

Book review

**Comprehensive Chemometrics. Chemical and Biochemical Data Analysis. S.D. Brown, R. Tauler, B. Walczak (Eds.): Elsevier (2009)
ISBN: 978-0-444-52702-8 (Hardcover, € 1360)**

The up-to-date state and new applications of chemometrics as an interdisciplinary and continue to growth area needs to be available to a wide range of new and experience practitioners. Many books devoted to the basic and highly tailored aspects of chemometric can be found elsewhere. Elsevier maintains a special series called "Data handling in science and technology". The first attempt inside this series should be attributed to "Chemometrics: A textbook" printed in 1988. This is a one-volume book of 488 pages. The second book of such a kind was published ten years latter. It consists of two volumes and of about 1600 pages. Now, some ten years later, it is a turn for the modern version of Comprehensive Chemometrics.

The presented recourse is a four-volume book comprising 90 chapters and about 2400 pages. In contrast to the previous ones the latter does not present a holistic text written by a small team of the authors but a set of reasonably self-contained chapters offered by about 160 researches. The editors planned a work "that would cover all of the major areas of chemometric research and a wide sample of current applications" and also "a resource that captures the practice of chemometrics in the early twenty-first century." Of course the editors tried to solve a very ambitious problem. The Comprehensive Chemometrics set of books consists of four volumes that are organized in $3+3+5+1 = 12$ sections.

The first volume outlines the fundamental principles, which constitute the basis of chemometrics, underlies all methods employed in the area. The volume consists of three sections. The first part (8 chapters, edited by L. Sarabia) is devoted to Statistics. The presented topics include: Theory of Sampling, various aspects of Quality Assurance, Resampling and Robust techniques, and Bayesian approach. The overall exposition is not even; very good written texts are mixed with the chapters that suffer from the serious mistakes. The second part of Volume 1 titled Experimental Design (7 chapters, edited by R. Phan-Tan-Luu) surveys different aspects of the technique, including Screening, Factorial, and Mixture Designs, and Response Surface method. Optimization is considered in the third part (5 chapters, edited by R. Leardi). It is opened with a short review on the relevant techniques. Specific topics include Sequential, Gradient, and Multicriteria Methods. The last chapter represents an excellent introduction into Genetic Algorithms.

Second volume consists of three parts. The first part (11 chapters, edited by J. Trygg) presents various aspects of data preprocessing as the inherit part of each multivariate data analysis. The selection of the appropriate method depends on the nature of the measurement data and the problem under consideration. Some repetitions, such as twice presentation of Savitzky – Golay algorithm (in chapters 2.02 and 2.03), can be attributed to the fact of its importance and wide-spread application in chemometrics. It was unexpected to find chapter 2.10 "Batch Process Modeling and MSPC" in the pre-processing part of the book. The title of the second part "Linear Soft-Modeling" (14 chapters edited by A. de Juan) seems to be too general. The input information for different methodologies of this part is presented as a single data with various degree of complexity: two-way, multiway, and multiset data. Chapter 2.13 is devoted to the principal component analysis (PCA) as the basis for all other techniques presented in this part. PCA is introduced in a rather traditional manner; unfortunately such an important aspect as different approaches for the determination of the number of principal components does not considered at all. This aspect has a special importance as a lot of curve resolution methods presented in the consequent chapters are started with determination of the rank of a bilinear systems based on PCA methods and have reference to chapter 2.13. The third part of volume 2 "Unsupervised Data Mining" (6 chapters, edited by D. Coomans) deals with classical pattern recognition methods. Four of six chapters are presented by mathematicians from the data mining community. Therefore unsupervised classification tools are presented wider than ordinary used in chemometric applications. The method description is done in a rigorous and clear manner with formulas and algorithms. We consider that this material can serve as valuable guide in further chemometric practice.

Third volume is divided into 5 sections. The first part titled Linear Regression Modeling (9 chapters, edited by J. Kalivas) begins with a nice review of this topic. Other chapters discuss various problems related to calibration: Diagnostics, Validation, Preprocessing, Variable Selection, Missing Data, Robust methods, Model Transfer, and Three-Way methodology. The title of the second section of volume 3 is Non-Linear Regression, (5 chapters, edited by L. Buydens). However, you will not find here the classical theory of the non-linear regression. The first chapter presents an introduction to the kinetics modeling. The next chapter discusses the SVM regression and classification. It is rather unusual to find SVM in the regression section instead of the next, Classification, part. Other topics presented in this section outline Locally Weighted Regression, Neural and Fuzzy approach, Classification and Regression Trees, and Projection Pursuit Regression. The last chapter gives a nice introduction into the Neural Networks methods. The third part of volume 3, Classification (5 chapters, edited by B. Lavine), begins with an overview of Basic Concepts. The methods presented here are limited to Statistical

Discriminant Analysis, Decision Tree Modeling, and Feed-Forward ANN. The last chapter discusses the problem of validation. We have already mentioned that SVM is presented in the regression section. Another strange fact is that very popular SIMCA approach has a rather sketchy exposition in this section. The next section is called Feature Selection, (4 chapters, edited by B. Lavine). It includes, besides introduction, the outlines of Uninformative Variable Elimination method, Genetic and wavelet algorithms in application for the variable and feature selection. The last section of volume 3, Multivariate Robust Techniques, edited by P. van Espen consists of only one chapter Robust Multivariate Methods in Chemometrics that presents excellent overview of this problem.

Volume 4 (15 chapters) is completely devoted to various application areas where chemometric methods are the inherent part of the researches. Chapters are presented by recognized specialists in these areas. Each chapter is a self-contained material and presents the specific subject in an individual manner. Some chapters are organized as a condensed guide which can be used by practitioners in their further studies. Chapter 4.01 presents the theory of sampling and also discusses test set validation problems. Chapter 4.02 deals with all aspects of multivariate statistical process control and fault detection. This chapter partly overlapped with chapter 1.04 (Statistical Control of Measures and Processes) and chapter 2.10 (Batch Process Modeling and MSPC), but due to the detailed and consecutive material presentation may be used as a self-reliant textbook on MSPC. Chapter 4.10 “Chemometric role within PAT context” is very close to chapter 4.02. Some chapters present mainly the reviews of modern state of the art in different areas. For example chapter 4.03 presents the review of chemometric applications in environmental studies and underlines specific models and approaches attributed to these studies. Chapter 4.04 is a review of chemometric applications in food sciences. The emphasis of this review is on possible mistakes and wrong conclusions that can be done by poor understanding of chemometric background. In the same manner the material is presented in chapter 4.12 - Chemometric Analysis of Sensory Data. Chapter 4.06 is devoted to the application of hyperspectral imaging with combination with PLS and ANN that is an important opportunity for rapid monitoring of biological and agricultural products. Chapters 4.07-4.09 describe various biological applications, and they may, as it is mentioned in the book, “contributes to the building of a bridge across the gap between the two scientific communities”, biochemists who are not trained in data analysis and data analysts who lack biological training.

Comprehensive Chemometrics is not a convention textbook in the sense that its contents are carefully distributed in the chapters. It is more like a very helpful dictionary or an encyclopedia, which chapters are organized in a self-sufficient manner, constituting together a number of good introductions for obtaining comprehensive and essential information in the

different areas of modern chemometrics. Every chapter can be studied separately; therefore it is possible that the reader can proceed directly to the chapter of her/his interest without any loss in information. The penalty for such a layout is that some information is repeated in different places in similar or different words. This, however, could help the reader to recall or rethink a topic in a different way following the author's preferences.

It has been about 20 years since the first chemometric handbooks appeared and now everybody has to acknowledge their significance in analytical chemistry. The chemometric concept in these books has been presented to the interested non-chemometrician in a manner that does not suppose a very good background in statistics or matrix algebra. The new collection provides an essential input in presenting and clarifying the state of the art in chemometrics, and it can be highly recommended to anyone working in this field. For this reason this book should find a merited place on bookshelves alongside its predecessors.

Another view on the book will be presented in the paper 'Chemometric View on 'Comprehensive Chemometrics'' that is in preparation now. In this research we try to give an objective evaluation of this recourse employing conventional tools of multivariate data analysis.

Alexey Pomerantsev, Oxana Rodionova
Institute of Chemical Physics RAS,
Kosygin str. 4, Moscow, 119991, Russia;
e-mail: forecast@chph.ras.ru