopy, *Analyst*, **124**, 163–167 (1999).
33. P. Corti, G. Ceramelli, E. Dreassi and S. Mattii, Near infrared transmittance analysis for the assay of solid pharmaceutical dosage forms, *Analyst*, **124**, 755–758 (1999).
34. M. Blanco, M. Alcalá, Content uniformity and tablet hardness testing of intact pharmaceutical tablets by near infrared spectroscopy. A contribution to process analytical technologies, *Anal. Chim. Acta*, **557**, 353–359 (2006).
35. M. Blanco and M.A. Romero, Near infrared transflectance spectroscopy. Determination of dexketoprofen in a hydrogel, *J. Pharm. Biomed. Anal.*, **30**, 467–472 (2002).
36. M. Blanco, M. Alcalá, M. Bautista, Pharmaceutical gel analysis by NIR spectroscopy determination of the active principle and low concentration of preservatives, *Eur. J. Pharm. Sci.*, **33**(4–5), 409–414 (2008).
37. P. Corti, E. Dreassi, C. Murratzu, G. Corbini, L. Ballerini and S. Gravina, Application of NIRS to the quality control of pharmaceuticals. Ketoprofen assay in different pharmaceutical formulae, *Pharmaceutica Acta Helvetiae*, **64**, 140–145 (1989).
38. B. Buchanan and D. Honigs, Trends in near-infrared analysis, *Trends Anal. Chem.*, **5**, 154–157 (1986).
39. A. Savitsky and M.J.E. Golay, Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.*, **36**, 1627–1639 (1964).
40. D.E. Honigs, G.M. Hieftje and T. Hirschfeld, A new method for obtaining individual component spectra from those of complex mixtures, *Appl. Spectrosc.* **38**, 317–322 (1984).
41. I.A. Cowe, J.W. McNicol and D.C. Cuthbertson, A designed experiment for the examination of techniques used in the analysis of near-infrared spectra. Part 2: Derivation and testing of regression models, *Analyst* **110**, 1233–1240 (1985).
42. T. Næs and T. Isaksson, Non-linearity problems in NIR spectroscopy, *NIR News* **4**(3), 14 (1993).
43. C.E. Miller, Sources of non-linearity in near infrared methods, *NIR News* **4**(6), 3 (1993).
44. R.L. Long, V.G. Gregoriov and P.J. Gemperline, Spectroscopic calibration and quantitation using artificial neural networks, *Anal. Chem.*, **62**, 1791–1797 (1990).
45. P.J. Gemperline, J.R. Long and V.G. Gregoriov, Nonlinear multivariate calibration using principal components regression and artificial neural networks, *Anal Chem.*, **63**, 2313–2323 (1991).
46. T. Næs, K. Kvaal, T. Isaksson and C. Miller, Artificial neural networks in multivariate calibration, *J. Near Infrared Spectrosc.*, **1**, 1–11 (1993).
47. B. Walczak and W. Wegscheimer, Non-linear modelling of chemical data by combinations of linear and neural networks methods, *Anal. Chim. Acta.*, **283**, 508–517 (1993).
48. Y. Liu, B.R. Upadhyaya and M. Naghedolfeizi, Chemometric data analysis using artificial neural networks, *Appl. Spectrosc.*, **47**, 12–23 (1993).
49. T. Næs, T. Isaksson and B.R. Kowalski, Locally weighted regression and scatter correction for near-infrared reflectance data, *Anal. Chem.*, **62**, 664–673 (1990).
50. T. Isaksson, C.E. Miller and T. Næs, Nondestructive NIR and NIT determination of protein, fat and water in plastic-wrapped, homogenized meat, *Appl. Spectrosc.*, **46**, 1685–1694 (1992).
51. J.H. Friedman and W. Stuetzle, Projection pursuit regression, *J. Am. Stat. Assoc.*, **76**, 817–823 (1981).
52. K.R. Beebe and B.R. Kowalski, Nonlinear calibration using projection pursuit regression: Application to an array of ion-selective electrodes, *Anal. Chem.*, **60**, 2273–2278 (1988).
53. N.B Vogt, Polynomial principal component regression: an approach to analysis and interpretation of complex mixture relationships in multivariate environmental data, *Chemom. Intell. Lab. Syst.*, **7**, 119–130 (1989).
54. S. Wold, N.K. Wold and B. Skagerber, Nonlinear PLS modelling, *Chemom. Intell. Lab. Syst.*, **7**, 53–65 (1989).

55. The European Agency for the Evaluation of Medicinal Products, 2003. Note for guidance. Available http://europa.eu/agencies/community_agencies/emea/index_en.htm accessed 30 November 2009.
56. ICH Guidelines: Q2A Text on Validation of Analytical Procedures 1994, Q2B Validation of Analytical Procedures: Methodology 1996. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use.
57. A.C. Moffat, A.D. Trafford, R.D. Jee and P. Graham, Meeting of the International Conference on Harmonisation's guidelines on validation of analytical procedures: Quantification as exemplified by a near-infrared reflectance assay of paracetamol in intact tablets, *Analyst*, **125**, 1341–1351 (2000).
58. R.N. Feudale, H.T. Woody, A.J. Myles, S.D. Brown and J. Ferre, Transfer of multivariate calibration models: A review, *Chem. Intell. Lab. Syst.*, **64**, 181–192 (2002).
59. M. Forina, G. Drava, C. Armanino, *et al.*, Transfer of calibration function in near-infrared spectroscopy, *Chem. Intell. Lab. Syst.*, **27**, 189–203 (1995).
60. M.R. Smith, R.D. Jee and A.C. Moffat, The transfer between instruments of a reflectance near-infrared assay for paracetamol in intact tablets, *Analyst*, **127**, 1682–1692 (2002).
61. Y. Wang, D.J. Veltkamp and B.R. Kowalski, Multivariate instrument standardization, *Anal. Chem.*, **63**, 2750–2756 (1991).
62. E. Bouveresse, C. Hartmann, D.L. Massart, I.R. Last and K.A. Prebble, Standardization of near-infrared spectrometric instruments, *Anal. Chem.*, **68**, 982–990 (1996).
63. M. Blanco and M.A. Romero, Near-infrared libraries in the pharmaceutical industry: a solution for identity confirmation, *Analyst*, **126**, 2212–2217 (2001).
64. N. Smola and U. Urleb, Near infrared spectroscopy applied to quality control of incoming material in the pharmaceutical industry, *Farmacevtski Vestnik*, **50**, 296–297 (1999).
65. S.H. Scafi and C. Pasquini, Identification of counterfeit drugs using near-infrared spectroscopy, *Analyst* **126**, 2218–2224 (2001).
66. M. Ulmschneider, G. Barth and E. Trenka, Building transferable cluster calibrations for the identification of different solid excipients with near-infrared spectroscopy, *Drugs*, **43**, 71–73 (2000).
67. Y. Ren, W. Li, Y. Guo, R. Ren, L. Zhang, D. Jin and C. Hui, Study on quality control of metronidazole powder pharmaceuticals using near infrared reflectance first-derivative spectroscopy and multivariate statistical classification technique, *Jisuanji Yu Yingyong Huaxue*, **14**, 105–109 (1997).
68. W.L. Yoon, N.C. North, R.D. Jee and A.C. Moffat, Application of a polar qualification system in the near infrared identification and qualification of raw pharmaceutical excipients. In Davies, A.M.C. and Giangiacomo, R. (eds), *Near Infrared Spectroscopy, Proceedings of the International Conference, 9th*, Verona, Italy, June 13–18, 1999, NIR Publications, Chichester, UK, pp. 547–550, 2000.
69. E. Hackmann, E. De Abreu and M. Santoro, Corn starch identification by near infrared spectroscopy, *Revista Brasileira de Ciencias Farmaceuticas*, **35**, 141–146 (1999).
70. Y. Woo, H. Kim and J. Cho, Identification of herbal medicines using pattern recognition techniques with near-infrared reflectance spectra, *Microchem. J.*, **63**, 61–70 (1999).
71. E. Hernandez Baltazar and B.G. Rebollar, Development of identification method by near-infrared spectroscopy: Acetaminophen tablets, *Revista Mexicana de Ciencias Farmaceuticas*, **33**, 42–47 (2002).
72. A. Candolfi, R. De Maesschalck, D.L. Massart, P.A. Hailey and A. Harrington, Identification of pharmaceutical excipients using NIR spectroscopy and SIMCA, *J. Pharm. Biomed. Anal.*, **19**, 923–935 (1999).
73. P.R. Khan, R.D. Jee, R.A. Watt and A.C. Moffat, The identification of active drugs in tablets using near infrared spectroscopy, *J. Pharm. Sci.*, **3**, 447–453 (1997).
74. V.W. Uhl and J.B. Gray, *Mixing: Theory and Practice*, Academic Press, New York, 1967.
75. R.E. Schirmer, *Modern Methods of Pharmaceutical Analysis*, vol. **2**, 2nd edition, CRC Press, Boca Raton, pp. 31–126, 1991.
76. R. Jensen, E. Peuchant, I. Castagne, A.M. Boirac and G. Roux, One-step quantification of active ingredient in pharmaceutical tablets using near-infrared spectroscopy, *Spectroscopy*, **6**(2), 63–72 (1988).
77. S. Lonardi, R. Viviani, L. Mosconi, *et al.*, Drug analysis by near-infrared reflectance spectroscopy. Determination of the active ingredient and water content in antibiotic powders, *J. Pharm. Biomed. Anal.*, **7**, 303–308 (1989).
78. E. Dreassi, G. Ceramelli, P. Corti, P.S. Lonardi and P. Perruccio, Near-infrared reflectance spectrometry in the determination of the physical state of primary materials in pharmaceutical production, *Analyst*, **120**, 1005–1008 (1995).
79. J.B. Reeves, Near- versus mid-infrared spectroscopy: Relationships between spectral changes induced by water and relative information content of the two spectral regions in regard to high-moisture samples, *Appl. Spectrosc.*, **49**, 295–903 (1995).

80. W. Plugge and C.J. Van der Vlies, Near-infrared spectroscopy as an alternative to assess compliance of ampicillin trihydrate with compendial specifications, *J. Pharm. Biomed. Anal.*, **11**, 435–442 (1993).
81. P. Corti, E. Dreassi, G. Ceramelli, S. Lonardi, R. Viviani and S. Gravina, Near Infrared Reflectance Spectroscopy applied to pharmaceutical quality control. Identification and assay of cephalosporins, *Analusis* **19**, 198–204 (1991).
82. J.K. Drennen and R.A. Lodder, Pharmaceutical applications of near-infrared spectrometry, in *Advances in Near Infrared Measurements*, vol. **1**, G. Patonay (ed.), JAI Press, Greenwich, CT, pp. 93–112, 1993.
83. E.W. Ciurczak, Following the progress of a pharmaceutical mixing study via near-infrared spectroscopy, *Pharm. Tech.* **15**, 140 (1991).
84. D.J. Wargo. and J.K. Drennen, Near-infrared spectroscopic characterization of pharmaceutical powder blends, *J. Pharm. Biomed. Anal.*, **14**, 1415–1423 (1996).
85. F. Cuesta, J. Toft, B. Van den Bogaert, D.L. Massart, S.S. Dive and P. Hailey, Monitoring powder blending by NIR spectroscopy, *Fresenius J. Anal. Chem.*, **352**, 771–778 (1995).
86. C. van der Vlies, K.J. Kaffka and W. Plugge, Qualifying pharmaceutical substances by fingerprinting with near-IR spectroscopy and the polar qualification system, *Pharm. Technol. Eur.*, **7**, 46–49 (1995).
87. P.A. Hailey, P. Doherty, P. Tapsell, T. Oliver and P.K. Aldridge, Automated system for the on-line monitoring of powder blending processes using near-infrared spectroscopy. Part I: System development and control, *J. Pharm. Biomed. Anal.*, **14**, 551–559 (1996).
88. S. Sekulic, H. Ward, D. Brannegan, E. Stanley, C. Evans, S. Sciavolino, P. Hailey and P. Aldridge, On-line monitoring of powder blend homogeneity by near-infrared spectroscopy, *Anal. Chem.*, **68**, 509–513 (1996).
89. S. Sekulic, J. Wakeman, P. Doherty and P. Hailey, Automated system for the online monitoring of powder blending processes using near-infrared spectroscopy. Part II: Qualitative approaches to blend evaluation, *J. Pharm. Biomed. Anal.*, **17**, 1285–1309 (1998).
90. M. Blanco, R. Gonzalez and E. Bertran, Monitoring powder blending in pharmaceutical processes by use of near infrared spectroscopy, *Talanta*, **56**, 203–212 (2002).
91. W. Plugge and C. Van Der Vlies, The use of near infrared spectroscopy in the quality control laboratory of the pharmaceutical industry, *J. Pharm. Biomed. Anal.*, **10**(11–12), 797–803 (1992).
92. L. Lin and C.W. Brown, Near-IR spectroscopic determination of sodium chloride in aqueous solution, *Appl. Spectrosc.*, **46**, 1809–1815 (1992).
93. S.R. Delwiche, K.H. Norris and R.E. Pitt, Temperature sensitivity of near-infrared scattering transmittance spectra of water-adsorbed starch and cellulose, *Appl. Spectrosc.*, **46**, 782–789 (1992).
94. J. Lin and C.W. Brown, Near-IR fiber-optic temperature sensor, *Appl. Spectrosc.*, **47**, 62–68 (1993).
95. J.E. Sinsheimer and N.M. Poswalk, Pharmaceutical applications of the near infrared determination of water, *J. Pharm. Sci.*, **57**, 2007–2010 (1968).
96. R.M. Issa, K.M. El-Marsafy and M.M. Gohar, Application of the near infrared spectrophotometry as an analytical procedure for the determination of water in organic compounds and pharmaceutical products, *An. Quim.*, **84**, 312–315 (1988).
97. W. Böhme, W. Liekmeier, K. Horn and C. Wilhelm, Water determination by near infrared spectroscopy, *Labor Praxis*, **14**, 86–89 (1990).
98. J. Ludvik, S. Hilgard and J. Volke, Determination of water in acetonitrile, propionitrile, dimethylformamide and tetrahydrofuran by infrared and near-infrared spectrometry, *Analyst*, **113**, 1729–1731 (1988).
99. H. Keutel, J. Gobel and V. Mehta, NIR diode array spectrometer in pharmaceutical production control, *Chemie Technik* **32**, 48–50 (2003).
100. R.M. Leasure and M.K. Gangwer, Near-infrared spectroscopy for in-process moisture determination of a potent active pharmaceutical ingredient, *Am. Pharm. Rev.*, **5**(1), 103–104, 106, 108–109 (2002).
101. A. Dunko and A. Dovletoglou, Moisture assay of an antifungal by near-infrared diffuse reflectance spectroscopy, *J. Pharm. Biomed. Anal.*, **28**, 145–154 (2002).
102. A.Z. David, I. Antal, Z. Acs, L. Gal and D. Greskovits, Investigation of water diffusion in piracetam by microwave moisture measurement and near-infrared spectroscopy, *Hung. J. Ind. Chem.*, **28**, 267–270 (2000).
103. X. Zhou, P. Hines and M.W. Borer, Moisture determination in hygroscopic drug substances by near infrared spectroscopy, *J. Pharm. Biomed. Anal.*, **17**, 219–225 (1998).
104. A. Stokvold, K. Dyrstad and F.O. Libnau, Sensitive NIRS measurement of increased moisture in stored hygroscopic freeze dried product, *J. Pharm. Biomed. Anal.*, **28**, 867–873 (2002).
105. G.A. Birrer, J. Liu, J.M. Halas and G.G. Nucera, Evaluation of a container closure integrity test model using visual inspection with confirmation by near infrared spectroscopic analysis, *J. Pharm. Sci. Tech.*, **54**, 373–382 (2000).
106. U. Buhler, E. Maier and M. Muller, Determination of the water content in lyophilizates with NIR, *J. Pharm. Tech.*, **19**, 49–55 (1998).

107. M. Savage, J. Torres, L. Franks, B. Masecar and J. Hotta, Determination of adequate moisture content for efficient dry-heat viral inactivation in lyophilized factor VIII by loss on drying and by near infrared, *Biologicals*, **26**, 119–124 (1998).
108. S.M. Han and P.G. Faulkner, Determination of SB 216469-S during tablet production using near-infrared reflectance spectroscopy, *J. Pharm. Biomed. Anal.*, **14**, 1681–1689 (1996).
109. M.W.J. Derksen, P.J.M. Van De Oetelaar and F.A. Maris, The use of near-infrared spectroscopy in the efficient prediction of a specification for the residual moisture content of a freeze-dried product, *J. Pharm. Biomed. Anal.*, **17**, 473–480 (1998).
110. T. Moy, A. Calabria and C. Hsu, Near infrared: a non-invasive method to determine residual moisture of lyophilized protein pharmaceuticals, American Chemical Society 221st National Meeting, BIOT-016. Meeting abstract, 2001.
111. J.A. Jones, I.R. Last, B.F. MacDonald and K.A. Prebble, Development and transferability of near-infrared methods for determination of moisture in a freeze-dried injection product, *J. Pharm. Biomed. Anal.* **11**, 1227–1231 (1993).
112. M.S. Kamat, R.A. Lodder and P.P. DeLuca, Near-infrared spectroscopic determination of residual moisture in lyophilized sucrose through intact glass vials, *Pharm. Res.*, **6**, 961–965 (1989).
113. X. Zhou, P. Hines and M.W. Borer, Moisture determination in hygroscopic drug substances by near infrared spectroscopy, *J. Pharm. Biomed. Anal.*, **17**, 219–225 (1998).
114. E.W. Ciurczak, R.P. Torlini, and M.P. Demkowicz, Determination of particle size of pharmaceutical raw materials using near-infrared reflectance spectroscopy, *Spectroscopy*, **1**(7), 36 (1986).
115. J.L. Ilari, H. Martens and T. Isaksson, Determination of particle size in powders by scatter correction in diffuse near-infrared reflectance, *Appl. Spectrosc.* **42**, 722–728 (1988).
116. M. Blanco, J. Coello, H. Iturriaga, S. Maspoch, F. González and R. Pous, Near infrared spectroscopy. In *Bridging the Gap between Data Analysis and NIR Applications*, K.I. Hildrum, T. Isaksson, T. Næs and A. Tandberg (eds), Ellis Horwood, Chichester, 1992.
117. P. Frake, I. Gill, C.N. Luscombe, D.R. Rudd, J. Waterhouse and U.A. Jayasooriya, Near-infrared mass median particle size determination of lactose monohydrate, evaluating several chemometric approaches, *Analyst*, **123**, 2043–2046 (1998).
118. A.J. O'Neil, R.D. Jee and A.C. Moffat, The application of multiple linear regression to the measurement of the median particle size of drugs and pharmaceutical excipients by near-infrared spectroscopy, *Analyst*, **123**, 2297–2302 (1998).
119. A.J. O'Neil, R.D. Jee and A.C. Moffat, Measurement of the cumulative particle size distribution of microcrystalline cellulose using near infrared reflectance spectroscopy, *Analyst*, **124**, 33–36 (1999).
120. M. Blanco and A. Peguero, An expeditious method for determining particle size distribution by near infrared spectroscopy: comparison of PLS2 and ANN models, *Talanta*, **77**(2), 647–651 (2008).
121. J.D. Kirsch and J.K. Drennen, Nondestructive tablet hardness testing by near-infrared spectroscopy: a new and robust spectral best-fit algorithm, *J. Pharm. Biomed. Anal.* **19**, 351–362 (1999).
122. M. Blanco, M. Alcalá, J.M. Gonzalez and E. Torras, A process analytical technology approach based on near infrared spectroscopy: tablet hardness, content uniformity and dissolution test measurements of intact tablets, *J. Pharm. Sci.*, **95**, 2137–2144 (2006).
123. M. Blanco, M. Alcalá, J.M. Gonzalez and E. Torras, Non-destructive dissolution testing by NIR spectroscopy, *NIRNews* **18**(6), 10–12 (2007).
124. M. Blanco, M. Alcalá, J.M. Gonzalez and E. Torras, Determination of dissolution profiles in intact pharmaceutical tablets by NIR spectroscopy, *J. Proc. Anal. Tech.*, **3**(5), 25–28 (2006).
125. Y. Roggo, N. Jent, A. Edmond, P. Chalus and M. Ulmschneider, Characterizing process effects on pharmaceutical solid forms using near-infrared spectroscopy and infrared imaging, *Eur. J. Pharm. Biopharm.* **61** 100–110 (2005).
126. J.D. Kirsch and J.K. Drennen, Determination of film-coated tablet parameters by near-infrared spectroscopy, *J. Pharm. Biomed. Anal.*, **13**, 1273–1281 (1995).
127. E.W. Ciurczak, Use of near infrared in pharmaceutical analyses, *Appl. Spectrosc. Rev.*, **23**, 147 (1987).
128. B.G. Osborne and T. Fearn, *Near Infrared Spectroscopy in Food Analysis*, John Wiley & Sons, New York, 1986.
129. P.K. Aldridge, C.L. Evans, H.W. Ward, S.T. Colgan, N. Boyer and P.J. Gemperline, Near-IR detection of polymorphism and process-related substances, *Anal. Chem.* **68**, 997–1002 (1996).
130. T. Norris, P.K. Aldridge and S.S Sekulic, Determination of end-points for polymorph conversions of crystalline organic compounds using on-line near-infrared spectroscopy, *Analyst*, **122**, 549–552 (1997).
131. R. Gimet and A.T. Luong, Quantitative determination of polymorphic forms in a formulation matrix using the near infra-red reflectance analysis technique, *J. Pharm. Biomed. Anal.*, **5**, 205–211 (1987).

132. P.E. Luner, S. Majuru, J.J. Seyer and M.S. Kemper, Quantifying crystalline form composition in binary powder mixtures using near-infrared reflectance spectroscopy, *Pharm. Dev. Tech.*, **5**, 231–246 (2000).
133. M. Otsuka, F. Kato, Y. Matsuda and Y. Ozaki, Comparative determination of polymorphs of indomethacin in powders and tablets by chemometrical near-infrared spectroscopy and X-ray powder diffractometry, *AAPS PharmSciTech.*, **4**, 147–158 (2003).
134. A.D. Patel, P.E. Luner and M.S. Kemper, Quantitative analysis of polymorphs in binary and multi-component powder mixtures by near-infrared reflectance spectroscopy, *Int. J. Pharm.* **206**, 63–74. Erratum in: *Int. J. Pharm.*, **212**, 295 (2000).
135. C.M. Deeley, R.A. Spragg and T.L.A. Threlfall, Comparison of Fourier transform infrared and near-infrared Fourier transform Raman spectroscopy for quantitative measurements: an application, *Spectrochim. Acta, Part A*, **47A**, 1217–1223 (1991).
136. M. Blanco, J. Coello, H. Iturriaga, S. Maspoch and C. Perez-Maseda, Determination of polymorphic purity by near infrared spectrometry, *Anal. Chim. Acta*, **407**, 247–254 (2000).
137. M. Blanco, D. Valdes, M.S. Bayod, F. Fernandez-Mari and I. Llorente, Characterization and analysis of polymorphs by near-infrared spectrometry, *Anal. Chim. Acta*, **502**, 221–227 (2004).
138. M. Blanco, M. Alcalá, J.M. Gonzalez and E. Torras, NIR Spectroscopy in the study of polymorphic transformations. *Anal. Chim. Acta*, **567**, 262–268 (2006).
139. R.J. Romanach and M.A. Santos, Content uniformity testing with near infrared spectroscopy, *Am. Pharm. Rev.*, **6**(2), **62**, 64–67 (2003).
140. R.C. Lyon, E.H. Jefferson, C.D. Ellison, L.F. Buhse, J.A. Spencer, M. Nasr and A.S. Hussain, Exploring pharmaceutical applications of near-infrared technology, *Am. Pharm. Rev.*, **6**(3), **62**, 64–66, 68–70 (2003).
141. E.W. Ciurczak and J.K. Drennen, Pharmaceutical applications of near-infrared-spectroscopy. In *Near-Infrared Applications in Biotechnology*, R. Raghavachari (ed.), Marcel Dekker Inc., New York, 2001.
142. M. Blanco, J. Coello, H. Iturriaga, S. Maspoch and C. Pezuela, Near-infrared spectroscopy in the pharmaceutical industry, *Analyst*, **123**, 135–150 (1998).
143. E.W. Ciurczak and J.K. Drennen, Near-infrared spectroscopy in pharmaceutical and biomedical applications. In *Handbook of Near-Infrared Analysis*, 2nd edition, D. Burns and E.W. Ciurczak (eds), Marcel Dekker Inc., New York, 2001.
144. J. Luypaert, D.L. Massart and Y. Vander Heyden, Near-infrared spectroscopy applications in pharmaceutical analysis, *Talanta*, **72**, 865–883 (2007).
145. L. Tian, Z. Tang, F. Liu, Y. Gou, Y. Guo and Y. Ren, Nondestructive quantitative analysis of aspirin by near infrared spectroscopy and artificial neural network, *Fenxi Shiyanshi*, **20**(1), 79–81 (2001).
146. M. Blanco, J. Coello, H. Iturriaga, S. Maspoch and N. Pou, Development and validation of a near infrared method for the analytical control of a pharmaceutical preparation in three steps of the manufacturing process, *Fresenius' J. Anal. Chem.*, **368**, 534–539 (2000).
147. P. Merckle and K.A. Kovar, Assay of effervescent tablets by near-infrared spectroscopy in transmittance and reflectance mode: acetylsalicylic acid in mono and combination formulations, *J. Pharm. Biomed. Anal.*, **17**, 365–374 (1998).
148. Y. Ren, Y. Gou, Z. Tang, P. Liu and Y. Guo, Nondestructive quantitative analysis of analgin powder pharmaceutical by near-infrared spectroscopy and artificial neural network technique, *Anal. Lett.*, **33**, 69–80 (2000).
149. M. Blanco, J. Coello, A. Eustaquio, H. Itturriaga and S. Maspoch, Development and validation of methods for the determination of miokamycin in various pharmaceutical preparations by use of near infrared reflectance spectroscopy, *Analyst*, **124**, 1089–1092 (1999).
150. R. Cinier and J. Guilment, Quantitative analysis of resorcinol in aqueous solution by near-infrared spectroscopy, *Vib. Spectrosc.*, **11**, 51–59 (1996).
151. M. Alcalá, J. León, J. Ropero, M. Blanco and R.J. Romañach, Analysis of low content drug tablets by transmission near infrared spectroscopy: selection of calibration ranges according to multivariate detection and quantitation limits of PLS models, *J. Pharm. Sci*, **97**(12), 5318–5327 (2007).
152. M. Blanco, M. Bautista and M. Alcalá, Preparing calibration sets for use in pharmaceutical analysis by NIR Spectroscopy, *J. Pharm. Sci.*, **97**, 1236–1245 (2008).

15

Near-Infrared Spectroscopy (NIR) as a PAT Tool in the Chemical Industry: Added Value and Implementation Challenges

Ann M. Brearley[1] and Susan J. Foulk[2]

[1] *Plymouth MN, USA*
[2] *Guided Wave Inc., Rancho Cordova, CA, USA*

15.1 Introduction

The drive to improve asset productivity, lower manufacturing costs, and improve product quality, which has been occurring throughout the chemical industry for the last several decades, has led to increased focus on process understanding, process improvement and the use of advanced process control. A key aspect of many of these efforts is improvement in the quality and frequency of the analytical data used to control the process or release product. Process analyzers are increasingly being used in order to improve the timeliness, frequency, or precision of analytical data. Multivariate spectroscopic analyzers (NIR, IR, Raman, etc.) can be used to monitor the chemical composition of a process stream noninvasively, in near real time, and are available commercially from numerous vendors. However, even after several decades of development and implementation, the technology is still relatively new to the chemical industry and there remains a steep learning curve for any given installation. A major reason for this is that process analytical chemistry (PAC) and chemometrics are not generally part of the standard educational curriculum for an analytical chemist, or for a process chemist or process engineer, or even for an analyzer engineer (the latter traditionally focusing on univariate sensors for pressure, temperature, flow, level, pH, etc.).

The objective of this chapter is to reduce the learning curve for the application of near-infrared (NIR) spectroscopy, or indeed any process analytical technology, to the chemical industry. It attempts to communicate realistic expectations for process analyzers in order to minimize both unrealistically positive expectations (e.g. 'NIR can do everything and predict the stock market!') and unrealistically negative expectations (e.g. 'NIR never works; don't waste your money'). The themes of this chapter are value and

Process Analytical Technology 2e Edited by Katherine Bakeev
© 2010 John Wiley & Sons, Ltd.

challenge: specifically business value and technical challenge, in that order. Realistic and thorough assessment of the business value to be gained by a process analyzer installation is critical to the success of the project. This assessment should be the very first step of the project since it serves to guide, motivate and prioritize all subsequent work. Technical (and nontechnical) challenges are part of any process analyzer installation. If not properly addressed, these challenges can become a huge sink for time and money and can even lead to project failure. The discussion and examples in this chapter are designed to help potential users of NIR (or other process) analyzers anticipate and be prepared for the most common challenges. NIR analyzers have enabled significant improvement of many industrial processes, and can do the same for many others, but only if they are implemented for the right (business) reasons by those with the knowledge and experience to properly address the challenges.

The first part of this chapter (Section 15.2) discusses the value of and challenges involved in implementing NIR technology in the chemical industry. It also describes a process for successful process analyzer implementation. Similar issues as they relate to the pharmaceutical industry are covered in Chapter 2. The second part of this chapter (Section 15.3) gives a number of examples of actual NIR applications in the chemical industry, which serve both to illustrate the points from the first part and to demonstrate the range of applications for NIR technology.

15.2 Successful Process Analyzer Implementation

The focus of this section is on how to 'do it right.' Implementation of a process analyzer (whether NIR or another technology) in a chemical manufacturing plant requires a large upfront investment of time, effort, and money. Analyzer implementations are complex projects, both technically and organizationally, requiring the coordination of resources and people across multiple organizational, functional, and geographic boundaries. If the implementation is done correctly, the analyzer can potentially save the business large amounts of money and also enable the plant to improve quality to a previously unattainable level. If the implementation is not done correctly, the business may lose its investment and acquire a lasting bad opinion of process analyzers, and the people involved in the project may get black marks on their records. This is clearly undesirable. It is important to keep the primary emphasis on project success rather than failure and focus on what the process analytical chemist needs to do to ensure a successful implementation. There are also many factors that increase the chances of failure of an analyzer implementation, and these are discussed in Section 15.2.5.

Section 15.2.1 describes a process for 'doing it right.' It is written broadly enough to apply to process analyzers in general, not just NIR analyzers, and to most industries. It can also be applied to the development of new methods using existing analyzer hardware (e.g. a new method or calibration model to monitor an additional analyte in the same process stream). After the overview of the guidelines, there is more detailed discussion in Sections 15.2.2–15.2.7 on what needs to be considered at the various steps.

15.2.1 A process for successful process analyzer implementation

The first issue that should be addressed when a process analyzer project is proposed is to define project roles: who will be doing what. There are four basic roles in an analyzer project: the customer(s), the sponsor, the project manager, and the receiver. These are described below.

1. *Customer(s):* The person(s) who will use the information from the proposed new method or analyzer. This could be a control engineer interested in improving control of the manufacturing process (tuning control loops, understanding dynamics, etc.), a process engineer interested in improving process yield or uptime (shortening start-ups or transitions, shortening batches), an R&D scientist interested in

improving understanding of the process, or a QA/QC (quality control) person interested in speeding up product testing and release. The customer must understand and be able to quantify the business opportunity or need.

2. *Sponsor:* The person who approves and pays for the project. This is generally a management-level person such as a plant manager or technical director, often in the line management of the customer.
3. *Project manager:* The person responsible for development, installation, and validation of the new method and/or analyzer. This person is often a process analytical chemist, an analytical chemist, or an analyzer engineer. Sometimes the technical aspects of the role may be contracted out to a consultant or vendor; in this case the project manager may be a process engineer or lab chemist; nevertheless s/he still has the responsibility for the success of the project. The project manager coordinates the efforts of people across organizational, functional, and often geographic lines, including vendors, analytical personnel, R&D personnel, operations personnel (engineers, chemists, process technicians), crafts people (electricians, etc.), QA/QC personnel, etc. S/he is responsible for communicating effectively so that all people involved in the project (including the sponsor and the customer) are kept informed of the project plan and their part in it.
4. *Receiver:* The person responsible for ensuring that the new method and/or analyzer is used and maintained correctly on site once it is installed and validated. This is nearly always an operations person, often a dedicated analyzer technician or a lab analytical technician.

Once the roles are established the implementation process can formally begin. The 10-step process described below can greatly increase the probability that the analyzer will be a success. Note that while the process is described as a linear series of 10 steps, in reality there may be iterations: for example, if feasibility testing (Step 3) shows that the technical requirements cannot be met with existing technology, one will need to loop back and revisit the technical requirements (Step 2). This process is not intended to be burdensome, but rather to keep the process analytical chemist focused on the various tasks that are needed for successful implementation. The length of time each step takes will depend on the magnitude of the proposed project. For example, a business value assessment may involve as little as a 30-minute discussion between two people, or as much as a month-long exercise involving dozens of people from multiple plants on multiple continents. The project manager does not personally carry out every step – some of the tasks can or must be delegated – but s/he is responsible for ensuring that each of the steps is done properly, documented, and the results communicated to all concerned.

Step 1 *Business value assessment*: Identify and quantify the business value for the proposed new method or analyzer. This is done jointly by the customer and the project manager. Ideally the value of the technology can be expressed concretely in monetary terms (e.g. dollars) as, for example, a net present value (NPV) or an internal rate of return (IRR). It is critical to include a realistic estimate of the costs of implementing and maintaining the analyzer, as well as the benefits to be realized from it. This assessment is used to prioritize effort and expenses and to justify any capital purchases needed. The various ways that process analyzers can contribute to business value are discussed in Section 15.2.2.

Step 2 *Technical requirements*: Determine the specific technical requirements for the new method and/or analyzer that will ensure that it captures the previously determined business value. This is also done jointly by the customer and the project manager. Avoid 'nice to have' and focus on what is necessary. Technical requirements may include accuracy, precision, specificity, turnaround time, response time, robustness, maintenance needs, etc. They should be focused on *what* the analyzer needs to do (e.g. determine the concentration of component X with 1.5% relative precision, once a minute, in a corrosive process stream at 80 °C), not *how* it will be done. Issues that need to be considered in setting technical requirements are discussed in Section 15.2.3.

Step 3 *Feasibility assessment*: Determine (via literature search, consultation, etc.) suitable candidate methods and assess the feasibility of these methods. The purpose of this step is to collect enough information to accurately define (or 'scope') and prioritize the project. The amount of information required may vary considerably, depending on the magnitude of the proposed project. Significantly less information is required to develop a new method for an existing off-line analyzer than is required to justify spending several million dollars on a major installation of on-line analyzers. (The capabilities and limitations of NIR process analyzers are discussed in Section 15.2.4.)

Step 4 *Project initiation and approval*: Define a project to develop and implement the proposed new method or analyzer and get it approved by the sponsor. The documentation required by the sponsor is usually company-specific, but will typically include: goal, business value, critical technical issues to be addressed, ramifications, proposed approach, resources needed (including personnel time), deliverables, and timing.

Step 5 *Off-line method or analyzer development and validation*: This step is simply standard analytical chemistry method development. For an analyzer that is to be used off-line, the method development work is generally done in an R&D or analytical lab and then the analyzer is moved to where it will be used (QA/QC lab, at-line manufacturing lab, etc.). For an analyzer that is to be used on-line, it may be possible to calibrate the analyzer off-line in a lab, or in situ in a lab reactor or a semiworks unit, and then move the analyzer to its on-line process location. Often, however, the on-line analyzer will need to be calibrated (or recalibrated) once it is in place (see Step 7). Off-line method development and validation generally includes: method development and optimization, identification of appropriate check samples, method validation, and written documentation. Again, the form of the documentation (often called 'the method' or 'the procedure') is company-specific, but it typically includes: principles behind the method, equipment needed, safety precautions, procedure steps, and validation results (method accuracy, precision, etc.). It is also useful to document here which approaches did not work, for the benefit of future workers.

Step 6 *Installation*: This is where the actual analyzer hardware is physically installed in its final location and started up. For both off-line and on-line analyzers, this may involve obtaining equipment (analyzer, sampling system, analyzer shelter, etc.), modifying facilities (piping, electrical, etc.), installing equipment, writing procedures, and training people. The project manager is responsible for ensuring that the method and analyzer are used correctly on site; s/he may choose to personally train all technicians and operators, to delegate the training but periodically audit their performance, or some combination.

Step 7 *On-line analyzer calibration*: Calibrating an analyzer entirely on-line is a last resort, for reasons discussed in Section 15.2.6. It is preferable to do at least the initial work to calibrate the analyzer off-line, or to transfer to the on-line analyzer a method developed on an off-line analyzer or on another on-line analyzer. However, sometimes this is not possible. On-line analyzer calibration is similar to standard analytical chemistry method development, except that getting sufficient variation in sample composition to build a robust calibration model may be difficult or may take a long time. (Ways to address the challenges involved in on-line calibration are discussed in Section 15.2.6 and in Chapter 14.)

Step 8 *Method or analyzer qualification*: Qualify the method on site: that is, assess the ability of the method *as practiced* to meet the original technical requirements such as speed, accuracy, and precision. This is best done jointly by the project manager and the receiver (who is the ultimate owner).

> *Note*: It is assumed that if the analyzer meets the previously agreed-upon technical requirements, it will then deliver the agreed-upon business value. Some companies may require that this be verified, for example, by tracking the savings over a specified period of time, and the results included in the project closure documentation.

Step 9 *Method maintenance*: Incorporate the new method or analyzer into the existing method maintenance systems for the site. This ensures that the method as practiced continues to meet the technical requirements

for as long as it is in use. This is done by the receiver. Method maintenance systems may include check sample control charting, intra- and/or inter-lab equivalence testing, on-site auditing, instrument preventive maintenance (PM) scheduling, control charting the method precision and/or accuracy, etc.

Step 10 *Project closure*: Document the results of the project in a technology transfer package or project closure report. The form of the documentation will be company-specific, but typically includes the documentation from the previous steps (business value assessment, technical requirements, method development report, etc.). It may also (especially for on-line analyzers) include engineering drawings, vendor manuals, standard operating procedures (SOPs), training materials, maintenance schedules, etc. Formal sign off on the project closure package by the customer(s), the receiver, and the sponsor may be required to formally end the project manager's responsibilities.

That's it! All done! Throw a party and celebrate success!

15.2.2 How NIR process analyzers contribute to business value

Process analyzer technology can contribute value to the business in two basic ways: by reducing risks and by reducing costs.

Risks can be related to safety, such as the risk of a runaway reaction, or to money, such as the risk of ruining a batch of an expensive product by mistakenly adding a wrong ingredient. Process analyzers can reduce safety risks by monitoring safety-critical process parameters in real time and automatically shutting down the process if a parameter goes outside of safe limits. On-line process analyzers can also reduce safety risks by reducing the need for routine process sampling, which reduces the exposure of process operators to potentially hazardous conditions. Process analyzers can reduce monetary risks by rapidly inspecting raw materials or in-process materials to assure that they are the correct material, or by detecting process deviations that can be corrected to prevent loss of a batch.

Cost reductions usually arise out of improvements to the process control for both continuous and batch processes. Process analyzers enable chemical composition to be monitored essentially in real time. This in turn allows control of the process to be improved by shortening start-up and transition times (for continuous processes) or batch cycle times (for batch processes). This is accomplished by improving the ability to respond to process disturbances, by enabling process oscillations to be detected and corrected, and by reducing product variability. Real-time monitoring of chemical composition in a process allows a manufacturing plant to:

- increase product yield by decreasing off-spec production
- increase product yield by reducing start-up times, transition times, or batch cycle times
- improve product quality (consistency)
- reduce product 'giveaway' (e.g. run closer to aim)
- reduce in-process inventory
- reduce consumption of expensive ingredients
- reduce production of waste streams that are expensive to treat or dispose.

Occasionally attempts to justify a process analyzer are made on the basis that it will reduce the number of grab samples that need to be taken from the process to be analyzed, and therefore reduce the sample load on the QC lab. This is not usually a good justification for an on-line analyzer for two reasons: (1) The reduction in lab analyses rarely offsets the cost of analyzer installation and operation; (2) An on-line analyzer relies on high quality lab data for calibration, validation, and long-term model maintenance, so the reference method can never be eliminated entirely.

15.2.2.1 Business Value Example

A fictitious example illustrates the large potential value of even small improvements in the control of a manufacturing process. Suppose one has a continuous process in which the final product (a polymer) is sampled and analyzed to be sure the copolymer composition is within specifications. A sample is taken from the process once every 2 hours, and it takes about 2 hours for the lab to dissolve the polymer and measure its composition. This process produces a number of different copolymer compositions, and it transitions from one product to another about twice a month on average. The 2-hour wait for lab results means that during a transition the new product has been within specification limits for 2 hours before the operators receive lab confirmation and are able to send the product to the 'in-spec' silo. Consequently, on every transition, 2-hours' worth of in-spec polymers are sent to the off-spec silo.

Suppose that implementation of NIR technology in this process (either off-line or on-line) allows the lab delay to be reduced from 2 hours to a few minutes (effectively zero). This means that, during a transition, the operators know immediately when the new product has reached specifications and can immediately send it to the in-spec silo. This reduces the transition time by two hours for every transition. Suppose this plant makes 15 000 pounds (6800 kg) of polymer per hour, and that off-spec polymer sells for US$0.70 less per pound (US$1.54 less per kg) than in-spec polymer. The 2-hour reduction in transition time will save US$504 000 year. This is a conservative estimate because it does not include other contributions that real-time monitoring will make, such as shortening the *beginning* of product transitions (by not sending the previous product to the 'off-spec' silo until it actually goes out of specification limits), shortening start-ups, and enabling quicker recoveries from process upsets.

Note: The costs of implementing this fictitious analyzer have not been included in the value calculation, though it is a critical aspect. The costs would include hardware, software, facility modifications, personnel time, travel, and ongoing maintenance costs (including the analyzer technician's time) over the expected analyzer lifetime.

15.2.3 Issues to consider in setting technical requirements for a process analyzer

This section describes the issues that need to be considered during Step 2 (technical requirements) of a process analyzer implementation. As stated earlier, the technical requirements should be focused on *what* the analyzer needs to do (e.g. determine the concentration of component X with 1.5% relative precision, once a minute, in a corrosive process stream at 80 °C), not *how* it will do it (e.g. with on-line s chromatography (GC) or off-line diffuse reflectance NIR), although the issues do overlap somewhat. It is critical at this stage to separate the 'would be nice' requirements from those that are truly necessary to gain the already-determined business value. For example, if the business value is gained by shortening product transition times by two hours (as in the fictitious example above), the new analyzer only needs to be fast; it does not need to be any more accurate or precise than the existing laboratory method. If the business value is to be gained by detecting and responding earlier to process upsets, then the analyzer needs to be fast, sensitive, and non-noisy; but it does not need to be quantitatively accurate.

The issues to consider in this step are presented in the form of questions to be asked. The questions should be asked more than once to people in different functions at many levels, since they may have different levels of knowledge about the business and the process, and different perspectives on the relative importance of various issues. The first answer is not necessarily the right answer.

- What type of measurement needs to be done? Does it need to provide quantitative information (concentrations) or qualitative information (identification, qualification {ok/not ok}, or trends)? If quantitative, what accuracy and precision are required? If qualitative, what level of errors can be tolerated?

- What are the analytes of interest? What are their typical concentration ranges? Do the concentrations of any of the analytes covary? What is the accuracy and precision of the reference method for each analyte?
- What other components are present in the process stream of interest: (a) under normal conditions? and (b) under process upset conditions? What are their concentration ranges? Do the concentrations of any components covary with any of the analytes of interest?
- In what type of process (batch or continuous, single product or multiple products, long campaigns or short ones) will the measurement be made?
- What will the measurement results be used for? Will they be used for process control (closed loop, open loop, feed-forward, feedback)? Will they be used as a safety interlock? (This puts very strict requirements on the analyzer reliability, and may even require duplicate analyzers.) Will the results be used to accept raw materials or to release product? Will they be used to sort or segregate materials? How frequently does the measurement need to be made? How rapid does one individual measurement need to be? How accurate does it have to be? How precise? How much analyzer downtime is acceptable?
- Are there external regulations (such as FDA, EPA, or USDA regulations) that apply to this process and any measurements made on it?
- Will the measurement be made on-line or off-line? For on-line analyzers, is there a grab sample port nearby? Is the sample stable once it is withdrawn from the process? (See Chapter 3 for more information on sampling issues.) Only an on-line analyzer can enable real-time process control. However, an on-line analyzer implementation will generally be more expensive, more complex, and involve more people in more functions than an off-line one.
- What is the nature of the material to be measured? Is it a gas, a liquid, a slurry or a solid? If solid, is it a powder, pelletized resin, fiber, film, or web? How large is it? Is it homogeneous? Is it clear, cloudy, or opaque? Is it colored? Viscous? Bubbly? Hot? Cold? Under high pressure? Under vacuum? Corrosive? Highly toxic? Is it moving, and if so, how rapidly? Does it need to remain sterile?

These questions will undoubtedly lead to more questions specific to the business and process at hand. They are not intended to be exhaustive, but to remind the process analytical chemist of the kinds of issues that need to be discussed during Step 2.

15.2.4 Capabilities and limitations of NIR

Before selecting a process analytical technology to implement, it is helpful to understand the capabilities and limitations of the technology. Good introductions to NIR spectroscopy and instrumentation can be found in Chapter 5, and to chemometric methods in Chapter 12. NIR pharmaceutical applications are covered in Chapters 13 and 14. Additional information on NIR techniques and applications can be found in Williams,[1] Siesler,[2] and Burns.[3]

Note: The focus of this book is on multivariate analyzers, but it is worth noting that there are still many process analyzer implementations in the chemical industry that are successfully using older filter NIR and IR technology. In some cases, business value can be gained by replacing the older technology with newer multivariate analyzers, but in other cases the cost of doing so exceeds the benefit. It is not necessarily true that the latest technology has the greatest value to the business, a fact which it is sometimes easy for technical people to forget. What is important about a technology is not how new it is, but whether it has the specific capabilities needed to capture business value (e.g. shorten transition times, increase first-quality yields, etc.).

In general, NIR can be used to monitor compounds that contain C–H, O–H, or N–H bonds. NIR is generally used to monitor bulk composition (components at levels of 0.1% and higher), rather than trace-level composition. However, NIR is very sensitive to water, alcohols, and amines, and under the right circum-

stances can be used to monitor them at levels well below 0.1%. NIR can be used for both quantitative and qualitative analyses. Qualitative methods can include identification or qualification of incoming raw materials, verification of product identity or quality, monitoring the trajectory of a reaction, and rapid sorting. The results of NIR monitoring (whether quantitative or qualitative) can be used to control a process (feed-forward or feedback control, closed or open loop control), to decide when a batch reaction is complete, to accept raw materials, to release product, or as a safety interlock (e.g. to shut a plant down automatically if the level of a critical parameter goes out of preset limits or if the analyzer itself malfunctions).

Two of the major benefits of NIR are *speed* and *precision*. NIR analyzers provide rapid results, generally in less than a minute. The speed of analysis is a result of the scan speed, the high signal-to-noise performance of the analyzers themselves, and because sample preparation is generally minimal. The speed advantage can be enhanced by putting the analyzer on-line or in-line. This also eliminates the need for routine sampling of the process, which can be time-consuming, may not provide representative samples, and may involve safety risks. The speed advantage is further enhanced by the ability of NIR to monitor multiple components simultaneously. NIR analyzers provide high-precision results, due primarily to the elimination of variability arising from sampling and sample preparation, as well as to the high stability of the analyzers themselves. Another potential benefit is the ability to multiplex certain types of NIR analyzers (those using fiber optics). This allows one piece of analyzer hardware to be used at multiple sample points (probes) in one process, or to sample multiple processes. This has the potential to greatly reduce the end cost per measurement for a multiple probe system.

However, NIR (like all spectroscopic methods) also has significant drawbacks. The major drawback is that it takes a significant investment of time, effort, and money to implement a spectroscopic analyzer. There are a number of reasons for this. Spectroscopic methods are secondary methods and must be calibrated against a primary or reference method. The calibration can take considerable time, depending on how long it takes to collect and analyze a suitably varied calibration sample set; and it takes additional time on an ongoing basis to validate and maintain the calibration. NIR can be used to monitor multiple components, but sometimes they are not the intended components (i.e. they are interferences). It takes time and chemometric skill to ensure that the calibration model accurately monitors only the components of interest. (From observation of the interferences encountered while monitoring the process, one can often increase knowledge of the process.) NIR (and other spectroscopic PAT tools) involves a steep learning curve not only for mastering the hardware, but also for mastering software, sampling issues, and chemometrics.

Because of the large investment required to implement NIR process analyzers, there must be a significant business value gained through the analyzer to make the effort worthwhile. This is why it is critical to assess the business value of a potential analyzer at the very beginning of a project. Failure to fully assess the value (both benefits and costs) at the beginning makes it difficult to prioritize effort, justify capital and other expenses, or engage the cooperation of already-busy people in other organizations and functions. The benefits and drawbacks of NIR are summarized in Table 15.1.

15.2.5 General challenges in process analyzer implementation

Challenges arise in implementing any new technology in a manufacturing process, some of which are difficult to anticipate. However, many of the challenges that occur during process analyzer implementations are common ones that have been experienced by other process analytical chemists in the past. Though they may or may not be show-stoppers, they will at minimum cause time and effort to be wasted. These challenges can often be avoided or minimized if recognized and addressed early. Note that some of the most serious challenges facing the process analytical chemist are *not* technical in nature. The outline for successful process analyzer implementation described in Section 15.2.1 (and in Chapter 2) is designed to allow the

Table 15.1 *Benefits and drawbacks of NIR*

Benefits	Drawbacks
Rapid (near real time)	Indirect (secondary method)
Multicomponent capability	Multicomponent capability (interferences)
High precision	Large investment (time, money)
Can be on-line (no sampling)	Steep learning curve
Can be remote (with fiber optics)	Typically not a trace level technology
Can multiplex multiple sample points	
Little or no sample preparation	
Nondestructive	

process analytical chemist to anticipate the avoidable challenges, and to recognize early and minimize the unavoidable ones.

The following factors increase the likelihood of *success* for an NIR process analyzer implementation.

- The project is undertaken for business reasons: the financial stake is sufficiently large to motivate the business to plan, support, and staff for success.
- The business calculates the financial stake for the project taking into account both the installation costs and the continuing support costs of the analyzer.
- The project is coordinated by someone in the business, typically a process engineer or chemist, a control engineer, or an analytical or control lab chemist, rather than by an 'external' resource.
- The analyzer is installed as part of a larger process improvement effort, such as installation of a distributed control system (DCS) and is part of the process control scheme.

The following factors increase the likelihood of *failure* for an NIR or other process analyzer implementation.

- The project is undertaken based on insufficient and/or overly optimistic feasibility studies.
- A technology is chosen that does not meet the technical requirements for the application. Sometimes this is through a lack of knowledge, but more often it is due to not having clarified the essential technical requirements before choosing the technology.
- The hardware chosen does not meet the technical requirements or the vendor is inexperienced in the application or unable to supply necessary support.
- Choice of vendor is particularly critical when analyzer implementation is contracted out. Avoid contract firms which have little or no experience with the specific technology being used. NIR analyzers have different challenges and issues than, for example, GC analyzers.
- There is a failure to recognize the plant-site requirements for NIR calibration and validation, such as the existence of appropriate sampling valves, well-designed sampling protocols, good laboratory reference methods, and variability in the analyte concentrations of interest. More details on these chemometric calibration issues can be found in Chapter 12.
- Model development is not thorough enough to allow the analyzer method to recognize and deal with interferences. This is often due to model development being carried out by someone with insufficient training in chemometrics or insufficient knowledge of the particular chemical process.

- The project is schedule-driven rather than data-driven.
  - Bad hardware decisions are sometimes irreversible.
  - Good model development takes time. Bad models can actually be worse than no models because they may convince the process operators and engineers that the NIR results are unreliable.
- The developer (the person who did the installation and model development) leaves too soon – voluntarily or otherwise. The developer generally needs to maintain involvement well into the long-term prove-out or qualification phase (Step 8). Plant personnel who 'inherit' process analyzer systems before qualification is complete may experience extreme levels of frustration, especially if they have no previous NIR experience and were not involved in the earlier stages of the project. Another way to state this factor is: The business fails to adequately staff the project.
- The project encounters unforeseen difficulties such as previously unsuspected impurities, unexpected fouling of the probes, or unfixable leaks. In addition, sometimes it is found that the reference method, which has been relied on for years, does not actually work.
  - Sometimes surprises occur due to unforeseen factors, but sometimes surprises result from gaps in communication. The process analytical chemist should seek the input of process chemists, engineers, and operators early in the project to maximize the chances of learning about relevant *known* problems. They should spend time at the plant talking to people very familiar with the process. They should ask a lot of questions, and note and pursue any inconsistencies or gaps in the answers.
  - Surprises can also result from inexperience on the part of the process analytical chemist. An example is not anticipating common difficulties such as fouling of probes or condensation in sampling lines. Less experienced people would be well-advised to talk over the proposed project with more experienced process analytical chemists at an early stage.
- There is a lack of communication or cooperation among project team members. Process analyzer work always involves a diverse team. Poor communication among team members can result if a single person is not assigned to be the project manager, if the project manager is not given enough time to do the necessary work, or if the project manager is inexperienced or has poor communication skills. Poor cooperation from team members can often be attributed to lack of clarity about the business value of the work, or to failure of management to adequately prioritize and staff the current projects.
- The project objectives, roles, and responsibilities are unclear.
- Insufficient attention is paid to the general plant environment (dirt, noise, heat, cold, vibration, being hosed down with water, being cooked with a blowtorch, being run into or climbed on by mistake, etc.). These kinds of problems are relatively easy to anticipate and prevent by visiting the plant frequently (if one does not already work there) and watching the process and listening to the people there.

15.2.6 Approaches to calibrating an NIR analyzer on-line

Ideally, an on-line analyzer will be calibrated before it is installed in the process. It may be possible to accomplish this by calibrating it off-line with process grab samples and/or synthetic samples. It may be possible to install the analyzer in a lab-scale reactor, or in a semi-works or pilot plant. It may be possible to transfer to the on-line analyzer a method developed on an off-line analyzer or on another on-line analyzer (e.g. at a different plant site). However, sometimes none of these are possible and the analyzer will need to be calibrated on-line. The challenges of on-line model development (calibration) and validation, as well as approaches to dealing with them, are discussed below. For information related to calibration transfer issues, please see Chapter 12 of this book.

1. *Limited variation in chemical composition:* This is often an issue at the point in the process where the analyzer is installed. A manufacturing plant's goal is to make only good product, with as little variability

as possible. In contrast, the analytical chemist's goal (especially when developing chemometric methods) is to have as much variation in composition as possible represented in the calibration sample set. A commercial-scale plant is rarely willing to produce 'bad' or off-specification product just to calibrate an analyzer, especially if the plant is sold out. However, most plants make some 'bad' product at start-up, at shutdown, or during product transitions. Also, some plants, though they are often reluctant to admit it, make bad product occasionally due to process upsets or 'excursions.' The amount of variation obtainable over a given period of time will depend on the plant: whether it operates continuously or in batch mode, whether it produces a single product or multiple products, how long the product campaigns are (i.e. the time between product transitions), how frequently the plant starts up and shuts down, and how often they have process upsets. One approach to increasing the composition range of the calibration model is to include as much start-up, shutdown, transition, and process upset data as possible. Another approach that can work, if the analyzer is monitoring multiple points in the process and the chemical composition is different at the different points, is to model several different points in the process together.

2. *Covariance among composition variables (analyte concentrations):* The covariance may be coincidental, or it may be fixed by the chemistry of the process and therefore unavoidable. An example of a fixed covariance is in the manufacture of poly (ethylene terephthalate) or PET, where the degree of polymerization (DP) and the concentrations of both carboxyl and hydroxyl end groups all vary together: for a given DP, an increase in hydroxyl ends necessarily means a decrease in carboxyl ends. Coincidental covariance is sometimes less severe when the process is changing, such as during start-ups, transitions, or excursions. One approach to solving this problem is to intentionally break the covariance in the calibration sample set by adding odd samples, or by adding synthetic samples. Another approach is to pay attention to the regression coefficients to be sure they make spectroscopic sense: the coefficients for a given analyte should be similar to the NIR spectrum of that analyte.

   Note: Monitoring an analyte indirectly, by relying on its coincidental covariance with another more easily detected analyte, is risky and should be avoided unless the covariance is 100% ironclad (which is rare).
3. *Infrequent samples:* The plant may not sample the process very often at the point where the analyzer is installed. This could be because the lab turnaround time is so long that the results are too late to be useful, or because the operators do not have time for sampling (they are understaffed), or because they are reluctant to don the personal protective equipment necessary for that particular sample port. Reluctance to wear the 'hot suit' needed to sample a polymer melt process is a common problem, especially in hot climates and during hot seasons of the year. In any case, solving this problem requires gaining the cooperation of the process operators and engineers (and their management). One approach is to negotiate a temporary increase in sampling, with the incentive that it will be possible to reduce or nearly eliminate sampling once the analyzer is calibrated.
4. *Nonexistent samples or reference values:* Samples may not exist because the plant does not exist yet (it is under construction and has not started up yet), or because the plant does not do any sampling at the process point of interest. Reference values may not exist because the plant lab is not set up to perform that particular analysis, or (worse) there is no established reference method for the analyte of interest. In any of these cases, one is left with no data to use for calibration. There are a number of ways to approach this challenge. If there are samples but no reference values, the plant samples can be sent off-site to be analyzed. The analyte concentration of interest can sometimes be estimated based on process conditions and/or the concentrations of other analytes. (This is one place where fixed covariance can come in handy.) If there is no plant yet, it may be possible to calibrate the analyzer elsewhere (different plant, semi-works, etc.). It may also be possible (or even necessary) to attempt 'lab value-less calibration,' in which one assumes that the concentration of the analyte varies linearly with the height

of an absorbance peak characteristic of that analyte (trend analysis). This method can provide qualitative data if the spectroscopy of the system is well-characterized and if there are no significant overlaps between peaks. The use of principal component analysis (PCA) scores to develop qualitative trending methods is another very common approach.

5. *Unstable samples:* The chemical composition of a sample removed from an industrial process often changes after it is removed from the process due to the sample losing or gaining moisture and/or continuing to react (i.e. not being effectively quenched). This is especially common with condensation polymers. Unstable samples result in an offset between the composition measured in the lab and the 'true' composition in the process, and the offset usually varies. Worse, the offset is usually not definable, since there is usually no way of knowing what the composition *would* have been before the sample was removed from the process. The offset appears as an increase in the variability of the lab method, but may be more accurately attributed to the inhomogeneity in the samples A common approach is to minimize the variation in the offset by ensuring consistency (especially in timing) in sampling and lab measurement procedures.
6. *Imprecise reference values:* Imprecise reference values can be due to laboratory method variability or sampling variability. Sampling is quite often the larger source of variability as is detailed in Chapter 3. Sampling variability arises from numerous causes, including errors in writing down the exact sampling time, sample inhomogeneity, and sample instability (as discussed above). Sampling variability will be largest when the process composition is changing the most, such as during start-ups and transitions. One approach is to use huge amounts of data during model development, in order to minimize the effect of the imprecision of any one data point. Another approach is to carefully (usually iteratively) filter out model outliers such as samples with clearly incorrect lab values (e.g. clerical errors), samples taken during transitions where the time stamp is inaccurate, or samples taken during periods of rapid change.

A brief word is in order here about accuracy and precision. The process analytical chemist needs to be aware that the other members of his/her team may not be clear on the difference between analyzer accuracy and analyzer precision, especially in the case of on-line analyzers. S/he should be prepared to explain the difference in language appropriate to their backgrounds. *Analyzer accuracy* refers to the error (difference) between the lab results and the on-line analyzer predictions for a given set of samples. It is often expressed in terms such as standard error of calibration (SEC) or standard error of prediction (SEP). *Analyzer precision* refers to the variability in the on-line analyzer predictions for the same sample over and over. Since most on-line analyzers are installed in flowing streams or other systems which change continuously with time, the analyzer never really sees the same sample over and over. Consequently, analyzer precision is usually estimated from the variability of the predictions over a short period of time while the process is lined out and stable. This gives a high estimate, since it includes both process and analyzer variability. Analyzer precision is often expressed in terms such as standard deviation (SD) or relative standard deviation (RSD) of the predicted value. Analyzer accuracy and precision are independent of each other: an analyzer can easily be accurate without being precise, or precise without being accurate. Analyzer accuracy and analyzer precision are also affected by different factors, as listed below.

The factors which limit on-line process analyzer *accuracy* include:

- error due to unmodelled conditions (bad model)
- sampling error (see Chapter 3)
- time assignment error
  - sample inhomogeneity error
- sample instability error
- nonrepresentative sample

- 'pure' lab error (i.e. the inaccuracy of the reference method itself)
- real process fluctuations
- instrument factors (instrument noise, nonlinearity, temperature instability, drift, etc.).

The factors which limit on-line process analyzer *precision* include:

- real process fluctuations
- instrument factors (instrument noise, nonlinearity, temperature instability, drift, etc.).

Note that there are more sources of variability for accuracy than for precision. This is the basis of the common rule of thumb which states that the precision of an on-line analyzer will typically be about ten times better than its accuracy.

15.2.7 Special challenges in NIR monitoring of polymer melts

Polymers (both for plastics and for fibers) are a major part of the chemical industry. The composition of polymer melts can be of interest in several kinds of systems, including polymerization reactors, polymer melt transfer lines, and extruders. However, there are additional challenges involved when using NIR analyzers in such systems due to a number of unique factors: sustained high temperatures, sudden changes in temperature, pressure extremes (high or low), polymer flow issues, and fouling.

Sustained high temperatures can cause physical deterioration of NIR optical probes if their components are not rated for the process temperatures. This can manifest itself, for example, as either a sudden or a slow drop in the optical efficiency of the probe: the latter is difficult to distinguish from slow probe fouling unless it is possible to remove the probe(s) from the process for testing. High temperatures (especially above 200 °C) also cause significant blackbody emission from the polymer and the process pipe itself. This is a huge source of 'stray' light. Spectra obtained with predispersive NIR analyzers need to be corrected for blackbody emission. High temperatures can also soften probe materials sufficiently so that the probes can be bent while being removed from the process. (With many polymer systems, the probe can be removed from the process only while the polymer is melted.) The probe body should be designed to withstand the process conditions.

Sudden large changes in temperature (either heating or cooling) can crack or break probe windows. In some polymer processes, for example, it is routine practice to deal with flow blockages by heating the section of pipe or piece of equipment with a blowtorch. Careful consideration of the relative thermal expansion coefficients of the window material and the probe materials during probe design can minimize the risk of probe damage in the process. Attention paid to thermal issues when writing or modifying SOPs when an NIR analyzer is implemented can also pay dividends.

High pressures (hundreds or thousands of psig where 1 psig = 6.9 kPa) can break probe windows, or break or loosen the seal between the window and the probe body. This is especially true when high pressures are combined with high temperatures. High pressures can also, in a worst case, push the entire probe out of the process, potentially with dangerous force. The probe itself should of course be designed and tested for the highest pressures expected to occur in the process. In addition, the mechanism by which the probe is secured to the process needs to be carefully designed and thoroughly reviewed for safety. Low pressures can also cause problems. Suction (negative pressures) can occur during a line shutdown or process upset if the polymer freezes. This can pull the window out of a probe, which not only destroys the probe and creates a process leak, but also puts a piece of glass into the process stream which can damage equipment downstream, and contaminate the product. The possibility of low or even negative pressures needs to be taken into account during probe design.

Polymer flow issues are concerns on the part of plant operations personnel that can arise when one proposes to put an in-line NIR probe (or pair of probes) into a polymer reactor or transfer line. These concerns tend to be plant or process-specific. Plant personnel are likely to be concerned if the probe will change the pressure drop in the line, create a cold spot in the reactor or line, protrude into the flow stream, or create dead spots (e.g. recessed probes or the downstream side of protruding probes). There may also be plant- or process-specific restrictions on where probes (or analyzers) can be located, on pipe sizes, on the use of welds, and on materials of construction. It is critical to involve plant operations personnel (including process operators) as early as possible in discussions about probe design and location.

Probe fouling is common in polymer systems. The symptoms are a gradual increase in the baseline absorbance over time (although care must be taken to ascertain that it is not actual haziness in the process itself, or optical degradation of the probes, which is causing the baseline increase). There are many ways to approach this problem, depending on the process. Changes in probe location, orientation (relative to the flow direction), or materials of construction can sometimes reduce fouling. Removal and cleaning of the probes at a set frequency is sometimes possible. Small baseline shifts can be eliminated by baseline correction or other preprocessing of the spectral data. However, probe fouling often causes changes in baseline tilt and curvature as well; these effects can sometimes be dealt with ('modeled out') in the calibration modeling until the fouling levels get too high (see Section 15.3.6). Note that both the baseline-correction approach and the modeling-out approach require high instrument y axis linearity, so that the peak intensities and shapes are not distorted. In addition, a thorough knowledge of both spectroscopy and chemometrics is required for correct implementation.

15.3 Example Applications

The applications described in this second part of the chapter are intended to illustrate the wide range of uses for NIR in the chemical industry. The selection of examples was intentionally limited to work done within industry and published in the open literature in order to keep the focus on work that has demonstrated business value. However, it has been the authors' experience that for every industrial NIR application published in the open literature there are at least two others practiced as trade secrets for business reasons. This is especially true for on-line applications that have progressed beyond the feasibility stage, since the NIR results can reveal a great deal of information about the chemical process itself – information which industry managers are often reluctant to share. Published industrial applications should therefore be considered merely the tip of the iceberg.

The examples also serve to illustrate the points made in the first part of the chapter, with a few exceptions. First, although all of the authors discussed the business value of the project, none of them quantified it (e.g. the analyzer saved US$1.5 million annually by shortening product transitions), presumably because the exact monetary values were considered proprietary business information. Second, because these were technical papers published in technical journals, the focus was on the technical challenges involved in the work, rather than the nontechnical ones. The nontechnical challenges (such as inconsistent management support or lack of communication among team members) are rarely discussed in the open literature, but they can be as much of a barrier to success as the technical ones, if not more; this is why they are emphasized strongly in Section 15.2. Finally, the passive voice used in technical writing makes it difficult to appreciate the large number of people who are generally involved in process analyzer projects.

15.3.1 Monitoring monomer conversion during emulsion polymerization

Chabot *et al.*[4] at Atofina Chemicals (King of Prussia, PA, USA) used in-line NIR to monitor monomer conversion in real time in a batch emulsion polymerization process. The *business value* of this monitoring

is twofold. First, emulsion polymerizations are complex processes and while much progress has been made in understanding them there are still many unknowns. Real-time monitoring allows the development of improved understanding of emulsion polymerizations, such as quantifying the effects of temperature, initiator type, initiator concentration, and other factors on the reaction dynamics. Second, accumulation of unreacted monomer in the reactor (due to factors such as fluctuations in polymerization rate, insufficient mixing, or poor control of feed rate or reactor temperature) can lead to runaway reactions. Real-time monitoring allows control of the process to be improved, reducing the safety risk associated with such a runaway reaction.

Chabot's work was done in a laboratory-scale batch reactor. Experiments at lab scale are a common first step in process analyzer work, since they allow technical feasibility to be demonstrated relatively quickly and inexpensively without having to interfere with production targets in a commercial-scale plant. In this work, moreover, one of the two business goals (improving process understanding) could be largely accomplished without going beyond lab-scale experimentation.

Several technical challenges were faced in this work. The monitoring had to be done in a complex, heterogeneous, unstable matrix, in the presence of high solid levels (~40% by weight), with large monomer droplets at the beginning of the reaction, very fine particles, and possible foaming. The matrix consisted of an aqueous mixture of seed latex particles, surfactant(s), chain transfer agent(s), one or more acrylic monomers (such as methyl-methacrylate, MMA), and initiator(s), under nitrogen or argon purge, at elevated temperatures of 65–75 °C. Furthermore, it was reasonably expected that an optical probe inserted into the reactor might become coated or plugged with polymer.

The emulsion polymerization process was monitored using an NIR analyzer (Foss NIRSystems Model 5000) equipped with a fiber-optic transflectance probe inserted directly into the reactor. The analyzer was calibrated for residual monomer level by withdrawing samples periodically during a run and analyzing them by the reference method, headspace GC. A high correlation (R^2 = 0.987) was found between the second-derivative absorbance at 1618 nm, attributed to the vinyl C–H group, and the residual monomer concentration as measured by GC. A single-wavelength multiple linear regression (MLR) model using the 1618 nm band in the second derivative predicted the residual monomer level with an accuracy (SEP) of 0.2% by weight, matching the accuracy of the primary GC method.

The technical challenges were dealt with as follows:

1. The use of a transflectance probe, rather than a transmission probe, allowed good quality spectra to be obtained despite the low light transmission due to the high solid levels.
2. Probe fouling was minimized when the probe was placed properly in the reactor relative to the liquid surface and to the mechanical stirrer. The ideal spot, although not explicitly stated, would presumably be deep enough into the emulsion to see a representative bulk sample and to avoid any 'dead' unstirred spots, but far enough from the stirrer blade to avoid impact.
3. The use of second-derivative spectra eliminated the baseline shifts observed during the reaction (which could be due to changes in the amounts or size distribution of particles, or small amounts of probe fouling).
4. It was observed that the NIR-predicted monomer concentration varied greatly during the first 45 minutes after monomer addition but then stabilized; this was attributed to slow solubilization of large monomer droplets into the water phase, and slow partition of monomer between the water phase and the latex phase, leading to large local variations in monomer concentration. This was avoided by allowing a 60-minute 'soak' period after monomer addition, before the temperature was increased to start the polymerization.
5. Finally, although temperature had a large effect on both the position (wavelength) and the intensity of the water absorption bands in the emulsion NIR spectra, careful experimentation demonstrated that the 1618 nm vinyl C–H band used in the calibration model did not shift in either position or intensity with

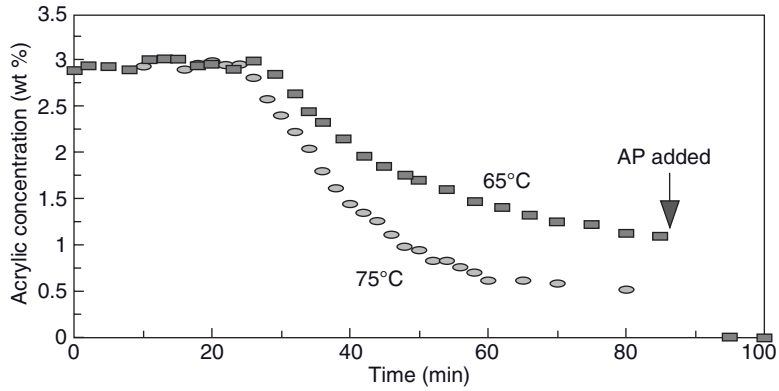


Figure 15.1 *Effect of temperature on the rate of emulsion polymerization as monitored by in-line NIR. Reprinted with permission from Chabot* et al. *(2000).*[4]

temperature, in the temperature range used in these studies (25–75 °C). Therefore, it was not necessary to correct the calibration model for temperature effects, either by the use of internal or external standards, or by including temperature variations in the calibration set.

Once calibrated, the NIR analyzer was used to investigate a number of factors expected to affect the polymerization kinetics, including reaction temperature, initiator type, and initiator concentration (relative to monomer concentration). These experiments, in addition to improving process understanding, also mimicked the effects of inadequate process control during a reaction. Figure 15.1 shows the effect of reaction temperature on kinetics. The reaction rate nearly doubles when the temperature is raised from 65 to 75 °C, and the concentration of unreacted monomer after 85 minutes is reduced from 1.1 to 0.5%. In-line NIR monitoring allows unusual behavior in either reaction rates or residual monomer levels to be detected and corrected immediately.

15.3.2 Monitoring a diethylbenzene isomer separation process

Chung *et al.*[5] at SK Corporation (Ulsan, Korea) and Hanyang University (Ansan, Korea) used off-line NIR to monitor the *p*-diethylbenzene (PDEB) separation process. The process consists of isolating PDEB from a stream of mixed diethylbenzene isomers and other aromatic hydrocarbons by selective adsorption on a solid adsorbent, followed by extraction with *p*-xylene (PX). Optimal control of the process requires accurate measurement of the concentrations of the *o*-, *m*-, and *p*-diethylbenzene isomers as well as the PX extractant. The business value of NIR monitoring lay primarily in reducing the analysis time from 40 minutes (by GC) to less than one minute, allowing the process to be monitored and controlled in (effectively) real time. There was some secondary benefit to replacing the GC method, which required constant attention to achieve reliable results, with the more rugged NIR method.

The challenges in this work concerned the fundamental limits of NIR spectroscopy. First, would NIR, with its typically broad and highly overlapped bands, have enough spectral resolution to distinguish the *ortho*, *meta*, and *para* isomers from each other and from the extractant? Second, would NIR, with its typically weak overtone and combination bands, have enough sensitivity to quantify the minor components of the stream, especially the *ortho* isomer, typically present at only ~1%?

The NIR monitoring was done off-line using a Foss NIRSystems Model 6500 analyzer in transmission mode and a quartz cuvette with 0.5-mm path length. The spectra had 10-nm wavelength resolution with data

Table 15.2 *PDEB separation calibration results*

Component	Concentration			
	Min	Max	MSECV (f)	SEP
o-Diethylbenzene	0.00	4.24	0.11 (6)	0.12
m-Diethylbenzene	0.04	56.27	0.15 (8)	0.16
p-Diethylbenzene	0.15	76.15	0.16 (7)	0.15
p-Xylene	0.47	99.38	0.28 (7)	0.27

Note: All concentrations are in volume %.

points taken every 2 nm. The analyzer was calibrated using 152 samples taken from a pilot plant simulated moving bed adsorption unit. The pilot plant was designed to operate under a wide range of conditions, which allowed the compositions of the 152 samples to be intentionally varied over a wide range as well. The composition of the 152 samples in the calibration sample set is shown in the first three columns of Table 15.2; the second and third columns of the table indicate the maximum and minimum concentrations (in volume %) found in the calibration sample set for the components listed in the first column.

The NIR spectra of the pure isomers and the extractant are shown in Figure 15.2. Although the spectra of the three isomers and the extractant are quite similar overall, there are distinct, reproducible spectral differences, especially in the 2100–2500 nm region. The authors did extensive studies involving mid-IR spectra, spectral simulations, and PCA in order to understand the origins of these differences to ensure that they were related to the isomers themselves and not to coincidental impurities.

Calibration models were developed using 106 spectra (70% of the total) chosen at random. The remaining 46 spectra were used to validate the models. The effects of number of PLS factors, wavelength range, and absorbance vs. second-derivative spectra were investigated. The most accurate models were obtained when the 2100–2500 nm region was included in the model, regardless of whether the rest of the spectrum was included – this makes sense since this region has the sharpest, most intense peaks. The last two columns of Table 15.2 show the results for PLS models developed using absorbance spectra in the 2100–2500 nm region only. Column 4 shows the accuracy of the model on the 106 calibration samples (MSECV: mean square error of cross-validation), with the number of PLS factors for the model in parentheses; column 5 shows the accuracy of the model on the 46 validation samples (SEP). The results are shown graphically in Figure 15.3.

The results indicate that NIR does indeed have both the resolution (specificity) and the sensitivity needed for this application. The NIR method was successfully implemented in the plant, replacing the conventional GC method and allowing real-time control and optimization of the PDEB separation process.

15.3.3 Monitoring the composition of copolymers and polymer blends in an extruder

McPeters and Williams[6] at Rohm and Haas (Spring House, PA, USA) used in-line transmission NIR to monitor the compositions of heterogeneous polymer blends and terpolymers in real time at the exit of an extruder. The business value of this monitoring comes from four sources. First, real-time monitoring minimizes waste production at start-up, since it is no longer necessary to wait several hours to get enough off-line lab data to demonstrate that the process has stabilized ('lined out'). Second, real-time monitoring allows ready determination of the extruder residence time, a key process parameter which is very tedious to determine otherwise. Third, real-time monitoring allows the detection and elimination of process 'cycling' or oscillation due to poorly tuned control loops; this greatly reduces process and product variability. Fourth,

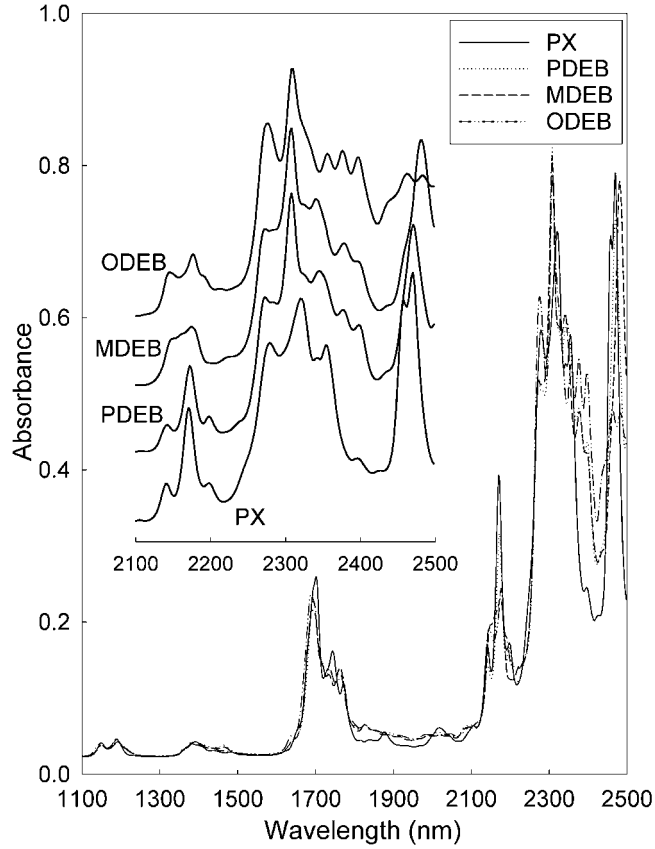


Figure 15.2 *NIR spectra of* o-*diethylbenzene (ODEB),* m-*diethylbenzene (MDEB), PDEB and PX. The 2100–2500 nm range is expanded in the upper left of the plot. Reprinted with permission from Chung* et al. *(2000).*[5]

real-time monitoring rapidly increases process understanding by enabling the effects (both the magnitude and the response time) of changes in process parameters such as zone temperatures, feed rates, vent rates, or screw speed to be observed immediately, and in situ.

The technical challenges in this work included high temperatures (~280–325 °C), high pressures (several hundred psig where 1 psig = 6.9 kPa), and NIR baseline fluctuations due to bubbles, particulates, impurities, or degraded polymer in the melt stream. The first two challenges were addressed using custom-made probes designed to withstand the simultaneous high temperatures and pressures. (Note that in addition to suitable probes, care must also be taken with the interface of the probes with the process, to ensure good optical alignment and reproducible path length, while avoiding leaks and other safety issues.) The baseline fluctuations were removed by linear baseline correction of the spectra, at 1290 nm and 1530 nm for the polymer blends, and at 1290 nm alone for the terpolymers.

The extruders were monitored using a pair of custom-made transmission probes inserted into the extruder die just downstream from the screws. Each probe consisted of a sapphire window brazed into a metal body; a quartz rod behind the window piped the light to the end of the probe; low-OH silica fiber-optic bundles connected the probe to the NIR analyzer, an LT Quantum 1200I. Optical path lengths were typically between 0.3 and 2.0 cm.

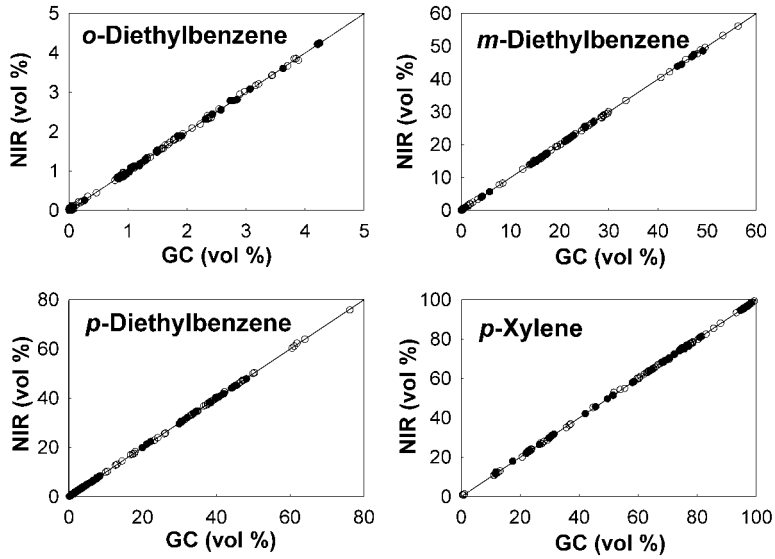


Figure 15.3 *Scatter plots showing the correlation between NIR (using the 2100–2500 nm region) and GC analyses for each component. Reprinted with permission from Chung* et al. *(2000).*[5]

The polymer blends were a heterogeneous mixture of an acrylate/styrene copolymer dispersed in a methacrylate/acrylate copolymer. The level of acrylate/styrene copolymer present, in wt%, is termed the 'blend level.' The NIR analyzer was calibrated over 2 days for blend level by obtaining spectra of seven different blends with blend levels ranging from 0 to 45%. The extruder temperature was 260 °C and the optical path length was 2.0 cm. The reference method in this case was the known weights used to prepare the blends. All spectra were used in calibration except the transition spectra. Calibration explored both MLR and PLS methods, on absorbance, first- and second-derivative spectra. Only the spectral region from 1200 to 1650 nm (in which the absorbances ranged between 0 and 1 AU) was used in modeling. The best model was an MLR model using two wavelengths, 1211 and 1595 nm, in the absorbance spectra; it had a correlation R^2 of 0.9986 and a standard error of 0.79%. Figure 15.4 shows the transition between the 10% blend and the 25% blend as predicted by the NIR model. Both the extruder dead time, which is about 5 minutes, and the extruder transition (or mixing) time, which is about 10 minutes, are clearly evident.

The terpolymers contained three components, one major and two minor. Component 1 (a methacrylate-related repeat unit) is the major component, with levels ranging between 60 and 90%. Component 2 (a methacrylate ester repeat unit) and component 3 (a hydroxyl-containing repeat unit) are the two minor components; component 3 levels are typically less than 1%. The NIR analyzer was calibrated for component levels using spectra corresponding to 18 pellet grab samples. The grab samples were analyzed for component 1 using elemental analysis, and for component 3 using titration. Again, only the spectral region from 1200 to 1650 nm (in which the absorbances ranged between 0 and 1 AU) was used in modeling. The best model was a three-factor PLS model using second-derivative spectra; it had a standard error of 0.98% for component 1 and 0.032% for component 3. Because of the strong sensitivity of NIR to hydroxyl functional groups, the spectral changes due to the small variations in the very low levels of component 3 are actually larger than those due to the much larger variations in component 1 and 2 levels. Thus the NIR model is more accurate in absolute terms for component 3 than for component 1. Figure 15.5 shows the ability of NIR to catch process cycling that occurs on a time scale too fast to be detected by grab sampling once or twice an

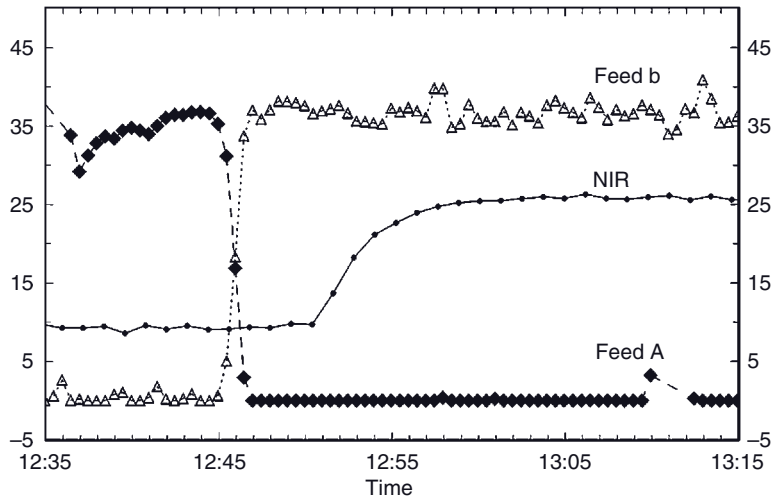


Figure 15.4 *Plot of the NIR predicted blend levels for the 10–25 wt% transition region overlaid with feed rates for the two blends. Reprinted with permission from McPeters* et al. *(1992).*[6]

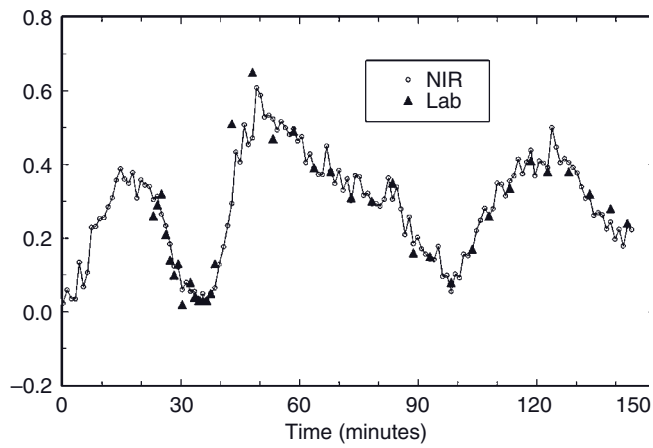


Figure 15.5 *In-line NIR predicted levels of component 3 (circles) and laboratory analysis for component 3 in pellets taken from the extruder process (triangles). Reprinted with permission from McPeters* et al. *(1992).*[6]

hour (at best). In this case, extra grab samples were taken, presumably to verify that the cycling was a real variation in polymer composition and not an NIR temperature or density artifact. The process cycling was traced to excessive cycling of the temperature controllers in one zone of the extruder.

15.3.4 Rapid identification of carpet face fiber

Rodgers[7] at Solutia (Gonzalez, FL, USA) used off-line diffuse reflectance NIR to rapidly and accurately identify the face fiber on carpets returned for recycling. The business value of this was twofold. It allowed rapid sorting of incoming carpets, an essential aspect of any recycling program. However, there were already

several commercial NIR-based instruments on the market for carpet sorting. The author developed an internal (to Solutia) system rather than using a commercially available system because of the limitations of many of the available systems from a fiber-producer's viewpoint, including the inability to include proprietary or unique samples in the calibration models, the inability to quickly update models to include new or modified products without returning the instrument to the vendor, the lack of qualitative information (to confirm identification, or to assess similarity between the evaluated sample and those used to develop the model), and the inability of many commercial systems to correctly identify carpets with widely varying color or carpet construction. Although it was not stated, the value of overcoming these limitations must have exceeded the cost of bringing method development, validation, and long-term maintenance in-house.

There were two classes of technical challenges faced in this work. The first challenge was how to prevent variability in carpet color (shade and darkness), carpet coloring method (dyes versus pigments), and yarn and carpet construction (e.g. cut versus loop pile) from interfering with accurate identification of the carpet face fiber. The second challenge was how to implement the NIR method at-line (using a fiber-optic probe that an operator can simply hold against the incoming carpet roll and minimizing the scan time) without affecting the identification accuracy. Both challenges were addressed by systematically exploring the effects of carpet and instrument factors.

The work was done using a large collection of carpet and fiber samples, containing nylon-6,6 (N66), nylon-6 (N6), polypropylene (PP), PET and acrylic polymer types. The samples in the collection varied widely in color, yarn and carpet construction, heat setting type, and dyeing/coloring methods. NIR diffuse reflectance spectra of the samples were obtained using a Foss NIRSystems Model 6500, operating in the 1100–2500 nm range only, with either the sample transport attachment ('static' mode, for laboratory analysis of carpet or fiber pieces) or a fiber-bundle interactance probe ('remote' mode, for at-line analysis of intact carpets). In the latter case, the probe was held directly against the fiber or carpet sample during scanning. Four spectra of each sample were obtained, repacking or rotating the sample, or moving the probe to a different location on the sample, between spectra. Foss NIRSystems software, either the DOS-based NSAS/IQ2 (Identify, Qualify, Quantify) program or the Windows-based Vision software, was used to create spectral libraries, develop calibrations, and test calibration performance. Wavelength correlation was used for identification of unknown samples, while wavelength distance was used for qualification. The results are briefly summarized here.

Polymer type: All of the polymers have very distinct, easily distinguished NIR spectra except N66 and N6 which are very similar. The use of second-derivative spectra is necessary to reliably distinguish the small spectral differences between N66 and N6.

Color effects: The effects of color were found to be specific to the fiber type and to the coloring method (dyes versus pigments). Color pigments caused large changes in absorbance across the entire spectrum, as shown in Figure 15.6. The use of second-derivative spectra removed the baseline shifts, but did not remove the variation in peak intensities. However, the positions of the characteristic absorption peaks for a given polymer were found not to vary significantly despite the color variations. This allowed accurate calibration models to be developed using second-derivative spectra as long as sufficient variation in carpet color was included in the calibration spectral library.

Construction effects: The effects of construction were less pronounced than the effects of color. In general, cut pile carpets had larger absorbances across the entire spectrum than did loop pile carpets, but the positions of the characteristic absorption peaks were not affected. As for color, accurate calibration models could be developed on second-derivative spectra as long as sufficient variation in carpet construction was included in the library.

Measurement speed effects: The effect of reducing the number of scans per spectrum, in order to increase measurement speed, was found to be minimal over the range from 20 scans down to 5 scans.

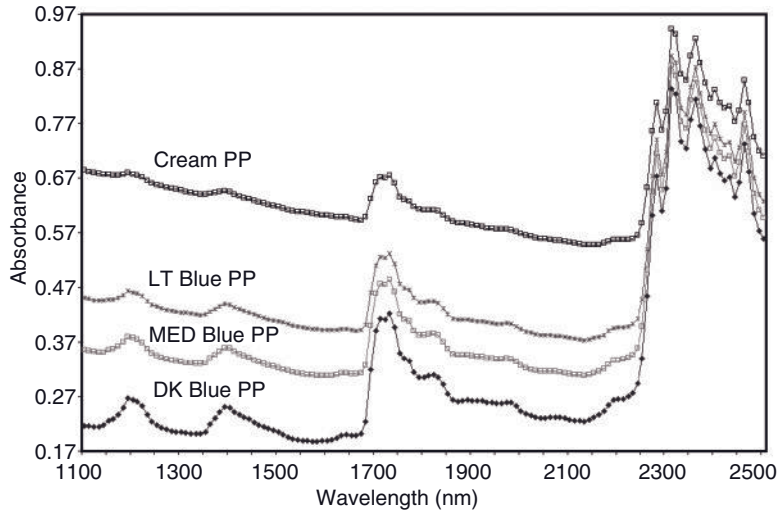


Figure 15.6 *Influence of solid pigments on the NIR absorbance spectra of colored polypropylene carpets. Note that as the carpets change from dark blue (bottom trace) to cream color (top trace) the baseline shift increases, the baseline tilt increases, and the intensity of the absorbance peaks decreases. Reprinted with permission from Rodgers (2002).*[7]

Table 15.3 *NIR identification of polymer type*

Sample type	Polymer type	Sources	Samples	% correct (absorbance model)	% correct (second-derivative model)
Fiber	N66	20	146	78	100
Fiber	N6	3	15	73	100
Carpet	PP	2	11	82	100
Carpet	N6	1	8	0	100
Carpet	N66	1	8	0	100

Sampling mode effects: The goal was to be able to measure at-line, with no sample preparation at all, using a fiber-optic probe or a remote sampling head. However, the area sampled by the fiber-optic probe is much smaller than for the sample transport module. It was found that the 'remote' (probe) spectra were very similar to the 'static' (sample transport) spectra, but the baselines were shifted significantly higher and the absorbance peaks consequently reduced in intensity; as before, the characteristic peak positions were not affected. Calibration models developed using spectra obtained with the fiber-optic probe performed equivalently to those developed with the sample transport module.

The performance of the NIR analyzer in 'static' mode is shown in Table 15.3. The use of second-derivative spectra in developing the identification method, along with the inclusion of sufficient diversity in color and construction, allowed for 100% accurate identification of unknown samples.

15.3.5 Monitoring the composition of spinning solution

Sollinger and Voges[8] at Akzo Nobel (Obernburg, Germany) used off-line transmission NIR to monitor the key quality parameters in cellulose fiber (viscose) spinning solutions over time. The business value lay both

in reducing the time and cost of analysis by consolidating four or five different lab methods into one NIR method, and in providing more timely information to the process, enabling tighter control of product quality, higher yields, and reduced waste.

The spinning solution consists of cellulose solubilized as cellulose xanthogenate in an aqueous solution of sodium hydroxide (NaOH) and carbon disulfide (CS_2). The spinning solution is extruded through nozzles into an acidic bath, regenerating cellulose as filaments. The key quality parameters for the spinning solution are total cellulose, xanthogenate content or 'ripeness' (number of CS_2 residues per 100 glucose units of the cellulose molecule), NaOH concentration, and trithiocarbonic acid (TTC) concentration.

There were many challenges involved in this work.

1. The spinning solutions are not stable over time, since CS_2 continues to react with the cellulose, increasing the xanthogenate content or 'ripeness' (which is why monitoring is needed in the first place). This was dealt with by carrying out the NIR and reference methods simultaneously, where possible, or by storing the samples at ~30 °C (where they are stable) until the reference analysis could be performed.
2. There are inherent intercorrelations among many of the quality parameters due to the chemistry. This was minimized in two ways. First, the calibration sample set was created by systematic variation of the total cellulose content (6–11 wt%), the total NaOH content (4–7 wt%), and the total CS_2 content (16–32 wt%) using a central composite design, along with repeated measurement of these solutions after different ripening periods. However, this did not completely remove all intercorrelations (such as between TTC and Na_2CO_3 levels). The remaining intercorrelations were minimized by careful choice of spectral regions to include in the calibration models, to ensure that only the parameter of interest affected the model.
3. The calibration models needed to be insensitive to variations in the levels of minor additives (<1 wt%) or process impurities. This was achieved by including samples from different production lines, containing varying concentrations (including zero) of different spinning additives, in the designed experiment sample set.

The spinning solution composition was measured off-line using a Foss NIRSystems Model 6500 VIS-NIR analyzer (wavelength range 400–2500 nm), equipped with a thermostatted cuvette holder (30 °C) and a cuvette of 1-mm path length. The analyzer was calibrated using the designed calibration set described above, containing 54 samples, using CAMO's Unscrambler software. First-derivative spectra were used to eliminate effects due to baseline shifts. The resulting PLS1 models were validated using a separate set of samples (which also included production line samples). The results are summarized in Table 15.4 (RMSEC: root

Table 15.4 *Multivariate calibration results for viscose spinning solutions*

Parameter	Units	Concentration range	Wavelength range (nm)	Number of PLS1 factors	Regression coefficient	RMSEC	RMSEP
Cellulose	wt%	6.0–11.0	2200–2400	5	0.993	0.119	0.131
γ number	mg/100 g	25–54	1120–1250, 1600–1800, 2200–2350	5	0.983	2.0	1.5
TTC	mg/100 g	240–1350	510–590	1	0.998	17.5	22.7
NaOH	mg/100 g	2400–5500	1350–1500, 1800–2100, 2300–2450	1	0.979	131	109

mean square error of calibration; RMSEP: root mean square error of prediction [validation]). The accuracy of the NIR predictions, expressed relative to the means of the respective calibration ranges, were between 2.3 and 3.3% for all parameters, which was considered acceptable.

15.3.6 Monitoring end groups and viscosity in polyester melts

Brearley *et al.*[9] at DuPont (Wilmington, DE, USA) used in-line transmission NIR to monitor carboxyl end groups and degree of polymerization (DP) in PET oligomer and prepolymer melt streams in a new polyester process. The business value was derived from several sources.

First and most importantly, real-time NIR monitoring enabled real-time control of the process. For a given product, the molecular weight and end-group balance in the prepolymer exiting the 'front end' or melt part of the process must be controlled at specified levels in order for the 'back end' or solid-phase part of the process to successfully produce the intended polymer composition. In addition, the variability in prepolymer composition must be controlled with very tight tolerances to keep the variation in final product composition within specification limits. Since the process dynamics in the front end were more rapid than those in conventional PET processes, the conventional analytical approach involving off-line analysis of samples obtained every 2–4 hours was not sufficient to achieve the desired product quality.

Second, real-time monitoring enabled particularly rapid development of process understanding, by providing otherwise-unattainable information on process dynamics and by drastically reducing the time needed to carry out designed experiments (since it was no longer necessary to remain at a given process 'state' for several hours until several lab results indicated that the process was lined out).

Finally, real-time NIR monitoring, once validated to the satisfaction of the process engineers and operators, significantly reduced the need for hot process sampling (with its attendant safety concerns) and lab analysis by allowing the sampling frequency to be greatly reduced (to near zero for some process points).

The PET melt composition was monitored using a Guided Wave Model 300P NIR analyzer, equipped with fiber-optic cable, transmission probes of 0.25-inch (0.6 cm) diameter, and DOS-based Scanner 300 software. The analyzer was located in an air-conditioned control room, with fiber optics running out to the process points. The transmission probes were inserted into the polymer transfer lines using custom-designed sapphire-windowed stainless-steel 'sleeves' to provide protection from process pressures. Custom software macros were written to scan the appropriate probes with the desired timing, apply the relevant models, and communicate the results to the plant DCS through a MODBUS interface. Calibration model development was done using CAMO's Unscrambler software. There were a number of technical challenges involved in this work:

1. An initial challenge was the desire to use NIR to control and optimize the process immediately upon plant start-up, which required the analyzer to be calibrated ahead of time in the absence of any process samples. The approach used in this case was to implement 'provisional' models, which were simply the absorbance at a characteristic wavelength multiplied by 1000. The provisional models were expected to be neither accurate nor precise, but merely to track the level of the functional group of interest. The characteristic wavelengths were determined based on where the functional group of interest absorbed in off-line NIR spectra of PET flake samples. Peak assignments were based on the NIR literature on polyesters, extrapolation from vibrational fundamentals in the mid-IR literature on polyesters, spectra of model compounds, and observed variations in the spectra with known changes in sample composition. The provisional model wavelengths, after baseline correction at 1268 nm, were 1590 nm for carboxyl ends, 1416 nm for hydroxyl ends, and 1902 nm for moisture. The provisional models were used for the first several weeks after plant start-up, until 'real' models could be developed, and worked remarkably well. They tracked the process behavior and the lab results, revealed process oscillations

due to imperfectly tuned control loops, and allowed designed experiments to be completed in minimal time.

2. A second challenge was to rapidly develop 'real' models as soon as possible after plant start-up, in spite of the relatively small variation in composition expected at a given process point. This challenge was solved by modelling all three process points (esterifier exit, pipeline reactor exit, and prepolymerizer exit) together. Inclusion of the process variability due to the start-up itself also helped. The calibration space for the first calibration set is shown in Figure 15.7.
3. A third challenge was the imprecision of the reference lab values for carboxyl ends and for DP. This was due both to lab method variability (with an occasional outright mistake) and to sampling variability. Sampling variability was due to time assignment error (labelling the sample to the nearest quarter-hour introduces significant errors during transitions), sample inhomogeneity (NIR sees a large volume of sample; the lab measurements are conducted on a very small sample), and continued reaction and/or moisture pickup/loss by hot samples as they are transported to the lab. This was dealt with in two ways: by using large numbers of samples (hundreds) in the calibration sample sets in order to average out the effects of any given lab number and by careful iterative exclusion of outliers during modelling.
4. A fourth challenge was fouling of the probes. The effect of fouling on the NIR spectra is shown in Figure 15.8. Low levels of fouling increase the baseline absorbance; this was easily remedied by baseline correction of the spectra. At higher fouling levels, however, the baseline begins to tilt and curve as well. These effects are very difficult to correct for or model out. It was found through experimentation that the steps listed below allowed development of NIR models that predicted polymer composition accurately despite fouling levels up to 2.3 AU at 1274 nm.
   Step 1 Include fouled spectra in the calibration data set.
   Step 2 Use raw spectra. (Do not use baseline-corrected, derivatized or pretreated spectra.)

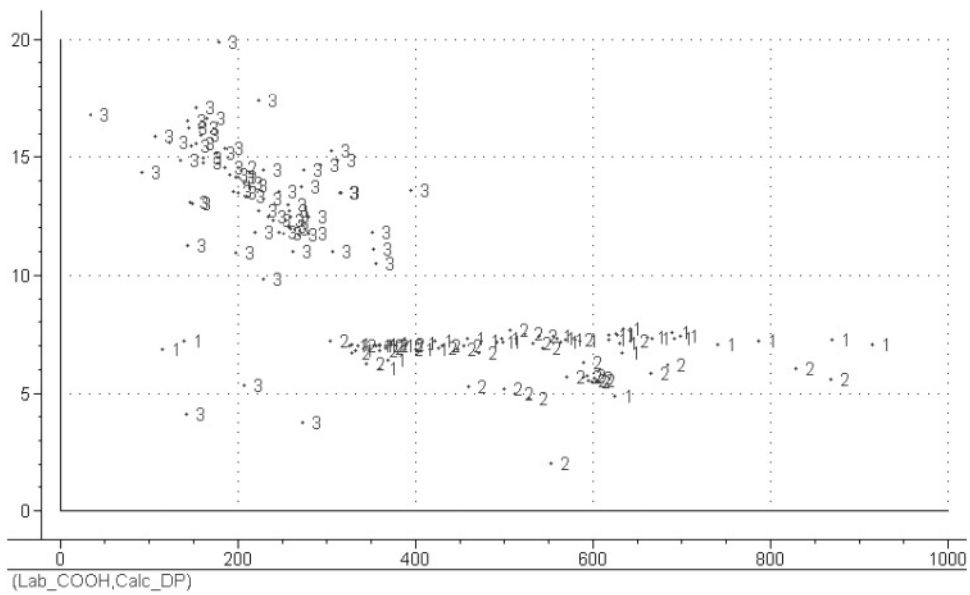


Figure 15.7 *Calibration space covered by the first calibration set. The x-axis is lab carboxyl ends in meq/kg and the y-axis is DP in repeat units. The samples are labelled by process point: 1 – esterifier, 2 – pipeline reactor and 3 – prepolymerizer. Reprinted with permission from Brearley and Hernandez (2000).*[9]

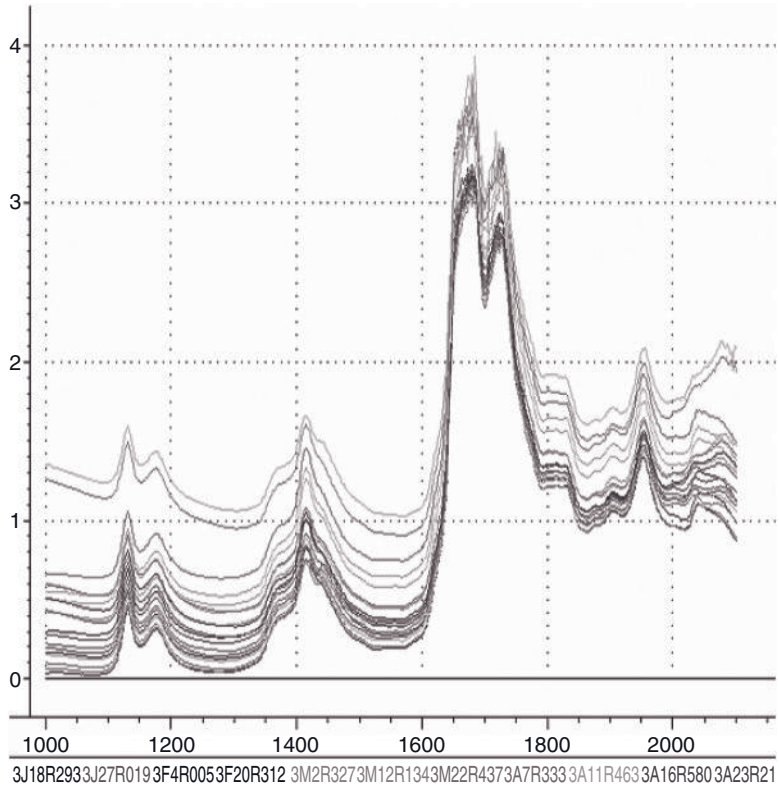


Figure 15.8 *NIR transmission spectra of PET prepolymer melt. The x-axis is wavelength in nanometers and the y-axis is absorbance. Reprinted with permission from Brearley and Hernandez (2000).*[9]

Step 3 Exclude the short wavelength edge of the spectra (e.g. wavelengths less than 1300 nm) since they are most affected by tilt and curvature.

Step 4 Exclude the long wavelength edge of the spectra (e.g. wavelengths above 2000 nm) since they become very noisy at high fouling levels.

Calibration models were developed using process grab samples. Each sample set was split in half randomly to give independent calibration and validation sample sets. The results for the best models are shown in Table 15.5. The real-time monitoring results as seen by the process engineers and operators are shown in Figure 15.9.

15.3.7 In-line monitoring of a copolymerization reaction

Beyers *et al.*[10] in the Polymer Research Division of BASF-AG used in-line transflectance NIR to monitor methyl methacrylate (MMA) and *N*,*N*-dimethylacrylamide (DMAAm) monomers in a copolymerization reaction. The work in this paper is of interest as it illustrates an example of calibration development done off-line with a very limited number of prepared calibration samples. The value of the measurement is to control the end properties of the products resulting from the copolymerization reaction. The end properties are related to many parameters including the intramolecular chemical composition distribution (CCD). The

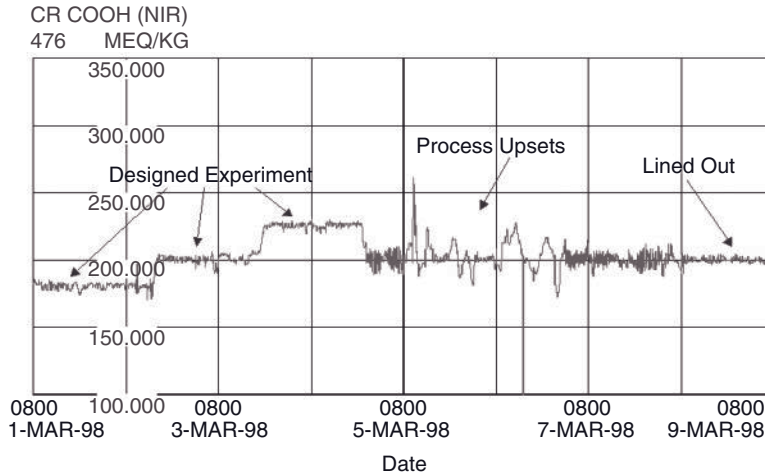


Figure 15.9 *Predictions of the NIR model for prepolymer carboxyl ends over an eight-day period. Three states of a designed experiment, as well as a period of process upset and the return to lined-out operation, are indicated. Reprinted with permission from Brearley and Hernandez (2000).*[9]

Table 15.5 *In-line monitoring results for polyester process*

Parameter	Model name	Range	Number of PLS factors	R_{val}	Accuracy (SEP)	Precision*
Esterifier: COOH ends	Ca4C3	100–700 meq/kg	3	0.9771	34 meq/kg	2.8 meq/kg
Prepolymerizer: COOH ends	Car5F5	80–460 meq/kg	5	0.9317	13.5 meq/kg	0.8 meq/kg
Prepolymerizer: DP	DP3B3	10–26 units	3	0.9935	0.54 units	0.03 units

R_{val} – correlation coefficient for the validation sample set.
SEP – standard error of prediction for the validation sample set.
* Precision is expressed as the standard deviation of the measurement over 30–60 minutes during a period when the process is lined out.

article states that the ability to control the CCD is directly dependent on the ability to control the instantaneous monomer concentrations in the reactor.

To control the CCD, the authors chose a closed-loop control strategy with in-line NIR used to measure the monomer concentrations. The closed-loop control strategy uses the NIR in-line measurement to provide close to real-time information about the concentration of both reactants and products in the reactor. One prerequisite of this control strategy is that the sampling rate be faster than the kinetics of the reaction.

Monitoring the MMA/DMAAm reaction is challenging because both monomers have very similar NIR spectra, and because other interfering substances are present in the reaction mixture. The authors prepared calibration samples gravimetrically and made the NIR measurements at reaction temperatures. The calibration sets consisted of only five or six samples, which is considerably fewer than standard recommendations for NIR chemometric model development. The final models for MMA and DMAAm were validated with an internal validation set as well as an external reaction validation. The performance of the models is summarized in Table 15.6. This table includes a measurement of the standard deviation of an external GC method

Table 15.6 *NIR calibration model performance for MMA and DMAAm in a reaction mixture*

Monomer	RMSEC (mol/L)	RMSEP (mol/L) internal validation	RMSEP (mol/L) external validation	GC STD (mol/L)
MMA	0.026	0.026	0.045	0.034
DMAAm	0.012	0.041	0.061	0.050

that was previously used to monitor the monomer concentrations. The RMSEP values for both internal and external validation sets compare well with the error in the laboratory GC method.

This paper by Beyers *et al.*[10] demonstrates that a thorough knowledge of the spectroscopy and chemometrics involved can often result in a simple solution to a complex problem.

References

1. P. Williams and K. Norris (eds), *Near-Infrared Technology in the Agricultural and Food Industries*, 2nd edition, American Association of Cereal Chemists, St Paul MN, 2001.
2. H.W. Siesler, Y. Ozaki, S. Kawata and H.M. Heise (eds), *Near-Infrared Spectroscopy*, Wiley-VCH, Weinheim, 2002.
3. D.A. Burns and E.W. Ciurczak (eds), *Handbook of Near-Infrared Analysis*, 2nd edition, Marcel Dekker, New York, 2001.
4. P. Chabot, L. Hedhli and C. Olmstead, On-line NIR monitoring of emulsion polymerization, *AT-Process*, **5**(1, 2), 1–6 (2000).
5. H. Chung, M.S. Ku, J. Lee and J. Choo, Near-infrared spectroscopy for monitoring the *p*-diethylbenzene separation process, *Appl. Spectrosc.*, **54**(5), 715–720, (2000).
6. H.L. McPeters and S.O. Williams, In-line monitoring of polymer processes by near-infrared spectroscopy, *Process Contr. Qual.*, **3**, 75–83 (1992).
7. J.E. Rodgers, Influences of carpet and instrumental parameters on the identification of carpet face fiber by NIR, *AATCC Review*, **2**(6), 27–32 (2002).
8. S. Sollinger and M. Voges, Simultaneous determination of the main constituents of viscose spinning solutions by visible near infrared spectroscopy, *J. Near Infrared Spectrosc.*, **5**, 135–148 (1997).
9. A.M. Brearley and R.T. Hernandez, *In-line monitoring of polymer melt composition; oral presentation at the Eastern Analytical Symposium*, Somerset, NJ, 2 November 2000.
10. C.P. Beyers, H.F.M. Boelens, L. Klumperman and J.A. Westerhuis, In-line reaction monitoring of a methyl methacrylate and *N*,*N*-dimethylacrylamide copolymerization reaction using near-infrared spectroscopy, *Appl. Spectrosc.*, **58**(7), 863–869 (2004).

16

Future Trends for PAT for Increased Process Understanding and Growing Applications in Biomanufacturing

Katherine A. Bakeev[1] and Jose C. Menezes[2]

[1] *CAMO Software Inc., One Woodridge Center, Woodbridge, NJ, USA*
[2] *Institute for Biotechnology and Bioengineering, IST, Technical University of Lisbon, Lisbon, Portugal*

16.1 Introduction

The field of process analytical technology (PAT) continues to evolve, driven by advances in instrumentation, data analysis tools, and regulatory agencies such as the US Food and Drug Administration (FDA) and the European Medicines Agency (EMA). This field, which when it first emerged was known as process analytical chemistry (PAC), is now almost exclusively referred to as process analytical technology (PAT) within the pharmaceutical industry. Regulatory bodies around the world are working to develop guidance to assist the regulated pharmaceutical industry to adopt PAT and reap the benefits of this as they work towards designing products and processes with quality embedded, and not tested post-process.[1–4] Historically PAT was considered as a means of having rapid measurements on a process to provide continuous quality assurance with more rapid and frequent measurements. It has been adopted in many industries and brought advantages which include: a reduction in the need to remove samples from a process to measure their properties, the ability to have the measurements tied into a control system to enable a logging of a process, and the possibility of real-time process control.

The use of PAT has expanded to industries beyond the traditional petrochemical, semiconductor, food and chemical industries to pharmaceutical, biochemical, paper and pulp and others who can benefit from the continuous quality assurance from on-line analysis, reduced sampling, and potentially reduced cycle times. How it is being used continues to develop with a greater emphasis put on implementing PAT from the early steps of development to enable increased process understanding and process control. One then can design products and processes with a greater understanding of the multivariate nature of the processes, the

Process Analytical Technology 2e Edited by Katherine Bakeev
© 2010 John Wiley & Sons, Ltd.

different sources of variability in the process, and how they can be controlled to produce consistent quality products. The process control can come from the combination of better process understanding as well as the ability to combine real-time measurements for feedback or feed forward control. The use of PAT from the development stage through manufacturing is also well aligned with recent regulatory guidance and with the principles of green chemistry.[5] In all industries PAT will be used only where it has a clear justification in terms of business and/or increased safety assurance, and if can provide the level of control needed for a given process. The most effective application of PAT on a pharmaceutical process (both primary and secondary) is an area which is still being defined and understood. PAT is not a direct replacement for end-product testing and often finds itself best applied (as the name indicates) at monitoring and controlling unit operations within the process, rather than direct measurement of specific attributes. A contributory factor to this is that the specifications for pharmaceutical products comprise mainly of minor (or low level) impurities, and the current sensitivity of PAT techniques is inadequate.

Instrument developments in terms of new technologies are adding to the potential of PAT, as are changes in terms of how PAT is viewed and implemented with a greater sophistication, reaching beyond real-time measurement and process monitoring, to serving as a key tool in process development and in process control. The previous chapters in this book have provided information about many of the technologies used for PAT, with numerous examples spanning many industries. In this chapter the ramifications of the quality initiatives on PAT will be discussed. Another area with increasing importance within PAT is multivariate data analysis (MVDA) and its growing use including multivariate statistical process control (MSPC), which can provide a holistic view of a process throughout various stages, with data from multiple sources. An industry that is greatly expanding its use of PAT is biomanufacturing and biofuels, and will be discussed in detail below. Emerging technologies with growing applicability will also be presented, along with the continuing challenges that lay ahead in the implementation of PAT.

PAT is maturing as new concepts to the pharmaceutical industry such as quality by design (QbD) are being clarified. There is still a long way to go to create common practices and definitions (e.g., how to build, define, and describe a design space) but this is an area of increased activity with promise to improve manufacturing efficiencies and product quality in all industries. A thorough discussion of how QbD can be used to ensure pharmaceutical quality is presented by Yu.[6]

16.2 Regulatory Guidance and its Impact on PAT

Guidance in terms of how to develop calibration models used in process analytical measurements based on near-infrared spectral data (ASTM E1655-05, ASTM E1790-04),[7,8] as well as on the validation of process analyzers (ASTM D6122-09, D3764-09)[9,10] have been in place for some time, derived from work over the past few decades, mostly in the petrochemical industry. These standards cover the development of calibration models from multivariate spectroscopic data, to the validation of analyzers. Within these standards, methodical, scientific approaches (including statistical analysis) for developing methods (measurements and calibration models), and in verifying and continually assuring performance of analyzers are presented, including risk assessment of the criticality of a measurement to define the level of validation necessary. They illustrate that PAT system implementation has evolved in a manner that supports process understanding, monitoring, and verification, as well as measurement of system performance. Additional codification of PAT use and implementation in terms of instrumental methods, as well as process and model validation specific to the pharmaceutical industry has been guided from the ASTM E55 committee (Manufacture of Pharmaceutical Products), whose membership has strong representation from the pharmaceutical industry.[11]

Guidelines specific to the pharmaceutical industry are developed by the International Conference on Harmonisation (ICH), which in recent years has introduced concepts and language that clearly underscore the value of PAT in advancing process understanding and providing a means for companies to mitigate risk in process and product development by thorough scientific experimentation and knowledge management. These documents give a structural backbone to the development of products and processes using tools of quality management that will encourage the integration of designed experiments to provide a multivariate understanding of the interactions of inputs (raw materials, formulation, process settings) and outputs (product attributes) such that the characteristics and performance of a product and/or process can be understood, along with the variability in the product/process. Sensor technologies, part of PAT, may also play a role in the development of a product/process understanding by providing in situ measurements which may be used to monitor a process from the development stages through to manufacturing, where controls can then be adopted.

The ICH Q8 guidance for pharmaceutical development introduces the term design space (DS), referring to a multivariate region of process operation wherein the parameter ranges (material attributes and process settings) are known to repeatedly result in products of assured quality. This DS is defined through a multivariate analysis of the variables that influence the product properties, and hence includes not just the parameters themselves, but the interaction between them that results in product that meets the standards for a safe, and effective product that is suitable for human use. This approach to process understanding is well aligned with the use of mechanistic models, design of experiments (DOE) and the use of PAT, all of which can deliver such results.

ICH Q8 also discusses the concept of QbD, which has its roots in the early quality and Six Sigma for manufacturing work of Juran and colleagues.[12]

The US FDA has put forth a quality by design initiative based on using scientific knowledge and understanding to develop well-understood and controlled processes. A thorough discussion of how QbD can be implemented in the pharmaceutical industry, focusing on the example of a generic drug is provided by Yu.[6] The key elements in QbD include the following activities: define the target product properties, design the product and process using scientific methodologies including design of experiments, and identify the critical product and process variables, and the sources of variation. With this information, the manufacturing process can be controlled and produce consistent quality. Examples of how QbD has successfully been used in the semiconductor industry have been presented. They illustrate how using chemical engineering principles is aligned with support of QbD in providing process understanding, and can be used also in the pharmaceutical industry, bringing business value.[13]

The QbD concepts parallel those put forth by Juran in response to the quality crisis of the 1980s that forced many Western companies to reevaluate their planning and development processes to plan for quality and success in business.

The concept of design space when it first emerged appears to have brought on a level of confusion as it challenged the familiar ways of working to define the operating ranges of a process. A multicompany committee working through the auspices of the International Society of Pharmaceutical Engineers (ISPE) has presented guidelines for defining a design space.[14] They discuss engineering and technical design processes that have been used in many types of industries, wherein risk-based analysis is used to determine design constraints and then determine appropriate controls. This underscores the necessity to draw from a multidisciplinary team in development that includes chemical engineers, and those experienced in PAT in addition to those with expertise in product development and formulation. Further engineering education emphasizing the use of science and technology in the pharmaceutical industry is proposed by Gernaey *et al.*[15] to provide a stronger link between PAT and scientific understanding based on first principles. The DS is defined in an iterative manner as additional knowledge is gained, but the initial definition of the DS for a product begins

during product development, and is linked to criticality through risk assessments performed on the product profile.

There are many tools and sources of information that can be beneficial in the definition of the DS for a product. Lepore and Spavins[14] suggest the following steps in defining a DS.

- use first principles if they adequately predict product performance
- understand the parameters that influence a process and make it scale-sensitive so that they can be controlled, and the scale-sensitivity be designed out
- use DOE to understand the interactions between multiple variables and to aid in designing their appropriate operating ranges
- analyze and present data in terms which provide information and knowledge (raw data itself is not knowledge)
- do not presume that the DS is equivalent to the proven acceptable range (PAR) for a process, as it involves the multivariate assessment of parameters.

Peterson[16] has presented a Bayesian approach to defining the DS, which provides design space reliability as well, as it takes into account both model parameter uncertainty and the correlation structure of data. To aid in use of this approach, he proposes a means of organizing information about the process in a sortable spreadsheet, which can be used by manufacturing engineers to aid them in making informed process changes as needed, and continue to operate in the DS.

The ICH Q9 (Quality Risk Management) and Q10 (Pharmaceutical Quality System) describe tools to aid in risk management and support effective quality management systems. It is anticipated that by following these guidelines the pharmaceutical industry will be able to demonstrate a thorough understanding of its processes, and hence be able to employ PAT tools to achieve real-time release (RTR) of products. This adoption of PAT, which is implemented for sound scientific grounds, can lead to large business benefits in terms of removing delays in manufacturing as intermediate products can be moved through the different stages of a manufacturing process based on real-time measurements that assure the product quality. If one can practice RTR, intermediate products can be used in the next manufacturing step without the requirement of waiting for off-line testing results. This concept supports that consistent quality is part of a well-understood process and therefore testing at the end of manufacturing may not be necessary, or add value. PAT is also aligned with green chemistry principles as its use can enable the reduction of manufacturing waste.[5] In fact, there is a draft guidance on continual process validation[4] that essentially states that in processes in which quality is ensured by proper design any produced batch can and should be considered a validation batch (as opposed to the current practice of three-batch process validation for regulatory approval of a process). It will be interesting to see how this draft evolves and after publishing how it will be interpreted by the industry and used in practice.

16.3 Going Beyond Process Analyzers Towards Solutions

Clearly there are many techniques that are used for process analysis, with developments continuing to expand the choices available. Just in this volume one observes that applications for analytical methods that have not been as widely adopted for process analysis including NMR spectroscopy, acoustic chemometrics and fluorescence are growing. An area that has the potential to bring greater value in quality, product development, and manufacturing efficiencies, is the application of multivariate data analysis to the data available from many stages of development and manufacturing, to improve process understanding. The best solution

for PAT is not to have sensors to measure every property at every stage of manufacturing. Not only would this be cost-prohibitive, but it would reflect that the process is not well understood, and therefore measurements are required throughout to verify the process health and the product quality. Taking a holistic approach to the design of the product and process can provide the understanding of the multivariate interaction of product and process inputs and outputs, and of the system variability. With this information, a control strategy can be developed and used to run a process under conditions that produce consistent quality.[17] Such an approach has been adopted in other industries, but closing the loop in terms of being able to define the attributes of raw materials to ensure product quality through the many stages of production requires a more thorough understanding of interactions along each step of the way to product.

PAT for process control has been realized in some industries, but that still needs to be achieved in the pharmaceutical and biotechnology industries. The lag within these regulated industries can be attributed to the late start that they have had in adoption of PAT relative to other industries, as well as the complexity of working in a regulated environment where process changes require a preapproval of the process with regulatory agencies. An example of the use of NIR spectroscopy for process control within the pharmaceutical industry has been published by scientists from Pfizer.[18] The application entails a PLS model of NIR spectral data to measure the water content in a continuous conversion reactor. Results are transmitted to the process distributed control system (DCS), and water adjustments to the process made as needed to keep the water between the necessary limits of 1.1 to 1.5%. By use of NIR, measurement frequency was increased to 200s, from hourly, giving a more stable process and greater product recovery, thus resulting in an estimated $500 000 cost savings during a six-month manufacturing campaign.

In an article which provides the historical background of the development of chemometrics and its utility, Munck[19] advocates a holistic approach to fully understanding a process from instrumental data and a full analysis of it. One must not overlook the necessity to revisit the raw data after modeling, to check if the data contains information that may have been lost by the data compression of the chemometric analysis. Part of the scientific understanding requires a visual inspection of the actual spectral data to see which regions carry the information that one is intending to model. Munck[19] illustrates how through an analysis of the near-infrared spectra of barley endosperm a fuller understanding of the genetic theory was obtained; a concept that is universal in developing process understanding. The importance of combining and transferring the fundamental chemical knowledge when developing chemometric models is stressed. What is required for solving problems through cheomometric modeling is quality data, mathematical models and knowledge of the chemical/process facts, theories and concepts. Munck's work is an example of how one can take the (spectral) data from a system or process to develop improved theories of science, which is an objective of the FDA's PAT initiative.

Another view of how PAT provides a holistic understanding of a manufacturing process is presented by Wold *et al.*[20] who describe four levels of PAT describing the chemometrics analysis of data and the complexity and objective of measurements.

The four levels of PAT can be defined as:

1. Off-line chemical analysis in labs to a large extent substituted by at-line analysis – e.g., spectroscopy, fast chromatography, imaging, and sensor arrays. Multivariate calibration used to convert PAT data to traditional space (e.g., concentrations, disintegration rate).
2. At-line PAT: data used directly without conversion back to the original variables space – e.g., MVI (Multivariate Identification), MSPC (multivariate statistical process control).
3. On-line PAT: data measured during batch evolution of each production step – e.g., on-line MSPC, BSPC (batch statistical process control).
4. Overall on-line PAT: Putting data from all steps and raw materials together for a total view of the process, which involves information from levels 1–3.

How one chooses to use PAT should be related to the levels of risk that one must address to keep a process in control. Wold[20] presents a breakdown of the risk involved in pharmaceutical manufacturing as:

- total risk
- chemical risks
- physical risks
- biological risks
- engineering risks
- analytical risk
- data analytical risk
- interpretation risk
- communication risk.

The perception and management of these risks is another aspect of PAT that must be addressed. The last three given above, included for completeness by Wold,[20] are not directly related to use of PAT, but more to how it is implemented. Indeed the communication risk is defined in how different parts of an organization see things, which requires that time be taken to define and agree project objectives in order to be able to define when a project is successfully completed, a point also addressed in Chapter 15 of this book. Risk management in the pharmaceutical industry is addressed in the ICH Q9 guidleine.[3] An additional dimension of this is the risk of acceptance of chemometric models, reflecting the lack of understanding of their development and use amongst personnel more familiar with the traditional analytical testing done on finished products.

It is vitally important that the multivariate nature of data related to a process be assessed to develop an understanding of a process and to assess quality. Process data together with appropriate chemometric models can provide information about: (1) product quality inferentially from process conditions; (2) process consistency (process signature, statistical process control); (3) analyzer reliability and (4) operational knowledge that can aid in scale-up and process transfers.[17]

16.3.1 Design of experiments for risk-based analysis

The development cycle of any process or product should begin with experimental design (DOE), at which time data are gathered and analyzed to add to process and product knowledge. DOE provides a risk-based analysis of the factors and responses that are used in defining critical to quality attributes (CQAs). An iterative cycle, as recommended for defining the design space, holds true in development of PAT solutions as well – going from design–collect–analyze–control. This iterative cycle closely matches the plan–do–check–act cycle, which is defined for continuous process improvement as seen in Figure 16.1.

Many aspects should be considered when using DOE tools for QbD studies. Most people familiar with DOE feel like going into QbD in one step and not in a stepwise manner which combines existing knowledge and the integration of mechanistic and empirical (DOE) models. However, defining a design space and eventually filing a QbD application should start with gaining process knowledge (and multivariate data analysis of the existing information is a key component) and end with a suitable control strategy to ensure operation within the design space. That is often neglected in classical DOE studies. Most DOE practitioners will go into defining DOE factors and jump to response surface modeling, without proper consideration of colinearity among factors (non-orthogonal factors) and many other interactions that can be handled by incorporating process knowledge in the analysis. DOE experts may often overlook that the main PAT objectives are hidden in the DOE workflow (Figure 16.2). Two other important challenges in using DOE concepts for QbD applications are (1) how to handle interactions between responses (CQAs) – something that is

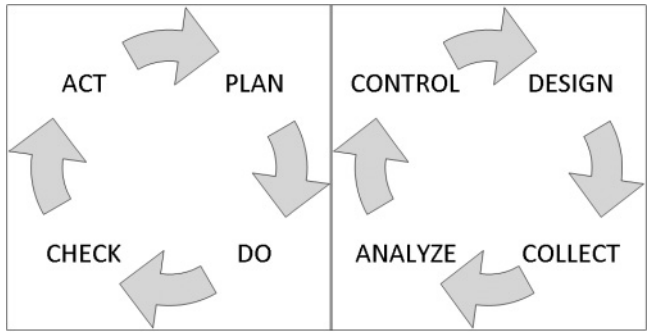


Figure 16.1 *Schematic of iterative problem solving process as given by the quality principles.*

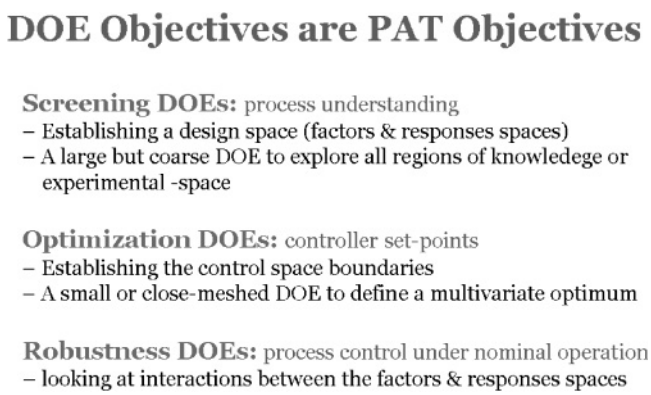


Figure 16.2 *DOE objectives and practice can be seen as the main PAT objectives.*

intrinsically done by most MVDA techniques common in the PAT toolbox, and (2) how to incorporate historical or process development data from several process scales. Here again there is the need for a good basis of process knowledge and how to define the governing phenomena in a dimensionless way. MVDA can guide and be of great usefulness in this exercise. Peterson's Bayesian approach[16] provides a framework for incorporation of prior knowledge.

16.3.2 Sample and process fingerprinting with PAT tools

One of the often-overlooked aspects in PAT is the intrinsic capability of some monitoring tools to combine different quality attributes. For example one of the major reasons why NIR spectroscopy is such a prevalent PAT monitoring technique is its capability of capturing both physical and chemical attributes in one multivariate measurement. The different attributes can be extracted from the data in numerous ways, which may include multivariate data analysis.

Such techniques, even without sophisticated chemometrics, provide a powerful and accurate way to assess the state of a process. For example we might define the end point of blending based on the standard deviation of successive NIR diffuse reflectance spectra; or one might preprocess the NIR spectra and after a simple PCA plot of the first few component scores monitor a process trajectory over time. In both cases, depending on preprocessing and wavelength regions selected, one is using different information present in

the sample matrix at different wavelengths and of different origin (chemical and/or physical). Process state and important process phenomena are better observed if the sample is analysed as a whole and matrix effects preserved (e.g., reaction stoichiometries become observable); this will enhance our capability to detect abnormal process/product conditions.

It is therefore expected that sample and process fingerprinting become more widely used. As to regulatory acceptance they are already included in the original FDA PAT Guidance.[2] While a sample's fingerprint (e.g., its NIR diffuse reflectance spectrum) can be used as criteria to end processing or take specific control actions, a process fingerprint (i.e., consecutive sample fingerprints) will define a process trajectory. Those process trajectories will be very useful in process development and scale-up (i.e., are valuable to establish process knowledge and demonstrate process scalability) but also process supervision and control. Both activities enabled by process level fingerprinting are directly at the core of PAT: process analysis and process control.

The criticism that is often raised against fingerprinting (namely by spectroscopists) is that it is a cover up for not knowing enough about a sample's or a process's chemistry and/or physics. Put differently 'fingerprinting is what you call your measurement when you don't know what you are measuring'; a gross exaggeration of a scientifically sound approach to using the data as a means to follow the process. One might compare fingerprinting with an impressionist's masterpiece while the real accurate measurement would be the snapshot of a good digital camera capturing all the features in the original scene including those in the impressionist's painting. From a process engineering point of view it makes no sense to rule out the use of sample or process level fingerprinting as long as the rationale used is properly validated.

16.3.3 Design and control spaces

A significant amount of what makes the PAT toolbox so effective comes from the work of MacGregor and coworkers,[17,21–23] namely batch and continuous process modeling and control with multivariate regression methods. Recently a significant contribution on design and control spaces development was proposed.[22] The current argument that QbD might be achieved by defining spaces made of process variables (process inputs or process parameters) that are not fixed but still constrained (similar to a multivariate PAR approach) is put forth by the authors and shown to lead to suboptimal results. The highest process capability indices (i.e., consistently meeting the multiple critical to quality attributes all the time) are only achievable if no constraints exist in the input space so that by manipulating the process variables one might arrive at exactly the same point in the response space (control space). According to MacGregor and Bruwer[22] this needs to be taken into account by regulators and also by companies developing QbD/DS filings. It will imply larger knowledge and design spaces and more careful consideration of all aspects involved (e.g., that by constraining the input space some variability may still remain in the CQAs at the end). The Bayesian approach to DS determination put forth by Peterson[16] considers DS reliability, robustness to noise, and use of prior information.

16.3.4 Chemometrics and process analysis

The analysis of process data with chemometric techniques (e.g., classification and regression methods) might be called process chemometrics. In that area there are still challenges to be overcome. Batch processes, especially those with physical and/or chemical transformations go through different phases of different dynamics due to different rate-limiting process phenomena. For example, in a bioprocess the physiological time is different from our physical time. Moreover that physiological time varies from batch-to-batch due to a number of reasons not all known a priori and not all measurable. As such when analyzing several runs of the same process we need to align similar events (e.g., synchronize the physiological time in a bioproc-

ess). Even though several authors have considered this problem for more than a decade and used different approaches (e.g., time warping to oversample shorter process phases in particular batches) the best solution so far has been to consider the separate analysis of specific process phases.[24–26] Again this implies that an iterative approach be used, as a process knowledge basis must exist to enable the definition of meaningful phases (i.e., phases in which specific phenomena dominate) and then focus on process data for each phase alone. Another somewhat related problem is how to deal with different runs of the same process exhibiting very large differences or different levels of variability (e.g., analyzing DOE runs or analyzing runs of the same process at different scales). Some approaches have been proposed to summarize dimensionless process data in terms of a reduced number of latent variables and do the analysis, processing modeling and scale-up from there.[23] This is clearly a very important area for industrial applications that needs further work.

16.4 Emerging Application Areas of PAT

16.4.1 Biofuels

Biofuels such as biodiesel and bioethanol are, for now, the main alternatives to fossil fuels for the most polluting activities related to transportation. Presently, feedstocks for biofuels production are all of natural origin and as such are subject to uncontrolled seasonal variations. For economic reasons the purchase of these feedstocks from different parts of the world follows the dynamics of commodities markets.[27]

For the moment most biofuels are not produced in biorefineries or by intensive bioprocessing. However, as the technology area matures the processes will become more biobased and more complex. The need for the integrated management of the three problem components (raw materials, process, end products) will make PAT approaches viable for large-scale commodity type bioproductions. Most refineries of fossil crude oil would claim that this is already done through the type of multivariate control and in-process monitoring they have used for several decades already. However, that is not really so. No QbD/PAT approach is yet used in refineries to release end products based on a multivariate assessment of the multiple quality specifications that must be obeyed simultaneously (i.e., by multivariate parametric release). Also, biorefineries will create many specific problems that are better handled by PAT than by classical methods (e.g., rapid microbial methods to assess containment of genetically modified organisms).

As such one could expect that as PAT becomes better understood and used within pharmaceutical and biopharmaceutical, other areas, especially related to energy production through biobased platforms, will be a natural area for PAT to develop into an integral part of manufacturing. Perhaps it will even become mainstream in areas that originated some of PAT's tools (e.g., MSPC) like in sophisticated chemical processing.

Overall biofuels production is prone to PAT implementations the same way a bio/pharmaceutical process is. If one examines process economics and its associated logistics the case in favor of PAT is then even greater. The final product cost structure shows that there is a strong case in favor of using PAT throughout (1) raw materials qualification, (2) production process monitoring and control and (3) end product multivariate parametric release. The logistics associated with the production of large quantities of a commodity type of product by batch operations, with strict and multiple quality specifications, increases the need to reduce end product variability as well as to use fast multiparametric quality control and lot release methods.[28]

Biofuels are produced in a sequence of large batch operations involving bio/chemical reactions, separation and purification steps, followed by formulation with specific additives. The final product must comply with multiple quality specifications despite the variability in raw materials and the complexity of unit operations used in their processing.

The opportunity to define solutions based on PAT methodologies is significant:[28]

1. Assessing raw materials' variability
   a. Developing fast qualification schemes for complex raw materials (QC, quality control)
   b. Defining supervisory systems for raw materials (QA, quality assurance)
   c. Defining mixed lots of raw materials to target a specific end product
2. Handling raw materials' variability during processing
   a. Monitoring quality attributes in-process, at-line or on-line (IPCs, in-process controls)
   b. Supervising processes (e.g., multivariate trajectories)
   c. Controlling critical to quality parameters during processing using QbD/DS methodologies
3. End-product quality control
   a. Utilizing multiparametric release for biofuels based on multivariate conformity indices
   b. Blending lots and formulating the end product with additives to obtain targeted specifications.

As the biofuels industry evolves to more complex processes (e.g., involving microbial biotransformations) PAT will become more important both in scale-up and process operation, specially to nondestructively assess process performance (e.g., oil content of microalgae in submerged cultivation) and accomplish product release cheaper and faster than today's one-CQA-at-a-time methods.

If PAT is to make a stronger contribution to the production of biofuels and of other non-health bioproducts, one area that needs to move forward is the acceptance by standards monographs (e.g., ASTM) of some PAT monitoring tools (e.g., parametric release of biodiesel with NIR). For instance, accepted and proven techniques for release of many solid dosage forms and other medicines are still not accepted in the biofuels area.

16.4.2 Biomanufacturing

Bioprocesses are strongly influenced by the quality of inocula and raw materials used. The number of parameters measured in a bioprocess is often only those related to what can be controlled (e.g., cell density, pH, extracellular concentration of substrates). Sampling frequency is very conservative since those parameters are manipulated at a low frequency. As such process development is time consuming and complex mainly because a large number of experiments must be carried out to generate enough process knowledge out of the many possible combinations of variables involved (strain or cell lines used, raw material conditioning and media formulation, inocula scale-up and fermentation operation). Finally, any PAT strategy developed for a dynamic process whose trajectory strongly depends on the starting conditions (e.g., inocula quality or previous development steps) must take into account information from previous processing stages if a precise end point is to be achieved consistently.[29]

One possible and obvious strategy to cope with those difficulties would involve: (1) monitoring intensification (more parameters measured more frequently in all relevant steps), (2) tools to pull together the information of several batches (developing a process model that can be used in building design and control spaces), and (3) systems engineering tools to analyse several runs of the entire process (thus describing the interactions between production steps).

Large amounts of highly informative data are of course much easier to obtain with on-line multivariate techniques (e.g., *in situ* spectroscopy or fermentation off-gas analysis) than with fast at-line univariate techniques. Process supervision takes into consideration several types of variables at once (i.e., it is multivariate) as opposed to monitoring. Process supervision is not only monitoring in perspective (i.e., the incoming monitored data point for a variable is plotted and compared to that variable's history along the batch); it takes up a higher level by comparing the running batch against previous batches (e.g., charting the running

batch trajectory over the nominal batch trajectory). While taking a batch perspective by analyzing several historical batches in a consolidated manner, PAT as a process supervision tool has significant potential in process analysis and understanding. If in addition to historical batches with built-in nominal variability (i.e., acceptable variability) non-nominal batches are available (viz., out of specification, abnormal or batches from DOEs) a broader operating region can be explored. In this case a reliable picture can be drawn from a process by establishing the interrelation among diverse process and product variables, highlighting different phenomena and process kinetics (process phases). This will define the process design space (i.e., operating region inside which the process is robust as to being able to cancel out the variability of its inputs from propagating to the process/product quality attributes).

As an overall picture emerges of the entire multistage process, based on a more accurate picture of its unit operations, integration of this distributed knowledge into a plant or process-wide perspective is of significant value in the PAT context. In fact, this can enable the anticipation and solving of problems that may have a negative impact downstream (e.g., accepting or rejecting a certain raw material lot or fermentation inoculum).

Finally, the use of PAT should not be limited to existing processes and products but is especially attractive in the R&D and scale-up of new processes and products. PAT is especially effective in scale-up. As PAT involves consideration of all monitored variables and not only an empirical selection of some of those variables, and since in-process monitoring techniques are normally multiparametric (e.g., near-infrared spectra of a whole sample), they will be more suited to capture scale effects present in the sample's matrix that show up clearly in a consolidated multivariate analysis of quality and operating variables, thus helping the skillful engineer or scientist to pinpoint and solve scale-up problems thus resulting in a much faster process prototyping and scale-up.

16.5 New and Emerging Sensor and Control Technologies

Though many tools for monitoring of processes have been successfully used to date, innovation and development continues bringing improvements to instrumentation and data acquisition. New tools also continue to emerge, demonstrating potential for measuring at lower levels, in different environments, and with greater specificity to analytes that may not be readily detected by current technology, some of which has been presented in the previous chapters.

Spectroscopic methods including NIR, Raman and Fourier transform infrared (FTIR) have been used for PAT for several decades, and are discussed in detail in previous chapters of this book, including some discussion of recent development in the instrumentation and applications related to these techniques. The potential value from use of these analyzers continues to increase as the multivariate data are explored, often in conjunction with process parameters. Many spectroscopic tools have expanding capability as they are coupled with imaging technology, allowing for spatial information on samples, which can be beneficial in understanding formulated product uniformity, and in detecting contaminants and counterfeits.

16.5.1 Terahertz spectroscopy

Terahertz, or far infrared spectroscopy, covers the frequency range from 0.1 to 10 THz (300 to 3 cm^{-1}) where torsional modes and lattice vibrations of molecules are detected. It is increasing in use in many application areas, including analysis of crystalline materials. Several dedicated commercial instruments are available which use pulsed terahertz radiation which results in better signal to noise than those using blackbody sources for radiation (and associated with the terminology far infrared spectroscopy). Work using extended optics of FTIR instrumentation as well as continuous-wave source THz has also been recently reported.[30]

Examples of terahertz spectroscopy as a PAT tool in the pharmaceutical industry include monitoring of polymorphism, tablet coatings, and crystallinity, even at the stage of product development to aid in process design.[31]

A recent review of terahertz spectroscopy used in the study of biosystems[30] illustrates that this spectral region has fingerprints directly related to the three-dimensional structure and solvate environment of crystalline solids. The review discusses the combination of experimental data with theoretical calculations to aid in the interpretation of the terahertz spectrum, covering studies of biological samples ranging from crystalline solids to more complex aqueous forms of small proteins and biomolecules. From this work one can see that terahertz spectroscopy has the potential to contribute to overall process understanding including information on crystallinity, solvation dynamics and a relation to first principles by relating the spectral data to theoretical models.

Applications of terahertz pulsed spectroscopy (TPS) and terahertz pulsed imaging (TPI) in the biomedical and pharmaceutical industries have been presented by Wallace *et al.*[32] TPS has been demonstrated to be sensitive to detecting polymorphic forms, and for measuring the crystallinity in active pharmaceutical ingredients, with the specific example of the forms 1 and 2 of ranitidine hydrochloride distinguishable from their terahertz absorption spectra. Imaging of cancer cell tumors has been done using TPI, with TPS also being used to aid in understanding the differences in the cancerous and normal cells. TPI has potential as an *in vivo* diagnostic tool for assessment of tumors and skin lesions.

TPS can be used for the real-time monitoring of polymeric compounding processes as demonstrated by Krumbholz *et al.*[33] An industrial hardened terahertz spectrometer was interfaced with a polymer extruder, allowing for real-time measurement of the additive content in molten polymers (Figure 16.3).

16.5.2 Integrated sensing and processing

Integrated sensing and processing (ISP) is a term given to smart sensors that are designed to optimize measurements by applying signal processing to raw data to extract features in the signal. The instrument

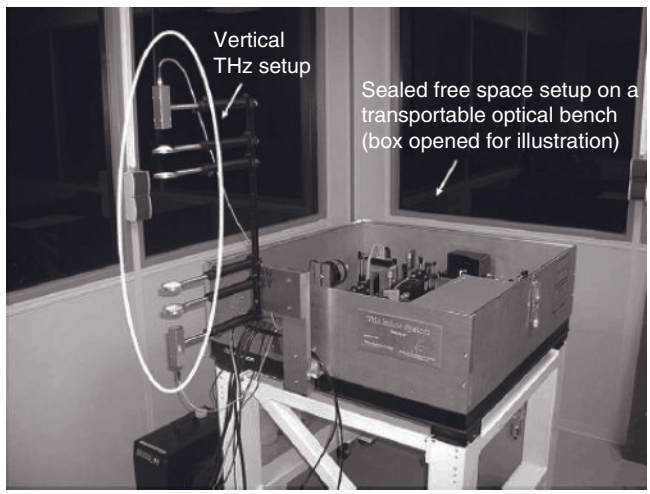


Figure 16.3 *Terahertz inline system developed for use in polymer processing. (Reprinted from Krumbholz* et al.[33] *with permission from Elsevier.)*

extracts the structured data, diminishing the need for further data processing. Using optical computing together with chemometric methods allows for the development of specialized sensors involving optical interference filters developed based on collecting data at the wavelengths most highly correlated to the chemical and physical properties of importance for a given system. Thus data acquisition can be minimized, while signal increased for a specific measurement, which can facilitate rapid measurements with high signal-to-noise in systems.[34]

Signal processing has applicability to many analytical methods, and serves as a complement to chemometric analysis, by focussing the measurement on the signals that are informative, rather than collecting data over an entire wavelength range. As data sets increase in size, ISPs will allow for their analysis without the need for taxing computational power. ISP has been demonstrated to increase the detection limit by about an order of magnitude for chromatographic analysis.[35] An application of ISP to dielectric spectroscopy,[36] expands the applicability of this technique for process automation.

16.5.3 Dielectric spectroscopy

Dielectric spectroscopy, also known as impedance spectroscopy, has been used for process analysis for some time, as it offers the ability to measure bulk physical properties of materials. It is advantageous to other spectroscopic techniques in that it is not an optical spectroscopy and is a noncontact technique, allowing for measurement without disturbing a sample or process. The penetration depth of dielectric spectroscopy can be adjusted by changing the separation between the sensor electrodes, enabling measurement through other materials to reach the substrate of interest. Because it measures the dielectric properties of materials, it can provide information not attainable from vibrational spectroscopy.

It has been shown[37–38] that radio frequency impedance (RFI) is an effective tool for monitoring cell density and cell growth of bioprocesses. The fermentation process, quite complex, is oftentimes difficult to sample and monitor. The RFI measurement could detect cell viability of *Escherichia coli* during the fermentation, serving as a qualitative measure of the metabolic load of the cell, and thus provide an *in situ* indicator of the optimal harvesting times.

The octane number of gasoline is very often measured *in situ* by using FTIR or NIR spectroscopy, enabling adjustments to the gasoline blending process to produce the desired grade. As the types of additives being used to boost gasoline octane performance now also include oxygenates and metallic ash, an alternative analytical technique with greater sensitivity to these is needed. Dielectric spectroscopy has been used to measure the octane number of clean gasoline noninvasively and nondestructively. Multivariate analysis with PLS calibrations were used to relate the dielectric spectrum to the octane number.[39] Dielectric spectroscopy can be used for process monitoring octane as a substitute method to existing technologies that may not detect all the additives used in formulating gasolines.

A novel analyzer for *in situ* monitoring by dielectric spectroscopy has been developed, with increased frequency range, and a measurement rate as short as 50 ms per spectrum. Its use has been demonstrated for measuring the methanol content of solid fuel cells, for measuring low levels of carbon monoxide at high temperatures, and for monitoring biochemical processes with high conversion rates. The system is equipped with an immersion probe and can be easily customized for applications by adapting the sensor module.[36]

16.5.4 Process chromatography

GC has been used for process analysis for many decades, along with many spectroscopic tools and univariate sensors. In recent years, developments in HPLC have made it now also available for on-line monitoring.[40–41] It has the advantage over spectroscopic methods in being able to detect trace levels of compounds, such as

impurities that may be formed during a reaction. Its use for on-line analysis has been hindered by the need for sample dilution, and the relatively long analysis time.

Reaction monitoring by on-line HPLC combined with MSPC has been presented by Zhu *et al.* for monitoring a steady state reaction over 35.2 hours.[41] Sampling was irregular during the reaction, ranging from 5 to 10 minutes, to give 197 chromatograms throughout the course of the reaction. The chromatograms at 245 nm were recorded over a 2.5 minute interval, and required preprocessing before multivariate analysis. The preprocessing included a baseline correction, followed by peak alignment using correlation optimized warping (COW), and then peak detection and integration. Window factor analysis (WFA) was then applied to the peak tables, as it allows for extraction of local information during an evolving process. The results from the WFA were used to define the normal operating conditions (NOC) of the reaction. Data that was representative of this condition were then used to develop models used for statistical control charts where in the distance of a sample from the NOC (the D statistic) and nonsystematic variations not explained by the model (the Q statistic) were computed. This works shows how on-line HPLC data need to be preprocessed. It also illustrates the ability to use these data in developing models capable of fault detection as well as for following the reaction progress in real time.

The feasibility of using on-line HPLC for making decisions on the pooling of chromatographic columns based on their quality attributes has been presented by Rathore *et al.*[40] Chromatographic separation of product is a common unit operation in bioprocessing, with decisions on pool purity and yield often based on pooling of results from offline analysis of fractions according to their absorbance at 280 nm. The approach of using on-line HPLC for column pooling decisions was demonstrated, allowing for an efficient measurement with comparable results for purity and yield. The disadvantages of using this approach at this time include the necessity of a higher level of process understanding for designing such an approach for use in a manufacturing environment, as well as a need to change the current practices in process and analytical development. With further work in this area, it is anticipated that control schemes based on real-time HPLC for pooling decisions can be developed to improve efficiencies and product quality in bioprocessing.

16.5.5 Mass spectrometry

Mass spectrometry (MS), with its inherent sensitivity, speed and molecular selectivity has been used for process analysis, including in monitoring off-gases of fermentation processes, monitoring drying processes, (again from the off-gases), and in environmental monitoring.[42] Because of its wide dynamic range it can be used to monitor a process covering large changes in component composition from the low ppm range to percents. MS is also attractive as one can monitor relative concentration from a straightforward trend of a given peak, without need of multivariate analysis. Calibration for quantitative analysis is achieved by measurement of calibration gases, which can be readily purchased. Advancements in MS instrumentation have been reviewed,[43] emphasizing the miniaturization that has been achieved, thus making available field-portable mass spectrometers.

16.5.6 Microwave resonance

Buschmuller *et al.*[44] have demonstrated that microwave resonance can be used effectively as means to monitor the moisture levels in a fluidized-bed dryer during the granulation process. The penetration depth of microwave resonance may be limited to a few microns, and hence this technique may not have any real advantages over NIR which has also been used for monitoring moisture in dryers, and has the advantage of providing chemical information such as solvent levels in addition to water, and other important properties such as polymorphic form, and particle size.

16.5.7 Novel sensors

16.5.7.1 Sensor Miniaturization: Microfluidic Flow Devices

Advancements in the area of micro total analysis systems (μ-TAS) continue towards a true 'lab on a chip,' which would include fully integrated analytical systems including results reporting, all on a single substrate. Microspectrometers (and the continued miniaturization of spectrometers in the mid-IR and NIR range) have been addressed (see Chapters 5 and 6). Gilbert[45] presents a review of the topic of microseparation for PAT, noting that there are not yet any reported applications of the use of such devices for process monitoring. Chip-based separation systems include liquid chromatography systems (LC), capillary electrophoresis (CE) and capillary electochromatography (CEC) approaches. Because LC is ubiquitous in analytical laboratories, one may expect it to also emerge for process analysis, since it is a preferred tool for laboratory analysis. But to this time, this is not the case as on-line HPLC requires sample preparation, and with this, large volumes of solvent. It has a relatively long analysis times (minutes versus milliseconds for some other available tools).

Research has been done showing that rapid pressure-driven LC analysis can be done with little solvent consumption, demonstrating this as a viable process analytical tool. Using electrokinetic nanoflow pumps LC can be miniaturized to the point of being a sensor system. Developments in terms of sampling to enable sampling directly from a process stream, to the separation channel on a chip are critical for the application of miniaturized process LC. The components (valves and pumps) required for hydrodynamic flow systems appear to be a current limitation to the full miniaturization of LC separations. Detection systems have also evolved with electrochemical detection and refractive index detection systems providing increased sensitivity in miniaturized systems when compared to standard UV-vis detection or fluorescence, which may require precolumn derivatization.

Due to their inherent simplicity, small size, efficiency and speed, the CE and CEC techniques are well suited for adoption as miniaturized process monitoring tools. Efforts have been made for such applications, coupled with continuous sampling from reaction mixtures. The CE lab on a chip has been shown to be a viable tool for real-time quantitation of a reacting system.

In the view of Gilbert,[45] the largest impact of on-line microseparation devices will be on high throughput screening for process development and monitoring of microreactors, where having automated analysis can increase the turnaround on analysis of such high volume experimentation. Thus it can have a large impact on the streamlining of product and process development, and contribute to an increased process understanding.

16.5.7.2 Biosensor for Fermentation Monitoring

A novel amperometric biosensor with an improved sampling methodology has been developed for monitoring of glycerol in bioprocesses.[46] The new flow injection analysis (FIA) system provides a simplified sampling system which avoids problems common to filtration sampling systems used for bioprocesses which suffer from membrane clogging, air entrapment, and biofouling. The biosensor uses glycerol dehydrogenase and diaphorase, which have a high selectivity and precision for measurement of process samples. Because of the high glycerol concentration during the initial phase of the fermentation to 1,3-butanediol, a dilution step was added which gives the biosensor a dynamic range of 0.2 to 50 g/L glycerol with a limit of detection of 0.1 g/L. The sampling approach with a dilution step is based on sequential injection of sample through two injection valves separated by a mixing chamber, as shown in Figure 16.4. This shows the possibility of direct sampling from a bioreactor without preseparation of the fermentation broth by filtration or dialysis. The glycerol analysis using this biosensor showed comparable reproducibility to off-line HPLC measurements. With an analysis time of two minutes it has been shown to be faster, as well as more sensitive than HPLC for glycerol monitoring.

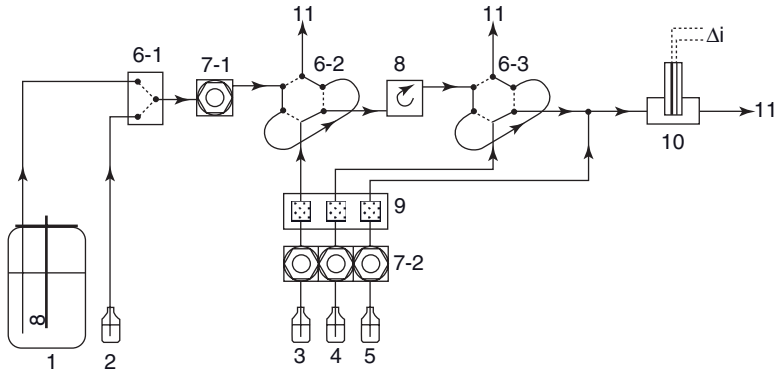


Figure 16.4 *Experimental set-up of an on-line FIA system: (1) bioreactor, (2) glycerol standard solution, (3) water for dilution, (4) buffer solution containing NAD⁺, (5) buffer solution containing ferricyanide, (6-1) switch valve, (6-2 and 6-3) injection valves, (7-1 and 7-2) peristaltic pumps, (8) mixing chamber, (9) degasser, (10) flow sensor containing biosensor, and (11) waste. (Reprinted from Sefcovicova* et al.[46] *with permission from Elsevier.)*

16.5.7.3 Quartz Crystal Microbalance

A quartz crystal microbalance (QCM)-based detection system was developed which has the ability to detect low levels of the pesticides imidacloprid and thiacloprod in aqueous solutions.[47] These pesticides, belonging to the class of neonicotinoids, are harmful to human health, and their levels in fruiting fruits and vegetables are regulated by European Union (EU) standards, hence rapid detection methods with high selectivity to the pesticides in water and in fruits are needed. A new type of molecularly imprinted monolayer (MIM) with two different templates was prepared and shown to have high selectivity for both the pesticides. This novel QCM has the potential to enable on-line monitoring of neonicotinoids in water or for the detection in fruits and vegetables. The development of specific sensors with high selectivity and sensitivity for particular compounds could also find applications in such areas as detecting contamination or adulteration of products, detecting pollutants, and monitoring water quality.

16.5.8 Inferential sensors

Inferential sensors, also known as soft sensors, are models that use readily measurable variables to determine product properties critical to prediction of product/process quality. Ideally the soft sensors are continuously monitored and controlled, or monitored on a relevant time scale. They need to make predictions quickly enough to be used for feedback control to keep process variability to a minimum.

Property prediction may be done using such routinely measured parameters as temperatures, pressures and flow rates when there is sufficient process knowledge to correlate these values to product quality. Sometimes process analyzers such as spectrometers are used to understand the process chemistry and kinetics, thus providing the ability to use soft sensors if the tool that helped elucidate the critical process variables is unavailable.

Model predictive control is concerned with continuous feedback of information with the objective of reducing the variability of product quality by changing the set points (narrowing the range) in a plant control loop.[48] By using model predictive control, projections on batch quality can be made and midstream corrections made to keep a batch within the target limits of the process.

Due to the complexity of bioprocesses, and the lack of direct in-process measurements of critical process variables, much work is being done on development of soft sensors and model predictive control of such systems. Soft sensors have long been used to estimate biomass concentration in fed-batch cultivations.[49,50] The soft sensors can be integrated into automated control structures to control the biomass growth in the fermentation.

Several statistics from the models can be used to monitor the performance of the controller.[48,49] Square prediction error (SPE) gives an indication of the quality of the PLS model. If the correlation of all variables remains the same, the SPE value should be low, and indicate that the model is operating within the limits for which it was developed. Hotelling's T^2 provides an indication of where the process is operating relative to the conditions used to develop the PLS model, while the Q statistic is a measure of the variability of a sample's response relative to the model. Thus the use of a multivariate model (PCA or PLS) within a control system can provide information on the status of the control system.

An example of the use of soft sensors is given by the automation of a penicillin production dependent on strict adherence to certain limits in the fermentation process since such biological systems are sensitive to changes in operational conditions.[51] An important issue in the use of soft sensors is what to do if one or more of the input variables are not available due, for example, to sensor failure or maintenance needs. Under such circumstances, one must rely on multivariate models to reconstruct or infer the missing sensor variable.[45]

16.6 Advances in Sampling: NeSSI

The New Sampling/Sensor initiative (NeSSI) was started from an industry-university consortium in 1999 sponsored by the Center for Process Analytical Chemistry (CPAC) at the University of Washington. Its goal is to provide solutions in terms of sampling challenges with new technology that require modularity, miniaturization and smart technology. It has succeeded in the development of a hardware standard being published (ANSI/ISA SP76.00.02), and many devices that can be used with this technology developed. There have been many outcomes also in the lab-on-a-chip and microfluidics areas from this consortium.[52] Sampling systems that allow for modularity and miniaturization continue to be developed through NeSSI as it works to consider all aspects of automated sampling systems.

16.7 Challenges Ahead

There are many hurdles in the implementation of PAT as a real-time solution in an industrial setting. Previous chapters in this book (see Chapters 2 and 15) have addressed many of them and offered tools and a systematic approach to aid scientists and managers in navigating this type of work. The challenges one faces include historical ones ('we tried this before (20 years ago); it didn't work'), cultural, organizational, financial, regulatory and technical. Organizationally it can be difficult to assemble a group with all the required skills to support work that can span numerous global R&D and manufacturing sites.

One challenge often cited is the need for more people trained to conduct PAT in an industrial setting. This is exacerbated by the lack of university degree programs that equip a professional with the skills to do this work (e.g., process analytical chemistry, chemometrics or multivariate process analysis). Granted it is a multidisciplinary activity, requiring involvement of many corporate functions, but strong foundations in analytical chemistry, process engineering, and multivariate analysis, in addition to project management will help a professional in this area. The topic of teaching process analytical chemistry has been discussed in the past[53] but still remains a field separate of itself. A very promising proposal that has the potential to benefit

industry and academia in terms of education and in merging information from PAT into development of mechanistic models has been made by Gernaey *et al.*[15] The training, done in collaboration with industry, is well aligned with the objective of developing robust and efficient manufacturing processes, where PAT tools are used to help create innovative manufacturing processes. It is proposed that case studies incorporating process data and final product properties be used to develop mechanistic models. Model simulations can be used to more fully understand the interaction of the process parameters (from multivariate as well as univariate sensors) and therefore monitoring and control systems can be designed.

One barrier to the advancement in PAT across various industries is the tendency for each industry to be insular and not necessarily look beyond its walls to see parallel science done that can be applicable in their own environments. Often the technological advancements in terms of new technologies have been discovered in disparate industries, and hence go unnoticed in new area. There is a biennial review of process analytical chemistry in *Analytical Chemistry* that presents a broad view of advances being made[54] and serves as a good resource. In looking towards the literature and discussions from other industries, one can learn of innovations that can be adapted to new problems without the need for completely new developments single handedly.

There are several consortia based at universities (CPAC, CPACT) which provide a forum for continuous learning about process analysis and new developments in terms of devices (NeSSI), analyzers (microinstrumentation), and data analysis. These consortia provide a forum for discussing needs, and a means for industries to work together in the development of tools which will support their PAT needs. They can help to bring together experts from numerous industries.

Some of the industrial challenges for the adoption of PAT in the pharmaceutical industry are how to implement this on existing products, where the investment of manufacturing equipment has already been made, and the motivation to improve the process (or perhaps even to better understand it) may not be deemed cost-effective (though there is a potential of both improving product quality and reducing manufacturing costs).

16.7.1 Continuous process validation

A recent advance in the pharmaceutical and biopharmaceutical manufacturing is the introduction of the concept of continuous quality verification (CQV) as an alternative to the previous mandated validation of a manufacturing process by following a recipe for three batches, showing production of quality product. Under the current system, the process as run during the validation is submitted for regulatory approval, and when accepted, defined as the process for the product life unless modifications were documented and approved by regulatory agencies. CQV is the approach by which the performance of a process is continuously monitored, evaluated and adjusted to maintain the control of the process to defined limits. It provides the opportunity for continuous process improvement, as one is no longer locked into an approved process (based on just three commercial-scale batches!). Guidance from the ASTM, the FDA and the ICH have been developed[3,4,11] providing a process that can be followed which is supportive of using the scientific knowledge one has about a process, and combining this with continuous monitoring (which may well include PAT applications) to demonstrate consistency and control, resulting in product of defined quality.

Process verification and control have been demonstrated in other industries, where closed-loop control has been adopted to make real-time adjustments to a process as a result of monitoring it, and having an understanding of the process parameters and how they interact.[13,17] There are similar unit operations used in the chemical industry (synthesis, crystallization, drying, etc.), with some possessing a greater complexity than pharmaceutical manufacturing processes. Demonstrating process knowledge allows for process control strategies to be developed, and enables the real-time release of product, without the necessity of additional post-production quality testing. Utilizing process monitoring tools will enable both reactive (feedback) and

proactive (feedforward) process control to consistently achieve quality product. It is anticipated that with the adoption of CQV, real-time release may be acceptable in the future for the more regulated pharmaceutical and biopharmaceutical industries as well.

An example of CQV of the batch cultivation of a vaccine has been demonstrated, where univariate (temperature, dissolved oxygen, pH) as well as spectroscopic tools were used to develop process models.[55] The measurements were used for a consistency analysis of the batch process, providing better process understanding which includes the understanding of the variations in the data. MSPC analysis of four batches of data was performed to monitor the batch trajectories, and indicated that one batch had a deviation in the pH. From the MSPC information, combined with calibration models for the composition of the process based on NIR spectral data, improved monitoring and control systems can be developed for the process, consistent with concept of CQV. The data from the univariate sensors and NIR were also fused for a global analysis of the process with a model comprised of all the measurements.

16.7.2 Data challenges: data handling and fusion

One very interesting area that will definitely see significant developments and that strongly relates to the opportunities provided by data fusion will be that of process fingerprinting (i.e., process as seen with multiple analytical instrumentation and not simple fingerprints from a single analytical record of one instrument). How to integrate the data from multiple instruments, presents a challenge, beginning with the need to align and merge the data, often collected at different time intervals, without losing important process characteristics. Approaches include projection methods such as PCA on the data from each block and plotting the PC scores of the each block to provide trajectories of the process.

Another area will be the use of correlated analytical signals obtained with different instrumentation which might be used to reinforce (adding) or cancel (subtracting) that information present in both and as such be used for process supervision and diagnostics; while, analytical instrumentation providing information that is complimentary (i.e., orthogonal and uncorrelated) might be used to expand process knowledge.

16.7.3 Regulatory challenges

The complexity of today's PAT filings for the pharmaceutical industry is significant (e.g., QbD-based) as compared to the initial filings that dealt mostly with QC protocols or simple applications to particular production steps (e.g., end-point determinations with NIR). There will be some years before bio/pharmaceutical companies internally build the capacity to manage this workflow but more importantly the regulating agencies get used to and confident with many of the concepts used (e.g., QbD). We are past simple MVDA of process historical data. The coming years will bring significant challenges for the regulatory authorities in terms of catching up with many of the techniques being used (e.g., extracting and using complex PAT monitoring tools in a process development/production context). Another area in which filings will become more frequent and on which the regulatory authorities might proactively contribute or actually delay developments, is that of predictive PAT applications based on correlating the performance of different parts of a process (i.e., managing process variability feedforward). That will require admission of variability levels, a good knowledge of process interactions and an overall QbD approach to the entire process.

16.7.4 Enterprise systems for managing data

With the growing volume of data associated with PAT, systems are needed to integrate and store it. Platforms like SIPAT from Siemens, xPAT from ABB, Discoverant from Aegis or PAT Suite from Umetrics (MKS)

claim to be able to deliver the seamless integration needed of several on-line instruments and multivariate data analysis, visualization, reporting, storage and retrieval.

Typical features (for SIPAT) include:

- data collection: capture process analytical and process data
- data mining: like MVDA model creation and validation
- real-time prediction of quality parameters
- integration with control systems (DCS)
- facilitating application of PAT in the broad sense: monitoring and control
- providing a common interface for all PAT tools (process analyzers, data mining tools …) reducing complexity, allowing standardization and assuring ease of operation
- a tool for use in manufacturing and development
- a modular and scalable architecture (allowing a scalable PAT rollout)
- easy and fast process validation.

These packages are 21CFR11 ready and can communicate with existing enterprise IT systems (e.g., ERP, MES or LIMS), but are developed with a high level of complexity.

A number of large pharmaceutical companies have been involved with the above vendors (names provided in their respective web sites) but there is still a lack of good reference sites and case studies that have been made public. One can guess that to make systems like these deliver what is expected / promised either a company will be strongly dependent on the vendor services or will need to allocate resources internally that are hard to find (i.e., PAT engineers or specialists).

16.8 Conclusion

One can see that the future looks bright for PAT. Developments continue in terms of technology that can be used, as well as in defining appropriate approaches to analyzing the vast volume of data that is generated. The implementation of PAT is maturing in the sense that it is now being approached holistically within the framework of product and process development. The information is being used to gain understanding of a process (product) and the variability within the process (product) to support controlled manufacturing maintained with process monitoring.

The future within the regulated industries is being facilitated by clarification of concepts in existing guidances and also new ones coming out in continuous process validation, which are supportive of the use of PAT throughout a process/product life cycle. With the clearer definition of QbD to develop well-understood and controlled processes, and the regulatory endorsement of making changes within a well understood operation space (the design space) we can look forward to better quality products. These can be made in an effective manner that includes monitoring and adjustment to keep the process running consistently in control.

References

1. US FDA, Pharmaceutical cGMPS for the 21st Century – A Risk-Based Approach. Final Report, September 2004, available http://www.fda.gov/downloads/Drugs/DevelopmentApprovalProcess/Manufacturing/QuestionsandAnswersonCurrentGoodManufacturingPracticescGMPforDrugs/UCM176374.pdf accessed 1 December 2009.

2. US FDA, PAT – A Framework for Innovative Pharmaceutical Manufacturing and Quality Assurance, 2004, available http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM070305.pdf.
3. International Conference on Harmonisation, http://www.ich.org/cache/compo/276-254-1.html, Guidances ICH Q8, Q9, Q10:
   ICH Q8(R1) Pharmaceutical Development, Nov. 2008
   ICH Q9 Quality Risk Management, Nov. 2005
   ICH Q10 Pharmaceutical Quality System, June 2008.
4. US FDA Guidance for Industry: Process validation: General Principles and Practice, Draft guidance, November 2008, available http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM070336.pdf accessed 1 December 2009.
5. B.W. Cue, J. Berridge and J.B. Manley, PAT and green chemistry: the intersection of benign by design and quality by design, *Pharm. Eng.*, **29**(2), 8, 10, 12, 14–18, 20 (2009).
6. L.X. Yu, Pharmaceutical quality by design: product and process development, understanding, and control, *Pharm. Res.*, **25**(4), 781–791 (2008).
7. ASTM E1655-05. Standard practices for infrared, multivariate, quantitative analysis, in *Annual Book of Standards*, vol. 3.06, ASTM International, West Conshohocken, PA, 2008.
8. ASTM E1790-04, Standard practice for near infrared qualitative analysis, in *Annual Book of Standards*, vol. 3.06, ASTM International, West Conshohocken, PA, 2008.
9. ASTM D3764-09, Standard practice for validation of the performance of process stream analyzer systems, in *Annual Book of Standards*, vol. 5.02, ASTM International, West Conshohocken, PA, 2009.
10. ASTM D6122-09 Standard practice for validation of the performance of multivariate process infrared spectrophotometers, in *Annual Book of Standards*, vol. 5.03, ASTM International, West Conshohocken, PA, 2009.
11. ASTM E2537-08 Guide for application of continuous quality verification to pharmaceutical and biopharmaceutical manufacturing, in *Annual Book of Standards*, vol. **13.01**, ASTM International, West Conshohocken, PA, 2008.
12. J.M. Juran, *Juran on Planning for Quality*, The Free Press, New York, 1988.
13. H. Wu, M.A. Khan and A.S. Hussain, Process control perspective for process analytical technology: integration of chemical engineering practice into semiconductor and pharmaceutical industries, *Chem. Eng. Comm.*, **194**, 760–779 (2007).
14. J. Lepore and J. Spavins, PQLI design space, *J. Pharm. Innov.*, **3**, 79–87 (2008).
15. K.V. Gernaey, J.M. Woodley and G. Sin, Introducing mechanistic models in process analytical technology education, *Biotechnol. J.*, **4**, 593–599 (2009).
16. J.J. Peterson, A Bayesian approach to the ICH Q8 definition of design space. *J. Biopharm. Stat.*, **18**, 959–975 (2008).
17. T. Kourti, Process analytical technology beyond real-time analyzers: the role of multivariate analysis, *Crit. Rev. Anal. Chem.*, **36**, 257–278 (2006).
18. H.W. Ward and F.E. Sistare, On-line determination and control of the water content in a continuous conversion reactor using NIR spectroscopy, *Anal. Chim. Acta*, **595**, 319–322 (2007).
19. L. Munck, A new holistic exploratory approach to systems biology by near infrared spectroscopy evaluated by chemometrics and data inspection, *J. Chemom.*, **21**, 406–426 (2007).
20. S. Wold, J. Cheney, N. Kettaneh and C. McCready, The chemometric analysis point and dynamic data in pharmaceutical and biotech production (PAT) – some objectives and approaches, *Chemom. Intell. Lab. Syst.*, **84**, 159–163 (2006).
21. T. Kourti, Abnormal situation detection, three-way data and projection methods; robust data archiving and modeling for industrial applications, *Ann. Rev. Control*, **27**, 131–139 (2003).
22. J.F. MacGregor and M.J. Bruwer, A framework for the development of design and control spaces. *J. Pharm. Innov.*, **3**(1), 15–22 (2008).
23. S. Garcia-Munoz, J.F. MacGregor, D. Neogi, B.E. Latshaw and S. Mehta, Optimization of batch operating policies. Part II. Incorporating process constraints and industrial applications, *Ind. Eng. Chem. Res.*, **47**(12), 4202–4208 (2008).
24. H.J. Ramaker, E.N.M. van Sprang, J.A. Westerhuis and A.K. Smilde, Dynamic time warping of spectroscopic batch data, *Anal. Chim. Acta*, **498**, 133–153 (2003).
25. K. Gollmer and C. Posten, Supervision of bioprocesses using a dynamic time warping algorithm, *Control Eng. Pract.*, **4**(9), 1287–1295 (1996).
26. S. Wold, N. Kettaneh, H. Fridén and A. Holmberg, Modelling and diagnostics of batch processes and analogous kinetic experiments, *Chemom. Intell. Lab. Syst.*, **44**, 331–340 (1998).

27. M.J. Haas, A.J. McAloon, W.C. Yee and T.A. Foglia, A process model to estimate biodiesel production costs *Bioresource Technol.*, **97**, 671–678 (2006).
28. J.C. Menezes, P. Felizardo and M.J. Neiva-Correia, The use of process analytical technology (PAT) tools in biofuels production, *Spectroscopy*, (9), 30–35 (2008).
29. J.G. Henriques, S. Buziol, E. Stocker, A. Voogd and J.C. Menezes Monitoring mammalian cell cultivations for monoclonal antibody production using near-infrared spectroscopy, in optical sensors systems in biotechnology, in *Advances in Biochemical Engineering/Biotechnology Series*, vol. **116** G. Rao (ed.), Springer, Heidelberg, Germany, 2009.
30. D.F. Plusquellic, K. Siegrist, E.J. Heilweil and O. Esenturk, Applications of terahertz in biosystems, *Chem Phys Chem.*, **8**, 2412–2431 (2007).
31. C.M. McGoverin, T. Rades and K.C. Gordon, Recent pharmaceutical applications of Raman and terahertz spectroscopies, *J. Pharm. Sci.*, **97**(11), 4598–4621 (2008).
32. V.P. Wallace, P.F. Taday, A.J. Fitzgerald, R.M. Woodward, J. Cluff, P.J. Pye and D.A. Arnone, Terahertz pulsed imaging and spectroscopy for biomedical and pharmaceutical applications, *Faraday Discuss.*, **126**, 255–263 (2004).
33. N. Krumbholz, T. Hochrein, N. Vieweg, *et al.*, Monitoring polymeric compounding processes inline with THz time-domain spectroscopy, *Polymer Testing*, **28**, 30–35 (2009).
34. J. Mendendorp and R.A. Lodder, Applications of integrated sensing and processing in spectroscopic imaging and sensing, *J. Chemom.*, **19**, 533–542 (2005).
35. C.Yung Fu, L.I. Petrich, P.F. Daley and A.K. Burnham, Intelligent signal processing for detection system optimization, *Anal. Chem.*, **77**, 4051–4057 (2005).
36. S. Doerner, T. Schenieder and P.R. Hauptmann, Wideband impedance spectrum analyzer for process automation applications, *Rev. Sci. Instrum.*, **78**, 105101 (2007).
37. C. Kaiser, T. Potozki, A. Ellert and R. Luttmann, Applications of PAT – process analytical technology in recombinant protein processes with *Escherichia coli, Eng. Life Sci.*, **8**, 132–138 (2008).
38. R. Fehrenbach, M. Comberbach and J.O. Pêtre, On-line biomass monitoring by capacitance measurement. *J. Biotechnol.*, **23**, 303–314 (1992).
39. L. Guan, X.L. Feng, Z.C. Li and G.M. Lin, Determination of octane numbers for clean gasoline using dielectric spectroscopy, *Fuels*, **88**, 1453–1459 (2009).
40. A.S. Rathore, M. Yu, S. Yeboah and A. Sharma, Case study and application of process analytical technology (PAT) towards bioprocessing: use of on-line high-performance liquid chromatography (HPLC) for making real-time pooling decisions for process chromatography, *Biotechnol Bioeng*, **100**(2), 306–316 (2008).
41. L. Zhu, R.G. Brereton, D.R. Thompson, P.L. Hopkins and R.E.A. Escott, On-line HPLC combined with multivariate statistical process control for the monitoring of reactions, *Anal. Chim. Acta*, **584**, 370–378 (2007).
42. T.B. Colin, Mass spectrometry in chemical process analysis, in *Spectroscopy in Process Analysis*, J.M. Chalmers (ed.), Sheffield Academic Press, Sheffield, UK, 2000.
43. A.M. Llamas, C.B. Ojeda and F.S. Rojas, Process analytical chemistry – application of mass spectrometry in environmental analysis: an overview, *Appl. Spectrosc. Rev.*, **42**, 345–367 (2007).
44. C. Buschmuller, W. Wiedey, C. Doscher, J. Dressler and J. Breitkreutz, In-line monitoring of granule moisture in fluidized-bed dryers using microwave resonanace technology, *Eur. J. Pharm. Biopharm.*, **69**, 380–387 (2008).
45. S.E. Gilbert, The future of liquid-phase microseparation devices in process analytical technology, in *Micro Instrumentation for High Throughput Experimentation and Process Intensification – a Tool for PAT*, M.V. Koch, K.M. VandenBussche and R.W. Chrisman (eds), Wiley-VCH, Weinheim, 2007.
46. J. Šefconičová, J. Katrlík, V. Štefuca, *et al.*, A filtration probe-free on-line monitoring of glycerol during fermentation by a biosensor device, *Enzyme Microb. Technol*, **42**, 434–439 (2009).
47. X. Bi and K. Yang, On-line monitoring imidacloprid and thiacloprid in celery juice using quartz crystal microbalance, *Anal. Chem.*, **81**, 527–532 (2009).
48. T. McAvoy, Intelligent 'control' applications in the process industries, *Ann. Rev. Control*, **26**, 75–86 (2002).
49. H. Zhang and B. Lennox, Integrated condition monitoring and control of fed-batch fermentation processes, *J. Process Control*, **14**, 41–50 (2004).
50. G.A. Montague, A.J. Morris and M.T. Tham, Enhancing bioprocess operability with generic software sensors, *J. Biotechnol.*, **25**, 183–201 (1992).
51. M.J. Araúzo-Bravo, J.M. Cano-Izquierdo, E. Gómez-Sánchez, M.J. López-Nieto, Y.A. Dimitraidis and J. López-Coronado, Automatization of a penicillin production process with soft sensors and an adaptive controller based on neuro fuzzy systems, *Control Eng. Pract.*, **12**, 1073–1090 (2004).
52. D.J. Veltkamp, New platform for sampling and sensor initiative (NeSSI), in *Micro Instrumentation for High Throughput Experimentation and Process Intensification – a Tool for PAT*, M.V. Koch, K. M. VandenBussche and R.W. Chrisman (eds), Wiley-VCH, Weinheim, 2007.

53. J.P. Chauvel, W.W. Henslee and L.A. Melton, Teaching process analytical chemistry, *Anal. Chem.*, **74**, 381A–384A (2002).
54. J. Workman, Jr., M. Koch, B. Lavine and R. Chrisman, Process analytical chemistry, *Anal. Chem.*, **81**, 4623–4643 (2009).
55. E.N.M. van Sprang, H.-J. Ramaker, M. Streefland, L.A. van der Pol, E.C. Beuvery and A.G. Smilde, Manufacturing vaccines: an illustration of using PAT tools for controlling the cultivation of *Bordetella pertussis*, *Quality Eng.*, **19**, 373–384 (2007).

Index

Page numbers in *italic* refer to tables or figures.

abundance, of sample components 255, 256, 263, 267–273
accelerometer 134, 281, 282, *283*, 287
accuracy
  of ammonia concentration monitoring 301
  of AOTF analyzers 126
  checking 90
  determination 8–10
  in DOE 367
  of fiber-optic based test 93–95
  in GLS weighting 376
  improving 425, 426
  of inline UV-vis spectroscopy 95
  of LIF instrumentation 346, 347, 349
  of new technologies 93
  in NIR applications 133, 444, 507, 509, 513, 516, *519*
  of packaging process 260–261
  in PAT calibration strategies 417–419
  of PDA devices 123
  in Raman spectroscopy 209
  of reference method 90, 425, 474, 499, 505
  requirement 25, 476, 495–499
  in sampling 37–40, 49, 51, 54, 61, 473
  vs precision 504–505
  wavelength/wavenumber 130, 131, 466
acoustic emission 281, 282, *283*, 284, 296, *299*, 447, 449, 455
ammonia
  acoustic chemometrics 284–286, 297–302
  Raman spectroscopy 214
analyzer
  accuracy 504
  precision 504, 505
  questionnaire (AQ) 25, 26
ANOVA 225, 355
artificial neural network (ANN) 387, 453, 467, 478, *484*
automation
  control 10
  in ISP 533
  of LIF instruments 347
  in manufacturing industries 3
  of online analyzers 181
  in PA/PAT 1, 5, *7*, 13, 22, *24*, 356, 379
  in penicillin production 537
  poorly suited methods 225
  in process control plan 26–29

band-pass filter 108, 122
biochemical manufacturing 344
biodiesel 151, 213
biofuels 331, 522, 529, 530
biomanufacturing 521, 522, 530
biomass 65, 339, 453, 454, 537
biopharmaceutical 30, 349, 350, *350*, 453, 529, 538, 539
bioprocesses
  biofuels 529
  chromatographic separation 534
  flow injection analysis 535
  LIF sensing 339, 341, *350*
  PAT strategy 530–531

Process Analytical Technology 2e Edited by Katherine Bakeev
© 2010 John Wiley & Sons, Ltd.

bioprocesses (*continued*)
physiological vs physical time 528
radio frequency impedance 533
Raman spectroscopy 213–214
soft sensors 537
UV-vis spectroscopy 96
bioreactor 213, 214, 347, 349, *350*, 535, *536*
blending
agglomeration process 257
costly 152
and domain size 257, 272–274
end point 527
of gasoline 326, 533
homogeneity 52, 246, 272–277, 450, 451, 473, 478–480
impact on product quality 152, 246, 530
NIR analysis 149, *150*
of powders 97, 214, 450–452
process IR spectroscopy *189*, *190*
Raman spectroscopy 214
validation of processes 479–480
breakthrough curves 92, 101
business value
assessment 25, 494–500, 502
NIR process analysis 494–500, 506–507, 508, 509, 512, 514, 516
QbD concept 523

calibration transfer 182, 184, 426–430, 476, 478, 502
cancer 532
carbamazepine 225
cascading libraries 471, 483
catalysis *82*, 215, 216
catalysts
heterogeneous 97
in hydrogenation reactions 218–219
in monochloroacetic acid production 216
and product properties 224, 228
Raman spectroscopy 215–216, 218–219, 224
recovery 92
solid 213
and stereochemistry of polymers 224
for transesterification 151
vanadium pyrophosphate (VPO) 97
cavitation 144, 296, *298*, *299*
cavity ring down spectroscopy 159
cell culture 214
cell density 453, 530, 533
cellulose 221, 316, 482, 483, 514, 515, *515*
cGMP 33, 34, 439, 478, 540
change control *7*, 29, 35
charge coupled device (CCD) 82, 85, 197
chemical interferants 157
chemical shift 303, 305, 308, 312, 318–320, 324, *325*
chemiluminescence 338, 340
classification
in chemical imaging data analysis 25, 251, 253–255, 257, *261*, 267
of counterfeit pharmaceuticals 217
of fluorophores 338–340
hierarchical cluster analysis (HCA) 405
LIF 341
in modelling 390–397, 423
pattern recognition methods *468*, *479*
PLS model 268, 272
in project scoping 10, *13*
SIMCA 284, 471
and validation 37, 76
cleaning validation 440, 454–455
cluster analysis 217, 225, 405, *406*, 407, 467, *479*
coatings
epoxy 196
NIR imaging 263–267, 482
NMR spectroscopy 247
optical 192, 224
particle size 481, 532
for pharmaceutical tablets 208, 210, *211*, 217, 230
protective 145, 173, 180, 183
Raman spectroscopy 210, *211*, 217, 224, 229–230
as source of errors 60, 184, 217
spectroscopic monitoring 452–453
terahertz spectroscpy 532
thickness 257, 452–453
colinearity 376, 379, 475, 526
color
and chemical properties 97
of dry solids 96–97
of extruded plastic 101–103
in image processing 251, 257, *259*, 270
measurement 84–85
in NIR spectroscopy 499, 513–514
in NMR spectroscopy 321
in Raman spectroscopy 198, 210
removal 94–95
in UV-vis spectroscopy 81, *82*
combustion 12, 157, 159, 166, 174, 176, *190*, 215
communications protocol 3, 181
compression 525
direct method 273
NIR monitoring 452, 472
NIR theory 113

on-line spectroscopy 440
powder flow 451–452
preparing for 440, 450
properties 447
and roller compaction 449
sticking 451
conformity index 480
contaminants, detection 160, 184, 199, 213, 246, 257, 261–263, 321, 454, 531
content uniformity 144, 247, 257, 258, 349, 449, 450, 452, 472, 484
continuous improvement 6, 11, 19, *24*, 30, 31, 35, 425, 440, 455
control strategy 9, 33, 519, 525, 526
cost accounting 25
cost effectiveness 27
cost penalty 389
costs
of acid 323–324
avoidance 2
in biofuel production 529
of DTGS detectors 116, 130
estimates 28, 495
of fiber-optic based systems 139, 144, 146
of innacurate predictions 430
of instruments 8, 12, *13*
of IR equipment 158, 160, 161, 164, 167, 169, 173–176, 188–189, 191, 192, 248
labor 19
of LIF eqipment 346, 347
maintenance 20
in modelling 418, 419
of NIR detectors 137
of NMR equipment 305
ownership 25, 27, 189, 200
in the pharmaceutical industry 30–31, 35
of Raman equipment 200, 220, 222
reducing/minimizing 4, 9, 49, 426, 427, 441, 493, 497–501, 525
of sampling or analysis 54, 71, 410, 515
of UV-vis sensors 92, 94
of waste 350
counterfeit 217, 257, *479*, 531
cross stream cutter 65, *66*
cross validation
in chemometrics 76–77, 289, 299, 301, 302, 379
error 379, 423, 424
full 289
genetic algorithm 424
MCR method 403
in NMR *328*, 379
in PAT 410, 411, *412*, 424
PINA analyses 325, *328*
segmented *290*, 299, *300*, 301
in support vector machines 389
time-consuming 388
crude distillation unit (CDU) 139, 147
crystal size 199, 226, 442
crystallinity *190*, 195, 199, 221, 308, 322, 532
crystallization
API 440, 442–443
isosbestic point 83
oiling-out 95–96
protocols 225
Raman spectroscopy 225, 226, 228
reflectance point 89
stress-induced 221
urea 286, 287, *288*, 289, 301

data acquisition
criteria 10
fast 122, 130, 172, 175, 176
improving 531, 533
in LIF 347
multiplexed 171
multivariate tools 440, 455
by near-line or at-plant analyzers 180
in process analysis *2*, 3, 10
in Raman spectroscopy 209, 225, 230
scanning regime 130
simplifying 433
software 299
data compression 362–363, 369, 376–377, 385, 424, 525
data mining 18, 540
decoloration, of wastewater 94–95
defense 191, 192, 339, *350*
degradation
of analyte 348
of furosemide 262–263
of model precision and accuracy 367
monitoring *189*, 479
optical 506
of pollutants 95
of signals 93, 131
of surfaces 187
density
in acoustic chemometrics 284, 296, 302
electronic 312
monitoring 322
NIR monitoring 448–450, 453, 467, 471, 474, 512
in NMR applications 322, 328

density (*continued*)
non-uniform 39, 253, 268
of powders 144, 450
in Raman spectroscopy 226, 228, 229
of sampling 267
of slurry 443
derivatives
of benzene 197
ethoxylated 149
fluorescent 337, 338, 340
mathematical 84, 90, 95, 253
Savitsky–Golay 258, 268, 371–372, 470, 475
derivatization 337, 339, 340, 349
design of experiments (DOE) 366, 523
design space 20, 33, 34, 522–524, 526, 528, 531, 540
detectors
affordable 81–82
diffraction grating monochromator 120–123
in fluorescent sensing 343, 345
indium arsenide 147, 176
in IR techniques 159, 163–165, 168–176, 182, 183, 192
low-noise 93
MEMS-based 176, *177*
multiplexed 9, 197
in NIR chemical imaging 247, 249, *250*, 251, 258, 261, 276, 277
in NIR techniques 108, *110*, *111*, 112, 114–118, 120–122, 126–132, 134, *136*, 143, 144
nonlinear responses 91, 475
performance characteristics 118, 183
in Raman spectroscopy 197, 198, 203–206, 230
range 126
reliability 183
response changes 427
response curves *117*
saturation 485
sensitivity 81–82, 108
see also charge coupled device (CCD); photodetector
diagnostics 3, 10, 27–29, 160, 184, 219, 539
dielectric spectroscopy 533
diesel *150*, *189*, 318, *319*, 320, 321, 330, *330*
differential scanning calorimetry (DSC) 221
diffraction grating 119–121, 123, 134, 135, 165, 171
diffusion 115, 303, 305, 310–312
direct calibration 380, 382, 383, 423
discriminant methods 467
dissolution
analysis 247
apparatus *94*
of drugs 87, 90, 257, 272, 439, 451
effects of crystal properties 442
prediction 452
profiling 246
rate/time 451, 482
of samples 480, 482
testing *82*, 93–94, 246, 273, 440, 482
distillation
analyzers *4*
high viscosity side-streams 139
NIR analyzers 149, *150*
process NMR spectroscopy 327–330
product composition 4
properties 152
Raman process monitoring 229
UV monitoring 96
see also crude distillation unit (CDU)
distributed control system 14, 27, 89, 501, 525
domain size 257, 272, 273, 275, 277
downhole exploration 322
drying
of API 442–444
continuous process validation 538
effect on quality 246, 442
in granulation procedures 446–449
monitoring by mass spectroscopy 534
monitoring by NIR spectroscopy 467
monitoring by Raman spectroscopy 229
powder 144
see also freeze-drying; loss on drying

electron spin resonance 321, 331
emission spectra 341, 342, 346
end point 2, *5*, 9, *82*, 144, 152, 218, 222, 441, 444, 448, 450–452, 480, 527, 530
Euclidean distance 390, 391, 421
extended mixture models 382, 383
extended multiplicative signal correction (EMSC) 375–376
external validation 410, 470, 520, *520*
extruder 96, 102, 218, 222, 445, 505, 509–512, *512*, 532

fennel oil 221
fermentation reactions
in biomanufacturing 530–531, 533–535, 537
IR spectroscopy *190*
LIF sensing 339, 347, *350*
on-line spectroscopic methods 440, 453–454, 530, 531, 533–535, 537
Raman spectroscopy 207, *208*, 213, 214

fertilizer 285, 297
fiber optics
  AOTF analzyzers 126
  cost 12
  in dissolution testing 93–94
  in IR spectroscopy 175, 184, 188, 192, 196
  in LIF sensing 341, 344, 345
  multiplexed 86, 89, 94, 100, 108, 131, 139, 500, *501*, 516
  in NIR spectroscopy 2, 126, 129, 131, 137, 139–147, *445*, 453, 464–465, 480, 507
  in Raman spectroscopy 51, 203, 204, 206, 207, 215, 220, 223, *446*
  in UV-vis spectroscopy 81, 85, 86, 87, 89, 90, 92–93, 96, 97, 99–100
FID 303–305, 309, *309*, 310, *310*, 312–318, 322, 323
film thickness 85, 99, 100, 103, 399
filtration 9, 137, 442, 535
fish oil 221
flow cell 2, 12, 86, 87, 93, 101, 138, 139, *141*, 147, 344, 348, 349
flow injection analysis 2, 340, 348, 349, 535
fluid bed dryer 443, 448, 449
fluorescence
  of API 349
  background 344
  CCD detectors 86
  derivativization techniques 340, 535
  discrimination power 337
  extrinsic 339, 349
  intrinsic 337
  and luminescence 338
  in process analysis 524
  quenching 347–348
  in Raman spectroscopy 198, *201*, 204–206, 210, 214, 216, 217, 230
  spectral profiles *342*
  see also laser induced fluroresence (LIF)
food industry
  chemometrics 354
  LIF monitoring applications *350*
  NIR techniques 108, 158, 354, 429
  NMR techniques 304, 316, 322–323, 331
  PA applications 3, 30
  PAT 521
  process sampling 65
  product formulation 31
  quality control 73
  Raman spectroscopy *190*, 191, 218, 220, 230
  UV-vis spectroscopy 92, 93
formulation
  of biofuels 529, 530
  blending 451, 481
  development 245, 246, 247, 258, 272
  dissolution testing 482
  of drug product 30, 31–32
  extruded 445
  NIR techniques 450, 481–483, *484*, 485
  and physical properties 224
  pilot plant 286
  polymorphs 482–483
  and product performance 440
  quality aspects 245–247, 440
  stability and bioavailability 444
  see also product formulation
Fourier compression 376
Fourier transform Raman (FT-Raman) 206
freeze-drying 229, 440, 453, 454
fuel cells 533
fuel properties
  of biodiesel 151–152
  of refinery hydrocarbon streams 148
fuels
  ambient air monitoring 176
  LIF monitoring applications 349, *350*
  mineral–diesel 151
  NMR applications 305, 318, 323–330
  process IR spectroscopy *189*
  see also biodiesel; biofuels
furosemide 262, *262*, 273, *274*, 278
fuzzy logic 447

gasoline
  blending *150*, 326–327
  composition monitoring *189*
  NIR spectrum 148
  NMR spectrum 320
  octane number 228, 533
  sulfur content 328–330
genetic algorithms 379
geostatistics 41, 70, 75
glucose 214, 219, 341, 454, 515
grab sampling
  inadequacies 54–56, 70
  in industrial environments *54*, 70
  methods 2, *54*
  not representative 39, 42
  in performance verification 29
  protocols 10, 11, 418
  use for calibration 498, 511
  validation 49

granulation
acoustic chemometric monitoring 282, 284–286, *288*, 295, 301
microwave resonance monitoring 534
pilot plant 473
quality aspects 246, 482
Raman spectroscopy 227–228, 229
wet 214, 446–449, 483, 485
green chemistry 522, 524
grism 205

health monitor 430–433
heterogeneity
1-D characterization 66–68, 71–75
constitutional 41–42, 51
counteracting 51, 78
cyclic and long-range 44, 48
distributional 42–45
Raman spectroscopy 209, 213, 222, 224
and representativity 41
and sampling problems 37–39, 42–44, 49, 55, 61, 62
technical challenge 32
heterogeneity NIR-CI 247, 251, 255, 256, 267, 275, *275*, 276
high throughput screening 535
histograms
in image processing 255–257, 270, 272, 274–277
in NIR-CI 265–267, 270, *271*, 272
holographic grating 108, *110*, 120, 205
hot melt extrusion 440, 445
hydrate 229, 444, 448
hydrogenation 96, 215, 218, 441
hypercube 251, *252*, 254, 263

identification
at-line approach 2
of carpet face fiber 512–514, 525
chemical 200, 246, 463
in COGM matrix *5*
of critical to quality attributes (CQA) 18, 246, 266
of equipment 27
multivariate (MVI) 525
NIR techniques 261, 266, 470–471, *479*
of pharmaceuticals 466–468, 476, *479*
of project 24
Raman techniques 217, 219, 221, 225
of raw materials 478, 500
spectral libraries 468–470
spectroscopic techniques 464, 466–468, 470–471
wavelength 127

immersion probe
in dielectric spectroscopy 533
in Raman spectroscopy 196, 206, 207, *208*, 209, 210, 214, 226, 229
in UV-vis spectroscopy 87, *88*, 91
implementation
of acoustic chemometric monitoring 295
approach 2–4
business criteria 12
of IR spectroscopy 160–162, 165–169, 175, 178, 182–184, 187, 191
of LIF techniques 347
of NIR techiques 107, 131–133, 136, 246, 445, 493–495, 498–501, 506
of NMR techniques 304, 315
Of NIR-CI 247
of PAT 522, 529, 537, 540
of process analytical technologies 17–36
in project life cycle 6–9
of Raman spectroscopy 199, 205, 207, 230, 231
of sampling unit operations 52
of UV-vis process analyser 89–91, 104
increments, compositing 39
increments in sampling 39–47, 49, 51, 52, 55–59, 62, 63, 66–71, *72*, 73, 78, 79
index of profitability 25
infrared focal-plane array 247
infrared reflection-absorption spectroscopy 454
initial rate of return 25
inspection 28, 39, 40, 58, 148, 261, 274, 323, 470, 479–481, 525
instruments
qualification *7*, 35, 466
standardization 184, 426–428, 430
interferometer
in IR techniques 159, 164, 165, *165*, 171–172, 173, 176, 178, 183, 188, 309
in NIR techniques 112, 119, 128, 129, 131, *132*, 133, 135
intermediate precision 476
isobestic points 83–84, 95

kurtosis 255, 256

lag parameter 67
Larmor precession 306, 312
laser induced breakdown spectroscopy 452
laser induced fluorescence (LIF) 338, 344–349, 351, 451
latent variables (LVs) 385–387, 389, 395, 409, 419, 470, 529

lean manufacturing 4, 5
linearity
  Beer–Lambert law 91, 184
  in classical least squares modelling 382
  dissolution test 94
  in NIR techniques *137*, 466, 506
  in photodetectors 115, 116, 122, 183
  in variable scaling 371
  see also colinearity
loadings
  manual 186
  plots *296*, *298*, *365*, *400*, *401*
  in principal components analysis (PCA) 296, 362–365, 383, 384, 386–388, 391, 398–400, 403, 414, 415, 430
loss on drying, instruments 2
lubrication 451
lyophilization 454

magnesium stearate 214, 273, *274*, 451
magnetic moment 306–308, *306*
Mahalanobis distance 29, 390, 391, 394, 405, *406*, 467, 469–471, 479
mass reduction 48, 49, *50*, 51, 52, 61, 72
master instrument 426–430, 466
mean-centering 253, 369, 370, 372
melt index 218, 228, 304, 322
methane 169, 215, 230, 323, 324
method development
  cost 513
  NIR process analyzer 496, 497
  outliers 430
  in process analysis *2*, 3, *6*, *7*, 8, 10, 11
  qualitative 466–471, 513
  quantitative 471–476
  sampling 32
method validation *20*, 34, 376, 495, 496
microfluidic flow devices 535
microreactor 330, 535
microwave resonance 449, 534
milling 227, 246, *445*, 446, 473
miniaturization 126, 158, 168, 180, 188, 534, 535, 537
mixing
  in bioprocesses 535, *536*
  chemical image analysis 274
  errors 46
  in granulation 214, 446, 448
  and heterogeneity 42, 43, 51, 52, 78
  insufficient 507
  of liquids 93
  parameters 284, 302
  in pharmaceutical applications 478, 480, 484
  of powders 450, 481
  reactor system 60, 61, 63, 64
  of solids 97
  spectral 254, 419
  speed 89, 214
  thorough 41
  time 511
model development
  for calibration 144, 146, 226, 502, 516
  in chemometrics 369, 519
  for moisture prediction 449
  for NMR techniques 317, 318
  outliers 413
  in process analysis 7
  for process analyzer implementation 501, 502, 504
  for Raman techniques 200, 224
  reference data 425
  software 432
model validation 384, 407, 408, 476, 522
modulus 322, 449
morphology 248, 257, 271, *272*, 273
multiphase fluids 282
multiplicative scatter correction (MSC)
  for powder samples and tablets 253
  pretreatment 146, 372–375, 399, 451, 475, 481
  reference spectrum 430
multivariate calibration
  in acoustic chemometrics 281, 282, 287, 296, 302
  in chemometrics 355, 366, 376, 426, 428
  in NIR spectroscopy 468, 471, 472, 474, 481, 485, 486, *515*, 525
  in PAT 38
  in process sampling 55, 59, 61, *64*
  in Raman spectroscopy 222
  in UV-vis spectroscopy 95, 96
multivariate curve resolution 218, 403, 404, 441, 454, 483
multivariate statistical process control (MSPC) *5*, *7*, 9, 301, 302, 443, 522, 525, 529, 534, 539

nanomilling 440, 444, *445*
nanoparticles 85, *445*
net analyte signal 383
net present value 25, 495
neural network *388*, 453, *484*
nugget effect 67, *69*, 70, 75

octane 228, 323, 324, *326*, 378, 381, 533
oil
  benzimidazole suspension–emulsion 226
  companies 304–305
  fluorescent sensing applications 349, *350*
  IR spectroscopy *189*
  NMR applications 305, 309, 322, 327, 328, 331
  trace concentrations 296
  in water 296, 297
  see also fuels; gasoline; refinery applications
oil seed content 126
optical density 88
optode 340, 341, 351
outliers
  in chemometrics 413–416
  detection 389, 419, 476
  diagnostics 29, 470
  exclusion 504, 517
  in 'happenstance' data 420–421
  in prediction 430–432
  t test 358

packaging
  design 228
  of instruments 117, 133, 135, 158, 160, 166, 173, 178, 192
  IR spectroscopy *190*
  NIR spectroscopy 260, 261
  NMR spectroscopy 304, 322
  Raman spectroscopy 210, 217
particle size
  in acoustic chemometric monitoring 285, 286, 289, 290–291
  determination 481–482
  and granulation 446–449, 485
  in nanomilling 444, *445*
  in NIR techniques 247, 253, 257, 268, 464, 471, 473
  quality aspects 246
  and Raman spectra 195, 464
  in Raman spectroscopy 196, 199, 210, 214, 223, 226
  reduction 52
  spectral libraries 468–469, 478
  in standard normal variate method 372
pattern recognition 254, 451, 467, 483, 486
PCA scores 295, 301, 362–365, 383, 384, 386, 391, 393–394, 398–401, 406, 414, 415, 419–421, 474, 504, 527
pesticides 263, 340, 536
photodetectors 81, 86
photometer 3, 8, 94, 169, *169*, 174, 191, 342, 344, 346
piecewise direct standardization (PDS) 477
PLS scores
  distribution 276
  images 255, 256, 258, 262, 268–272, 274, *275*, 276, *277*
  orthagonality 399
  values 274, 275
PLS scoring, in NIR-CI 268, 272
pollutant 92
polyethylene
  melt index 304, 322
  NIR techniques 150, 398–402
  NMR techniques 322
  Raman spectroscopy 210, 215, 221
polymers
  business value 498
  chemometrics 380, *382*, *385*, 386, 393, 397, 398
  film thickness determination 103
  LIF techniques *350*
  melts 96, 505–506, 516
  monitoring *190*, 509, 532
  NIR techniques 245, 378
  NMR techniques 316, 317, 322, 323, 329
  pharmaceutical grade 445
  Raman spectroscopy 195, 196, 199, 218, 221–225, 228, 229
  types 513, *514*
  unstable samples 504
polymorphism 225, 226, 442, 464, 467, 468, 473, 478, 482, 483, 532
polyols 149, *151*, *393*
pore size distribution 303, 308
portable instruments 161, 200
powder blending 450, 451
power spectrum 284
precision
  analyzer questionnnaire (AQ) 25
  in biosensors 535
  in chemometrics 367, 389, 425, 426
  in fluoresence techniques 337, 343, 345–347, 349, 367
  of instruments and methods 8–10, 12
  in IR methods 160, 166, 171, 173
  in NIR methods 130, 131, 149, 474, 476, *484*, 493, 495–500, *501*, 504, 505
  in pharmaceutical applications 452
  in Raman methods 209, 213
  in sampling 37, 39, 41, 54, 55
  in UV-vis methods 94, 95, 104
preprocessing
  in chemometrics 341, 363, 369–376, 426, 427, 429, 432

of chromatograms 534
in NIR spectroscopy 251, 253, 268, 451, 506, 527
in Raman spectroscopy 214
in UV-vis spectroscopy 90
principal component analysis (PCA) 224, 254, 281, 295, 362, 441, 451, 470, 504
principal component regression (PCR) 84, 214, 254, 383, 451, 472
principal components 224, 362–363, 376, 384, 385, 390, 398, 443, 471, 474, 480
probe fouling 32, 137, 142, 447, 448, 505–507
process analytical scientist 22
process chromatography 533–534
process control
acoustic chemometric methods 285, 290
algorithms 208
automation 22, 27
facilitation 2
improvements 497
IR system 161
models 3, 508
NIR systems 107, 153, 444, 499, 501, 525
NMR system 303, 304, 318, 322
on-line methods 349, 454, 521, 530
and PAT 521, 522, 525, 528
and process analysis 4, *5*, 18, 19, 24–26, *32*
Raman techniques 208, 209, 211, 213, 215, 220, 222–223, 230
rapid 356
requirement 246
and sampling 67
strategies 8, 9, 440, 538–539
tools 440, 455
UV-vis system 96, 104
see also multivariate statistical process control (MSPC)
process monitoring
acoustic chemometric techniques 281–299
in chemical industry *32*
distributed control system (DCS) 27
as industrial driver 19
IR techniques 157–160, 162, 168, 172, 187
multivariate modelling 33
NIR techniques 107, 133, *153*, 276, 464
and PAT 522, 529, 531, 533, 535, 538, 540
permanent installations 20
in the pharmaceutical industry 444, 461
Raman techniques 195, 212, 224, 229
sampling 38, 48, *53*, 68, 70
simulation 29
spectrofluorometric 344
UV-vis techniques 82, 95
process state 285, 286, 295
process understanding 1–3, 6, 8, 11, 14, 18
process validation *7*, 35, 524, 538, 540
product formulation *7*
project management 5, 22, 24, 152, 597
proportional integral derivative (PID) 315

qualification
of installation/operational/performance 29, 191
of materials 469, 470–471, 476, 479, 498, 500, 513, 529, 530
for process analytical chemists 23
of system/process/instruments *7*, *21*, 35, 466, 496, 502
quality by design (QbD) 1, 14, 33, 245, 440, 522, 523
quantitative data 310
quartz crystal microbalance (QCM) 536
quenching 95, 341, 347, 348

Raman techniques
applications 160, 212–230, 442–443, 445–446, 448–449, 451, 452, 454
current state 230–231
fexibility 196–197
imaging 247, 248, 452
instrumentation 203–209
introduction *4*
limitations 198–200
in PAT 531, 532
in pharmaceutical industry 464, 466–467, 470, 483
quantitative 195, 209–212
in reaction monitoring 95
sampling rate 197
spectral properties 159, 197, 200–203, 464
stability and robustness 198
random sampling 67, 70
raw materials
in biomanufacturing 530–531
heterogeneity *66*
NIR techniques 246, 464, 467, 468, 470–472, 478, *479*, 497
origin 73, *74*
in PAT 523, 524, 525, 526, 530–532
pharmaceutical *32*, 33, 463
quality control/screening 161, 179, *190*, 192, 463, 499, 530
Raman spectroscopy 209, 212
variations 29, 417, 530

Rayleigh scatter 201–203, 205, 343
reaction end point 218, 222, 441
reaction monitoring
  IR techniques 159, 161, 168, *189*, *190*
  NMR techniques 305, 312, 322, 330
  in the pharmaceutical industry 441, 535
  Raman techniques 212, 224
  UV-vis spectroscopy 95–96, 97
reference
  standardization 466
  standards 427, 465
refinery applications 139, 147, 305, 529
reflectance 528, 529
  in chemometric techniques 355, 369, 372–375, 427
  in fluorescence techniques 344
  in IR spectroscopy 162, *164*, 169, 175, 186, 187, 248, 442
  in NIR spectroscopy 108, 139, 144, 146, 249, *250*, 252, 253, 372, 392, *393*, 444, 450, 464–466, 472, 474, 480, 481, *484*, 485, 498, 512, 513, 528, 529
  in UV-vis spectroscopy 83, 84, 86, 89, 96, *98*, 102, 103
regulatory agencies 20, 30, 33, 185, 439, 463, 522, 526, 539
relaxation time 303, 308, 309, 322, 330
reliability
  in chemometrics 356, 379, 424, 425, 432
  definition 27
  importance 20, 27
  in IR techniques 158, 159, 171, 179, 180, 182, 185, 192
  key factors 29
  in NIR techniques 153, 470, 499
  and PAT 525, 527, 529
  of sampling 73, 248
repeatability 8, 126, 127, *137*, *153*, 321, 368, 425, 466, 476
replication experiment 52, 72, 73, 78
representative sample
  defining 38–40
  and grab sampling 42, 54
  importance 75, 209
  obtaining 45, 56, *58*, 59, 289, 472, 500
  in Raman spectroscopy 209, 213, 223
reproducibility
  in NIR methods 120, 121, 131, 133, 135, *137*, 139, 144, 146, 171, 173, 179, 248, 249, 255, 464, 483, 509, 510
  in NMR methods 317, 331
  in pharmaceutical products 442, 450, 470, 472, 476
residual variance 467, 469–471, *479*
resonance energy transfer 341
resonance frequency 108, 303, *307*, 312, 313, 315, 316
return
  on assets 25
  on investment 11, 25, 191, 200
riffle splitter 49
risk-based approach 33, 439, 440, 523, 526–527
risks
  analysis 35
  assessment 4, *7*, 25, 522, 524
  high 4, 8, 30
  in luminescent detection 347–348
  management 1, 439, 524, 526
  in manufacturing 1–4, 25, 31, 32, 526
  mitigation 31, 199, 218, 455, 523
  of modifications 35
  of performance failure 265, 266
  of physcial sampling 213
  reducing 1, 497
  in regulatory approval 31, 34
  safety 203–204, 213, 218, 497, 500, 507
  to the public 454
  tolerable 39
robustness
  analysis 34
  of data analysis 76
  enhancing 443
  of fiber-optical dissolution test 94
  of FTIR devices 133
  in manufacturing environment 8, 9, 20, *21*
  of NIR technologies 135, 146, 482
  as quality parameter 476
  of Raman methods 199, 205, 215, 218
  statistical 254, 529
robustness, technical requirements 495
roller compaction 446, 449

sample conditioning 2, 138, *143*, 144, 163, 181, 184, 340, 349
sampling
  accessories 157, 170
  bias 37–41, 44, 51, 55–57, 59, 61, 77, 78
  in chemometrics 284, 289, 291, 302, 356, 368, 383, 416–418, 426, 427, 432, 433
  error 37–48, 52–53, 59, 60, 66–67, 70–71, 75–76, 207, 450, 504
  fundamental principle 51
  and heterogeneity 66–75
  in-line/on-line/at-line 18

in IR spectroscopy 157, 160–165, 170, 175, 178, 179, 181, 182, 184–187, 191
maintenance 27, 29
mass reduction 48–51
in NIR spectroscopy 108, 112, 130, 136, 137, 139, 142–144, 153, 248, 249, 267, 268, 270, 472, 496–497, 499–504, 511, 514, 516, 517, 519
in NMR spectroscopy 315–317, 320–322
in on-line PAT applications 442–444, 449–451, 453
and PAT 522, 531, 535, 536, 538
practical aspects 51–57
in Raman spectroscopy 196, 197, 200, 203–207, 209, 210, 213, 216, 219, 225
reactors and vessels 58–66
spatial dimension 40
system 27, 28, 137, 157, 161, 187, 317, 321, 356, 432, 496, 536, 538
in UV-vis spectroscopy 81, 87, 89, 92, *94*, 96, 100, 102
see also grab sampling; representative sample; theory of sampling (TOS)
sampling systems
abnormal responses 432
automated 27, 187, 537
continuous mode 157, 161
cost 317
design and implementation 27–28, 496
in fermentation monitoring 535
in NMR 317, 321
score plots 286, 291–293, *294*, 301
score space 405
scores
in chemometrics 430, 474, 504, 524, 528
constraints 403
see also PCA scores; PLS scores
security 160, 168, 191, 192, 220, *350*, 432, 433
selection rules 114, 197, 202, 203, 219
sensitivity
in chemometrics 371, 383, 392
environmental 135
in fluorescent sensing 343–347, 351
of instruments 12, 82, 86, 108, 118, 173, 175–176
of measurements 8–10
in NIR techniques 108, 276, 451, 463, 464, 483, 484, 508, 509, 511
in NMR techniques 304, 315, 321, 323
of PAT techniques 523, 534–537
in Raman spectroscopy 199, 204–206, 210, 217, 219, 228
sensors
acoustic 282–283, 284, 296, 447
definition 12
for fermentation monitoring and control 453
LIF 341, 344
need for 213
NIR 449, 453
NMR 303, 313, 315, 323
noninvasive 2, 281, 282, 297
novel 12–13, 535–537
in PAT 523, 525, 526
positioning 286–287, 301, 447
in sampling 37, 38, 45, 51, 54–62, 64, 65
in the semiconductor industry 166
simple 18, 434
smart 5320533
univariate 19, 168, 285, 539
UV-vis methods 92, 96, 100
shim 95, 168, 313, 315
sill 4, 67, *69*, 70, 73–75, 79
SIMCA 284, 391, 396, 397, 471, *479*
skew 199, 225, 255, 256, 265, *265*, 266
slurry 89, 97, 136, 222, 443, 499
soft sensor 536, 537
solid dosage forms 246, 247, 264, 530
solubility *185*, *186*, 321, 442, 444, 480, 482
soybean oil 213
specificity
of an analyzer 355
of catalysts 224
desired level 470
of dissolution test 94
of monitoring applications 366
of NIR methods 108, 248, 249, 509
of NMR methods 312
as quality metric 392
as quality parameter 476
site 75
as technical requriement 495
specimens 39, 42, 56, 57, *57*, *58*, 60, 61
spectral correction 251, 252, 268
spectral image 251, 260
spectral libraries 225, 464, 467, 468, 470, 483, 513
spectral match value 471
spectral preprocessing 214, 251, 253, 451
spin relaxation 333
spray coating 440
standardization 8, 29, 173, 184, 426–430, 466, 477, 478, 540
statistical image analysis 251
statistical process monitoring 95

Stokes Raman shift 201
stratified random sampling 67, 70
stray light 85, 91, 110, 111, 120, 122, 126, *137*, 505
styrene-butadiene rubber 322
sublibrary 469
sulfur 100, 197, 230, 321, 323, 324, *325*, 328, 330
supersaturation 442
support vector machine (SVM) 389, 397
surface enhanced Raman spectroscopy 197
surface tension 297
swim lane flowchart 23, *24*
system qualification *21*, 35
systematic sampling 66, 67, 70

tablet hardness 448, 449, 452, 482
technology transfer 10, 12, 22, 33, 34, 497
terahertz pulsed image mapping 452
terahertz spectroscopy 532
theory of sampling (TOS)
  and data quality 75–76
  and heterogeneity 41–45
  mass reduction 49, *50*, 51
  as missing link in PAT 38, 75–76
  principles 37–48, 63
  representativity 39, 49, 52
  and sampling bias/error 38, 40, 44, 47–48, 51, 77
  spatial dimensions 40
trapdoor sampling 58
tunable diode laser absorption spectroscopy 454, 455
tunable laser 158, 164, *165*, 172–174, 176
turbidity *13*, 85, 96
turbulence 62, 199, 287, *298*

unit dose 450–452
unit operations
  in bioprocessing 524
  in chemical industry 538
  controlling 522
  crystallization 442–443
  definition 32
  in diesel production 330
  effect on final product 246, 257, 529
  equipment 31
  in PAT 531
  powder binding 97
  sampling (SUO) 37, 51, 52
univariate sensors 19, 285, 493, 533, 538, 539
unsupervised method 467
upward flux sampling 62
urea *150*, 285, 286, 287, *288*, *289*, 291, 292, 301

validation
  of blending processes 479
  and calibration 474, 518
  in chemometrics 37, 38, *55*, 76–78, 281, 284, 291, 299, *300*, 384, 407–413
  criteria 10
  cross-model 425
  of equipment 9, 10, 495, 522
  of filling of clinical trial blisters 260
  of homogeneity 60
  importance 20
  of NIR models 476, 501, 502–505, 519–520
  proof of concept 8
  of spectral library 469–470
  test set 75, 301, 302
  see also cleaning validation; cross-validation; method validation; process validation
vapor density 444
vapor phase 115, 160, 162
vapor pressure 228, 297, *298*
vapor stream 444, 454
vapors
  analysis 92, 100, *101*, 176, *189*, *190*
  applications 191
  deposition 229
  detection 174, 178
  flammable 180, 181
  solvent 229
  UV signatures 92
variable selection 420, 421–426, 468
variogram 66–71, *68*, *69*, *73*, *74*, 78, 79
variography 38, 52, *53*, 66, 72, 75, 78
vegetable oil 151
verification 3, *5*, 6, 9, 29, 60, 72, 78, 350, 500, 522, 538
viscosity
  in acoustic chemometrics 296, 297
  control 224
  of fuel 152
  high 139
  index 328
  in luminescence techniques 348
  monitoring 12, 516
  in NIR techniques 453, 471, 474
  nonlinear 330
  of polyester melts 516
  prediction 228
  of process stream 136, 188, 317
  of sample 143, 144
  threshold 321

wastewater
  decoloration 94–95
  treatment 82, 92, 94
water quality *5*, 344, 347, 349, 536
water vapor 160, 419, 454
wavelet compression 376
wide-area illumination (WAI) 196
wireless 161, 362
work process 17, 18, 21, 23, *24*, 33, 34, 36

X-ray diffraction 124, 215, 483

zero-filling 317