Journal of the Serbian Chemical Society
JSCS@tmf.bg.ac.rs • www.shd.org.rs/JSCS
J. Serb. Chem. Soc. 79 (4) 509–516 (2014)
UDC 543+061.3(4)
EuCheMS news

DIVISION OF ANALYTICAL CHEMISTRY
EUROPEAN ASSOCIATION FOR CHEMICAL AND MOLECULAR SCIENCES

EUCHEMS NEWS

European Analytical Column No. 42

ROMÀ TAULER[1], FEDERICO MARINI[2], BEATA WALCZAK[3], LUTGARDE BUYDENS[4], RICHARD G BRERETON[5], WOLFGANG BUCHBERGER[6] and PAUL WORSFOLD[7]*

[1]IDAEA-CSIC, Barcelona, Spain (Roma.Tauler@idaea.csic.es), [2]Department of Chemistry, University of Rome "La Sapienza", Rome, Italy (federico.marini@uniroma1.it), [3]Institute of Chemistry, University of Silesia, Katowice, Poland (beata.walczak@us.edu.pl), [4]Radboud University, Nijmegen, The Netherlands (l.buydens@science.ru.nl), [5]University of Bristol, United Kingdom (r.g.brereton@bristol.ac.uk), [6]Analytische Chemie, Universität Linz, Altenbergerstrasse 69, A-4040 Linz, Austria (Wolfgang.Buchberger@jku.au) and [7]School of Geography, Earth and Environmental Sciences, Plymouth University, Plymouth PL48AA, United Kingdom

INFORMATION FROM THE EUCHEMS DIVISION OF ANALYTICAL CHEMISTRY (DAC)

Euroanalysis XVII, held in Warsaw, Poland, August 25–29, was the prime event within the DAC portfolio of activities for 2013. It attracted more than 700 participants from 64 countries. Maciej Jarosz as chair and Ewa Bulska as co-chair, together with their team of local organizers, provided an excellent environment for scientific discussions and networking. One of the highlights of this conference was the Robert Kellner Lecture given by Jürgen Popp from Jena, Germany.

Further events organized in co-operation with DAC included the In Vino Analytica Scientia Conference 2013 held in Reims, France, July 2–5, 2013, and the 4th Danish Metabolomics Seminar in Copenhagen, November 15, 2013.

The 44th Annual Meeting of DAC took place within the Euroanalysis conference in Warsaw on August 25, 2013. Delegates and Observers from 22 countries attended the meeting. They had the privilege to witness the DAC Tribute to Bo Karlberg and Adam Hulanicki for their achievements for DAC and the Euroanalysis series (a report about the DAC Tributes was published in the EuCheMS Newsletter November 2013).

* Corresponding author. E-mail: pworsfold@plymouth.ac.uk

The delegates of the Annual Meeting unanimously elected Paul Worsfold (UK) as Chair of DAC for a second term (2014–2016). The current Secretary (2013–2015) is Wolfgang Buchberger (Austria). The delegates also approved the other members of the Steering Committee for 2014, namely Jiri Barek (Czech Republic), Slavica Razic (Serbia), Christian Rolando (France) and Charlotta Turner (Sweden). In 2013, the Steering Committee held a meeting in Nicosia (Cyprus) on April 7, which was followed by a Mini-symposium, with lectures given by the members of the Steering Committee at the University of Cyprus. Everyone enjoyed the great hospitality of Constantina Kapnissi-Christodoulou who acted as the local coordinator. Another meeting of the Steering Committee took place prior to the DAC Annual Meeting in Warsaw.

Recently, an Open Call for the Robert Kellner Lecture 2015 and the DAC--EuCheMS Award 2015 has been made. The Robert Kellner Lecture is intended for a mid-career person with significant achievements in analytical sciences during the last five years, whereas the DAC-EuCheMS Award is for lifetime achievements of a senior person. Both awards are sponsored by Springer. Nominations should be sent to the Secretary of DAC by October 31, 2014.

The next Annual Meeting of DAC will be held on Sunday, August 31, 2014, at the 5th EuCheMS Chemistry Congress (August 31–September 4, 2014) in Istanbul. Various sections of the EuCheMS conference are closely related with analytical topics, and analytical chemists are encouraged to attend and to submit contributions.

Please make a note now of the date for the next Euroanalysis conference, Euroanalysis XVIII, which will be held in Bordeaux, France, September 6–10, 2015, under the auspices of the Societé Chimique de France. It is also important to note that in 2017 Euroanalysis XIX will be organized in Stockholm by the Swedish Chemical Society.

Further information about ongoing DAC activities can be found on its website, which has recently been moved from the old web address to the EuCheMS website and is available at www.euchems.eu/divisions/analytical-chemistry. Thanks go to Slavica Ražić (a member of the DAC Steering Committee) who has spent a lot of time and effort setting up this webpage, which contains a wide range of historical information and topical news of relevance to DAC members and the wider Analytical Chemistry community.

Currently, DAC operates five Study Groups devoted to major topics of particular importance, namely “Education in Analytical Chemistry“, “Bioanalytics“, “History”, “Quality Assurance and Accreditation”, and “Chemometrics” (see http://www.euchems.eu/divisions/analytical-chemistry/news-current-activities-conferences-and-events/study-groups-and-task-forces.html). These Study Groups are evaluated after a period of three years and may be renewed. Besides, DAC has set up a Task Force on “Archaeometry and Cultural Heritage in Anal-

ytical Chemistry" which will run for another year. In this European Analytical Column, Roma Tauler, Federico Marini, Beata Walczak, Lutgarde Buydens and Richard G. Brereton provide the Chemometrics Study Group's view on the role of chemometrics in Analytical Chemistry.

CHEMOMETRICS AND ANALYTICAL CHEMISTRY

Definition of chemometrics

Chemometrics can be defined as the science of extracting information from chemical systems using data analysis methods. It is characterized by the application of mathematical, statistical and computer methods to solve problems in different chemical disciplines and related fields, such as chemical engineering, biochemistry, medicine, environmental sciences and biology, and is related to other "-metrics" fields, such as psychometrics, biometrics and econometrics. Chemometrics is an application driven discipline that is widely used in industry and in academic and research institutions. Chemometric techniques are widely used in Analytical Chemistry and are acquiring increasing acceptance in emerging "omics" fields and Biological Sciences.

The discipline of chemometrics is distinguished from other computational branches of chemistry designed for theoretical studies (Computational Chemistry) or for the curve fitting determination of parameters using pre-established physical, chemical or empirical models (hard modeling approaches). Historically, when measured data were scarce (and often only univariate measurements were available) and computational resources limited, science progressed slowly in the description of natural systems. Advances were consequently strongly influenced by the theoretical postulation of appropriate models, based on experimental data obtained in the laboratory and under very controlled conditions, to describe the behavior of the system under study. The investigation of natural systems and of real life problems and situations was beyond the scope of these rather limited data analysis approaches. Theoretical Computational Chemistry, on the other hand, investigates the intimate nature of chemical systems at the atomic level and has developed in an independent way. In this area, the amount of available experimental data was also scarce due to the difficulty in acquisition, at least until very recently. In contrast to these two "hard" modeling approaches, the concept of "soft-modeling" can be widely applied in chemometrics. In this case, instead of the *a priori* postulation of a physical (chemical) model in which parameters are tuned (adjusted) according to an optimal fit to measured experimental data, no underlying physical model is initially proposed. Soft modeling approaches propose a simple empirical model describing the behavior of the data. A typical example is the bilinear extension of the Beer Law in UV–Visible absorption spectrometry for multi-wavelength, multi-component and multi-sample data analysis. In contrast to traditional hard modeling curve fitting

approaches, soft modeling approaches imply that one has lots of multivariate data measured over many samples that have common sources of variation (see below). Knowledge of the laws governing the system being studied (*e.g.*, the theory behind the separation mechanism in chromatography) is not required in soft modeling chemometric approaches, which are, therefore, more widely applicable than the traditional hard or "parametric" methods, since the assumptions made are less restrictive. On the other hand, these characteristics can sometimes be considered a limitation (*e.g.*, because the lack of a hard model may hamper the interpretation). To overcome this drawback, mixed "hard-soft" or hybrid approaches are being developed for different applications, *e.g.*, studying the kinetics or thermodynamics of chemical reaction systems. In these cases, the wealth of information contained in multivariate measurements (such as spectroscopic data) is combined with a knowledge of reaction orders, mass balances, rate laws and mass action thermodynamic laws to optimally describe the behavior of the experimental data. The variability of these data is not only associated with a systematic component that can be described by physical and chemical laws, but also with measurement uncertainties and interfering, unknown sources of variance. An example of this situation is the description of complex chemical reaction systems in industrial environments.

Chemometrics in analytical science

Although there may be common aspects and similarities among different computational approaches and with "-metrics"-based soft modeling disciplines in other scientific disciplines (*e.g.*, biometrics, econometrics and psychometrics), the unique feature of chemometrics is the model free analysis of experimental measurements and data to extract information about systems. The current widespread applicability of chemometric methods is precisely because of the rapidly expanding availability of experimental measurements from new analytical instruments and from the widespread use of computer systems to analyze the data. This focus of chemometrics on the analysis of experimentally measured data is the reason why it is intimately related to Analytical Science. Since Analytical Science in general, and Analytical Chemistry in particular, can be described as measurement sciences applied to materials systems (including chemical systems as well as biological, earth, physical, environmental and all other natural systems), chemometrics should be considered a cornerstone and a powerful reinforcement of the theoretical foundations of these systems. Chemometrics uses computational, mathematical, statistical and logical tools to achieve similar goals in Analytical Chemistry and should be considered an important and integral part of the subject. Analytical chemists have devoted considerable effort into the development of new instrumental methods of analysis and new methods of sample preparation, pretreatment and separation but data analysis (including the acquis-

ition, treatment and interpretation steps) should also be considered as a key step in the analytical process. Due to the limited availability of appropriate data handling and data analysis methods, most effort in analytical problem solving has been directed towards the development and improvement of physical (instrumental) and chemical (separation) solutions to the ubiquitous twin challenges of selectivity and sensitivity. In the past, simple numerical calculations and traditional univariate statistical analysis were considered sufficient to address these challenges but, even with more sophisticated analytical methods and instrumentation (*e.g.*, mass spectrometry), new challenges continually arise, especially in the analysis of natural samples. This is particularly the case with the explosion of massive and megavariate analytical data from the "-omics" platforms, which cannot be processed using these traditional data analysis tools.

Chemical concepts

Chemometrics has been applied to solve both descriptive and predictive problems in the experimental sciences, especially in chemistry. In descriptive applications, the properties of chemical systems are modeled with the objective of understanding the underlying relationships and structure of the system (*i.e.*, model understanding and identification). In predictive applications, quantitative numerical values (*e.g.*, concentrations of the analytes) and properties of chemical systems are modeled with the intent of estimating new values and properties. Many early applications involved multivariate classification, multivariate calibration and numerous quantitative predictive applications.

Compared with data analysis disciplines in other fields, chemometrics depends on certain highly characteristic steps. From the beginning, finding techniques for optimal data preprocessing and pretreatment have been extraordinarily important and they are still at the core of the best strategies to extract reliable information from instrumental data. From the well-known Savitzky–Golay differentiation and smoothing filters, to the more recently developed wavelet transforms, asymmetric least squares baseline correction and optimized warping techniques for signal alignment and correction, together with the more traditional preprocessing methods of data centering, normalization and scaling, at present a plethora of possibilities exist for handling many of the problems currently encountered in real life experimental measurements.

Another very important concept in chemometrics is that of "latent variables" (originally developed 80 years ago in the psychometric and factor analysis literature), *i.e.*, variables which are not directly measured (they are hidden, not explicit), which need to be uncovered for a proper interpretation of what is observed and for future predictions. This concept is in the core of many of the more widely used methods in chemometrics, such as Principal Component Analysis and Partial Least Squares, and of Factor Analysis and Pattern Recognition methods in

general. This also implies recognition of the fact that, in most cases, the ultimate object of investigations can only be measured indirectly.

Another core concept related to latent variables is that of Mixture Analysis, which recognizes that existing analytical methods are imperfect, in the sense that they are not (and that they probably cannot be) totally selective, and therefore that the measured signal is usually a mixed signal which needs to be "unmixed" into its component parts before one can use and interpret its content.

An additional concept that is not unique to Chemometrics but which, in this context, has become highly important if not fundamental is the concept of validation. For predictive modeling, the results and method performance should be validated properly and no new approach should be accepted until this is clearly shown. Validation, as the name suggests, is the step in which the reliability of the analytical approach to the data is established. In its most commonly adopted meaning, it involves taking out samples that are not part of the training set and then checking how well the model performs when applied to these external samples. However, the concept of validation is broader and includes an evaluation of the appropriateness of the model in describing the data, the identification and treatment of outlying observations, the chemical interpretability of the results, as well as more technical issues, such as whether the solutions are stable or whether the algorithm has truly converged to the global optimum. Accordingly, it follows that the number of possible strategies for validation is very large.

The number of instrumental data sources that are now available significantly increases the possibility of having the same system simultaneously investigated by multiple methods and instruments. The concept of data fusion and the design of novel chemometric techniques for this purpose are opening up new possibilities to correlate and interpret more deeply the nature of complex systems. Data sources can now be massive and, in addition, natural systems (such as environmental or biological systems) are intrinsically complex. Their investigation and analysis are very challenging and require the development and adaptation of new methods and strategies for data analysis.

Emerging trends

Since the historical development of chemometrics, new situations are constantly emerging. Although one may think that many of the initial goals of chemometrics have already been reached and that chemometrics can now be absorbed by other disciplines, such as "-omics" or bioinformatics, there should be a clear scientific recognition of the extraordinary efforts and successes achieved by this discipline. More and more, the worldwide spread of chemometrics solving new problems and providing new solutions to both new and old analytical challenges can be observed. The development of new chemometric methods and solutions is still required to address emerging challenges in the measurement sciences.

Nowadays chemometrics has reached a mature state and is accepted in many universities, research institutions and research departments of large chemical industries throughout the world. For instance, it is routinely applied as a fundamental part of Process Analytical Technologies (PAT) in the oil, food and pharmaceutical industries. New application areas have appeared in different domains, such as molecular modeling, QSAR, and chemoinformatics. It has a strong impact on the development of new tools in hyperspectral image analysis, environmental analysis and especially in the "-omics" (genomics, proteomics and metabolomics) branch of analytical science. "-Omic" disciplines are changing the face of analytical science, and will continue to do so over the next few years, and chemometrics has an important role to play in this development. It is only with this intimate relationship between chemometrics and "-omic" analytical methodologies that the challenges posed in the analysis of overwhelmingly large instrumental data sets can be overcome and extraction of the required information be enabled. This will then allow the differentiation of natural sources of variation from the effects caused by, *e.g.*, stressors and treatments.

Publications

Apart from scientific journals uniquely devoted to chemometrics such as the *Journal of Chemometrics* (Wiley) and *Chemometrics and Intelligent Laboratory Systems* (Elsevier), most routine applications of existing chemometric methods are published in broad analytical journals (*e.g.*, *Analytical Chemistry*, *Analytica Chimica Acta* (which was the first journal with a separate chemometrics section), *Analytical and Bioanalytical Chemistry*, *Analyst*, *Talanta* and *Applied Spectroscopy*). Moreover, several important books/monographs on chemometrics have been published over the last 30 years, either dealing with specific or more general aspects of the discipline. These include: Malinowski's "Factor Analysis in Chemistry" (Wiley, 1989, 2002), the two milestone chemometrics textbooks from Sharaf, Illman and Kowalski (Wiley, 1986) and Massart *et al.* (Elsevier 1988, 1998), "Multivariate Calibration" by Martens and Naes (Wiley, 1989), and more recent publications such as "Comprehensive Chemometrics", a multi-authored chemometrics "encyclopedia" in four volumes (Elsevier, 2009), the "Practical Guide to Chemometrics" edited by Paul Gemperline (CRC Press, 2006), the monographs by Richard Brereton ("Chemometrics Data Analysis for the Laboratory and Chemical Plant" (Wiley, 2003) and "Chemometrics for Pattern Recognition" (Wiley, 2009) and the "Multi-way Analysis with Applications in the Chemical Sciences" by Smilde, Bro and Geladi (Wiley, 2004).

In addition, the interest in using chemometrics in different fields is witnessed by application-oriented monographs such as "Chemometrics in Environmental Chemistry" by Einax (Springer, 1995), "Multi- and Mega-Variate Data Analysis"

by Eriksson *et al.* for omics data analysis (Umetrics Academy, ISBN-13: 978-91--973730-2-9) and Marini's "Chemometrics in Food Chemistry" (Elsevier, 2013).

Since chemometrics involves the use of mathematical, statistical and computer methods, a very important aspect of this discipline from its very beginning has been the attention placed on software development and dissemination. The forefathers of chemometrics took particular care in writing computer routines to perform essential chemometric computations and to organize them in packages that could be easily used by less expert researchers: Kowalski's ARTHUR, Pirouette (Infometrix, WA, USA), Wold's SIMCA (UMETRICS, Umeä, Sweden), Forina's PARVUS and Martens' Unscrambler (CAMO, Oslo, Norway) have played an important role in the dissemination of the discipline and most still exist, even if now developed by software companies, flanked by new contributions such as Eigenvector's PLS-Toolbox (Eigenvector Research, WA, USA).

CONCLUSIONS

In conclusion, although applications of chemometrics are not restricted to Analytical Chemistry, its "home" within this discipline is well justified, and in many situations the goals of both coincide. In addition, in the same way that Analytical Chemistry has expanded to encompass all branches of Analytical Science, the same is true of Chemometrics, which can easily be extended from chemically related measurements to most other types of measurements performed throughout science and engineering, including chemical engineering.