COMPREHENSIVE CHEMOMETRICS

Chemical and Biochemical Data Analysis

Volume 3

Linear Regression Modeling • Non-linear Regression • Classification • Feature Selection • Multivariate Robust Techniques

Section Editors: J. Kalivas, L. Buydens, B. Lavine, P. van Espen

Editors-in-Chief

Steven D. Brown

Romà Tauler

Beata Walczak

COMPREHENSIVE CHEMOMETRICS

Chemical and Biochemical Data Analysis

Editors-in-Chief

Steven D. Brown
University of Delaware, Newark, DE, USA

Romà Tauler
Institute of Environmental Assessment and Water Research (IDÆA), Spanish Council of Scientific Research (CSIC), Barcelona, Spain

Beata Walczak
University of Silesia, Katowice, Poland

ELSEVIER

AMSTERDAM BOSTON HEIDELBERG LONDON NEW YORK OXFORD
PARIS SAN DIEGO SAN FRANCISCO SINGAPORE SYDNEY TOKYO

Elsevier
Radarweg 29, PO Box 211, 1000 AE Amsterdam, The Netherlands
Linacre House, Jordan Hill, Oxford OX2 8DP, UK

First edition 2009

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Catalog Number: 2008943480

ISBN: 978-0-444-52702-8

For information on all Elsevier publications
visit our website at books.elsevier.com

Printed and bound in Slovenia

09 10 11 12 11 10 9 8 7 6 5 4 3 2 1

COMPREHENSIVE CHEMOMETRICS

Chemical and Biochemical Data Analysis

Related Titles of Interest

JOURNALS
Chemometrics and Intelligent Laboratory Systems
Analytica Chimica Acta
Talanta

BOOKS AND SERIES
Chemometrics in Spectroscopy
Chemometrics Tutorials
Statistics in Spectroscopy

Data Handling in Science and Technology Series
Chemometrics: A Textbook
Design and Optimization in Organic Synthesis
Data Analysis and Signal Processing in Chromatography
The Data Analysis Handbook
Experimental Design: A Chemometric Approach
Handbook of Chemometrics and Qualimetrics
Multivariate Analysis of Data in Sensory Science
Nature-inspired Methods in Chemometrics: Genetic Algorithms and Artificial Neural Networks
Scientific Data Ranking Methods
Statistical Design – Chemometrics
Wavelets in Chemistry

Preface

Some 40 years ago, the first publications appeared on the use of computer-aided mathematics to analyze chemical data. With those publications, the modern field of chemometrics was launched. Both the speed and power of computers and the sophistication of analytical instrumentation have made great leaps in the intervening time. The ready availability of chemometric software, coupled with the increasing need for rigorous, systematic examination of ever-larger and more sophisticated sets of measurements from instrumentation, has generated great interest in reliable methods for converting the mountains of measurements into more manageable piles of results, and for converting those results into nuggets of useful information. Interest in application of chemometrics has spread well beyond chemists with a need to understand and interpret their measurements; now chemometrics is helping to make important contributions in process engineering, systems biology, environmental science, and other disciplines that rely on chemical instrumentation, to name only a few areas.

As applications of chemometrics continue to grow, so too does the methodology of chemometrics itself. At 40, chemometrics is a scientific field with many mature areas, but it is also a field where change continues to occur at a rapid pace, driven both by advances in chemical instrumentation and measurement and by close connection of chemometrics with the machine learning, statistics, and signal processing research communities. The interfacial location of chemometrics, falling between measurements on the one side and statistical and computational theory and methods on the other, poses a challenge to the new practitioner: gaining sufficient breadth and depth to use chemometrics effectively.

The four volumes of *Comprehensive Chemometrics* are the result of a meeting in Oxford in September 2005, where the editors planned a work that would cover all of the major areas of chemometric research and a wide sample of current applications. Our goal was to produce a reference work that would serve both the new and the experienced practitioner. We divided the coverage of methodology into sections: Statistics, edited by L. Sarabia; Experimental Design, edited by R. Phan-Tan-Luu; Optimization, edited by R. Leardi; Data Preprocessing, edited by J. Trygg; Linear Soft-Modeling, edited by A. de Juan; Unsupervised Data Mining, edited by D. Coomans; Linear Regression Modeling, edited by J. Kalivas; Non-Linear Regression, edited by L. Buydens; Classification Feature Selection, edited by B. Lavine; and Multivariate Robust Techniques, edited by P. van Espen. The Editors-in-Chief oversaw a section on applications, where several of the newer directions in chemometrics are explored in depth.

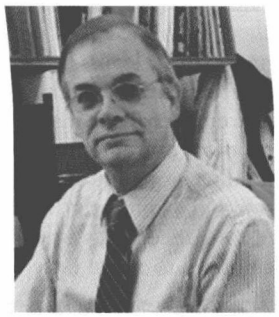
What has resulted from this collaboration is a resource that captures the practice of chemometrics in the early twenty-first century. The four volumes in this work include about 100 chapters, making this the most wide-reaching and detailed overview of the field of chemometrics ever published. *Comprehensive Chemometrics* offers depth and rigor to the new practitioner entering the field, and breadth and varied perspectives on current literature to more experienced practitioners aiming to expand their horizons. Software and data sets, both of which are especially valuable to those learning the methods, are provided in many chapters. The coverage is not only comprehensive, but it is also authoritative; authors contributing to *Comprehensive Chemometrics* are among the most distinguished practitioners of the field.

Comprehensive Chemometrics would not have been possible without the work of the Section Editors named above, who helped us identify authors and who were tireless in reviewing submissions. We also owe special thanks to Adrian Shell, Senior Acquisitions Editor at Elsevier, for supporting the project and seeing the project off, to Claire Byrne, our Developmental Editor, for her patience and persistence in seeing this project to completion, and to Hazel Harris, our Project Manager, for keeping the production schedule. Finally, we extend special thanks to all of our authors whose efforts have made the work the valuable reference that it is.

Steven D. Brown, Romá Tauler and Beata Walczak

September 2008

Editors in Chief



Steven D. Brown obtained the Ph.D. degree in analytical chemistry in 1978 from the University of Washington, working with Bruce Kowalski. The same year he was appointed Assistant Professor at the University of California, Berkeley, and he held a joint appointment at Lawrence Berkeley Laboratory. In 1981, he moved to Washington State University and, in 1986, to the Department of Chemistry and Biochemistry at the University of Delaware, where he is presently Willis F. Harrington Professor.

He has served as Chair of the Department of Chemistry and Biochemistry for 5 years, as a Section president of the American Chemical Society, and as President of the North American Chapter of the International Chemometrics Society. He is one of the three Founding Editors of the *Journal of Chemometrics* and has served for 20 years, first as its North American Editor and then for 12 years as its Editor-in-Chief. He was winner of the first EAS Award in Chemometrics in 1986.

His research interests concern a wide range of problems in chemometrics. The focus of his research has been the development of new instrumental methods through use of multivariate mathematical methods for multicomponent analysis, including calibration transfer, and the novel use of data fusion methods.

Romà Tauler Ferré obtained his PhD in Chemistry at the University of Barcelona in 1984. He is Research Professor at the Institute of Environmental Assessment and Water Research (IDÆA), CSIC, in Barcelona (Spain), since July 2003. Previously he was Associate Professor of Analytical Chemistry at the Analytical Chemistry Department of the University of Barcelona during 1987–2003. He performed postdoct stays at Institut für Anorg. u. Anal. Chemie, Univ. of Innsbruck, Innsbruck (Austria) in 1985 and 1989 and a year's sabbatical as a research scientist at the Center for Process Analytical Chemistry (CPAC), Department of Chemistry, University of Washington, Seattle, USA, in 1992. At present, he is the Editor in Chief of the journal *Chemometrics and Intelligent Laboratory Systems* and of this Major Reference Work on Chemometrics. He has published more than 200 research papers, most of them in the field of chemometrics and its applications, and in particular in the development and applications of new multivariate resolution methods. In the recent years he has focused more on the investigation of environmental problems.

Beata Walczak graduated in chemistry from the Faculty of Mathematics, Physics and Chemistry, Silesian University, Katowice, Poland, in 1979. Since then, she has been working in the Institute of Chemistry, Silesian University, where now she is the head of the Department of Chemometrics. She has been involved in chemometrics from the early 1990s and her main scientific interest is in all aspects of data exploration and modeling (dealing with missing and censored data, dealing with outliers, data representativity, enhancement of instrumental signals, signal warping, data compression, linear and nonlinear projections, development of modeling approaches, feature selection techniques, etc.).

She has authored and co-authored around 140 scientific papers and 250 conference papers, and has delivered many invited lectures at numerous international chemistry meetings. She acts as Editor of *Chemometrics and Intelligent Laboratory Systems* and 'Data Handling in Chemistry and Technology' (the Elsevier book series), and also as a member of the editorial boards of *Talanta, Analytical Letters, Journal of Chemometrics* and *Acta Chromatographica.*

Contents of All Volumes

Linear Soft-Modeling

Edited by A. de Juan, *University of Barcelona, Barcelona, Spain*

Unsupervised Data Mining

Edited by D. Coomans, *James Cook University, Townsville, QLD, Australia*

Volume 3

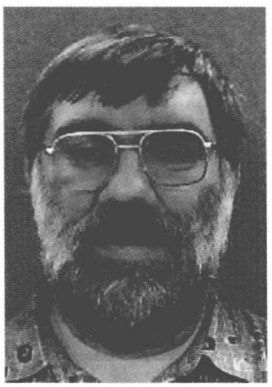
Linear Regression Modeling

Edited by J. Kalivas, *Idaho State University, Pocatello, ID, USA*

Section Editors

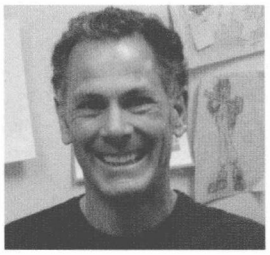
Lutgarde Buydens is head of the Chemometrics Department at the Institute for Molecules and Materials of the Radboud University of Nijmegen, the Netherlands. She originally studied pharmacy at the Brussels free University and did her Ph.D. with professor Massart on a QSAR subject. After a postdoc position at the University of Illinois in Chicago with the group of Professor William Dunn III, she started at the University of Nijmegen where she became a full professor in 1995. She received the first Elsevier Chemometrics award in 1992. Her research interests include the (further) development of chemometrical techniques within several application areas. Chemometrical techniques for combining different (spectroscopic imaging) data is her major research area.

Professor Buydens is author of more than 200 publications in international scientific journals and co-author of 4 scientific books, including the 2 volumes of the *Handbook of Chemometrics and Qualimetrics*. Since 1998, she has been the editor of the journal *Analytica Chimica Acta* and has also been the co-editor of various books and special issues of international scientific journals.



Danny Coomans was the first Ph.D. student of chemometrics of the late Professor D. L. Massart at the Vrije Universiteit Brussel, one of the founders of the field of chemometrics. He obtained the degree in 1982. He is still part-time associated with the department. He is full-time Professor in Statistics and Intelligent Data Analysis in the School of Mathematics, Physics and IT at the James Cook University, Townsville, Australia. His research interests are in multivariate statistics, statistical pattern recognition and data mining, wavelet-based feature extraction, computer intensive methods, and distributed computing. From an applied point of view, his areas of expertise include chemometrics, environmetrics, and bioinformatics with special interest in the analysis of spectral databases, microarrays, QSAR data, and aspects of industrial laboratory quality control. He has published over 250 research papers and has contributed to monographs related to chemometrics.

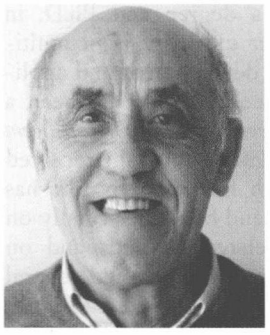
Anna de Juan has been an Associate Professor in the Department of Analytical Chemistry at the University of Barcelona since 2003, teaching chemometrics at undergraduate and graduate levels. She holds a degree and Ph.D. in chemistry from the University of Barcelona and her expertise is in multivariate curve resolution (MCR) methods: theoretical development and application to bioanalytical and analytical problems. Since 2002 she has been a member of the Editorial Advisory Board of *Chemometrics and Intelligent Laboratory Systems* and since 2006 of *Analytica Chimica Acta*. In 2004, she received the 4th Chemometrics Elsevier Award together with Karl Booksh. She has published around 60 papers in international journals and books, especially on multivariate curve resolution developments and related methods and on applications to the study of bioanalytical processes, image analysis, and general analytical applications.



John H. Kalivas is a Professor in the Department of Chemistry at Idaho State University, a primary undergraduate institution in Pocatello. He is author and co-author of over 80 professional papers, book chapters, and books. He is a member of Sigma Xi, the Society for Applied Spectroscopy, and the Council on Undergraduate Research and he serves on the Editorial Board of the *Journal of Chemometrics*, *Applied Spectroscopy*, and *Analytical Letters*. Dr. Kalivas received the B.S. degree (1978) in chemistry from California Polytechnic State University, San Luis Obispo, and the Ph.D. degree (1982) in chemistry from the University of Washington, Seattle.

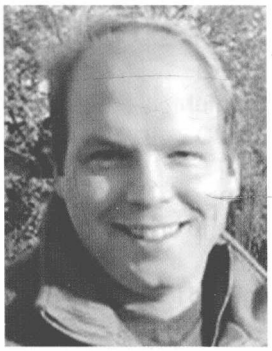
Barry K. Lavine is an Associate Professor of Chemistry at Oklahoma State University in Stillwater, OK. He has published around 90 papers in chemometrics and is on the editorial board of several journals including the *Journal of Chemometrics*, *Microchemical Journal*, and *Chemoinformatics*. He is the Assistant Editor of Chemometrics for Analytical Letters. Lavine's research interests encompass many aspects of the applications of computers in chemical analysis including pattern recognition, multivariate curve resolution, and multivariate calibration using genetic algorithms and other evolutionary techniques.

Riccardo Leardi graduated in pharmaceutical chemistry and technology in 1983. Since then, he has been working in the section of Analytic Chemistry of the Department of Pharmaceutical and Food Chemistry and Technologies of the Faculty of Pharmacy of the University of Genova, and his research field is chemometrics. His interests are mainly devoted to problems related to food, environmental and clinical data, and to experimental design and process optimization. In the last years, his research focused mainly on genetic algorithms and on three-way methods. He is author of around 80 papers and around 80 communications in national and international meetings; he has been an invited speaker in 13 international meetings and in several industries and research centers. He is Review Editor of *Journal of Chemometrics* and Editorial Adviser of *Analytica Chimica Acta*. In November 2002, he started his activity of chemometric consultancy.

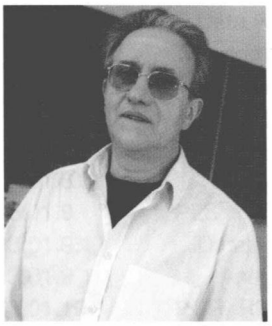


Roger Phan-Tan-Luu was Professor at the University Paul Cezanne of Marseille, France. In 1970, he founded the Laboratory of Methodology of Experimental Research in which he developed both the methodological approach and the algorithms that enable construction of efficient designs. He was at the start of the establishment of trainings about the Methodology of Experimental Research in several universities and industries in France and abroad. He is author and co-author of numerous scientific articles dealing with experimental designs. He belongs to the scientific committees of several international congresses. In 1992, Roger Phan-Tan-Luu was awarded Doctor Honoris Causa of the University of Umeå (Sweden).

Luis A. Sarabia received his Ph.D. in Statistics from the University of Valladolid (Spain) in 1979. Since 1974, he is teaching Statistics and Mathematics mostly to graduate and postgraduate students of Chemistry. At present, his research is centred on Chemometrics as a member of the Chemometrics and Qualimetrics group of the University of Burgos. His research activities include development of software and implementation of nonparametric and robust statistical methods, genetic algorithms, neural networks, etc. He is also involved in multivariate/multiway regression methods, methodology of the experimental design, quality assurance, and validation. He is the author of about a hundred papers on these matters.



Johan Trygg obtained his Ph.D in Organic Chemistry from Umeå University in 2001 (S. Wold). After receiving his Ph.D. at Umeå University he was awarded the prestigious Knut & Alice Wallenberg scholarship, and spent 2 years abroad as postDoc, first at the Institute for Molecular Bioscience (IMB), Brisbane, Australia, and later at Imperial College, London, both positions in the field of metabonomics. Johan Trygg brings with him years of analysis of high-complexity systems using an ever-broadening adaptation of chemometrics analysis, and provides expertise on the analysis and interpretation of highly complex data sets. Johan Trygg is an Associate Professor of Chemometrics at Umeå University, Sweden. He is acting group leader of Chemometrics & Bioinformatics at Umeå Plant Science Centre (UPSC) at Umeå University.



Pierre Van Espen was one of the first Ph.D. students at the University of Antwerp. He is currently Professor in the Department of Chemistry at the University of Antwerp and *Profesor Invitado* at the University of Havana, Cuba. He accomplished some pioneering work on the analysis of X-ray spectra by means of the development of software for the PDP 11/45 computer. At a later stage, this resulted in the AXIL (analysis of X-ray spectra by iterative least squares regression) package, which gained international acceptance and widespread use through the International Atomic Energy Agency (IAEA). Apart from the AXIL package, his work also comprises application and extension of chemometric methods for the analysis of data from X-ray spectrometry.

Contents of Volume 3

Contributors to Volume 3

M. C. U. Araújo
Universidade Federal da Paraíba, João Pessoa, PB, Brazil
Chapter 3.05 p. 233

F. Arteaga
Catholic University of Valencia San Vicente Ferrer, Valencia, Spain
Chapter 3.06 p. 285

B. R. Bakshi
Ohio State University, Columbus, OH, USA
Chapter 3.12 p. 453
Chapter 3.13 p. 463

T. R. Blank
Instrumentation Metrics, Chandler, AZ, USA
Chapter 3.18 p. 571

K. S. Booksh
University of Delaware, Newark, DE, USA
Chapter 3.09 p. 379

B. Bronk
Air Force Research Laboratory, Wright Patterson Air Force Base, OH, USA
Chapter 3.09 p. 379

S. D. Brown
University of Delaware, Newark, DE, USA
Chapter 3.08 p. 345
Chapter 3.17 p. 541

V. Centner
EffiChem, Lysice, Czech Republic
Chapter 3.21 p. 609

H. Chen
Ohio State University, Columbus, OH, USA
Chapter 3.12 p. 453

D. Coomans
James Cook University, Townsville, QLD, Australia
Chapter 3.23 p. 647

J. Czege
Air Force Research Laboratory, Wright Patterson Air Force Base, OH, USA
Chapter 3.09 p. 379

C. E. Davidson
Science and Technology Corporation, Edgewood, MD, USA
Chapter 3.22 p. 619

D. A. Donald
BSES Limited, Gordonvale, QLD, Australia
Chapter 3.23 p. 647

Y. L. Everingham
James Cook University, Townsville, QLD, Australia
Chapter 3.23 p. 647

N. M. Faber
Chemometry Consultancy, Beek-Ubbergen, The Netherlands
Chapter 3.03 p. 91

J. Ferré
Rovira i Virgili University, Tarragona, Spain
Chapter 3.02 p. 33