Anal. Chem. **2008,** *80,* 4519–4531

Chemometrics

Barry Lavine* and Jerome Workman

Department of Chemistry, Oklahoma State University, Stillwater, Oklahoma 74078, and Luminous Medical Inc., Carlsbad, California 92010

Review Contents

Calibration	4521
Resolution	4525
Pattern Recognition	4526
Literature Cited	4530

This review, the seventeenth of this series and the fifteenth with the title of Chemometrics, covers the most significant developments in the field from January 2006 through December 2007. As in the previous review (*1*), breakthroughs and advances in the field are highlighted, and trends within the field are evaluated. As is the current restriction for such reviews, the limit for citations is less than 200 references, which continues to pose a challenge since the number of citations in the field of chemometrics continues to grow. If you include the broader areas of image data analysis, digital and signal processing, photonics, biophysics, drug discovery, sensors, in silico processing for high throughput screening, structure and toxicity analysis, genetic analysis, and the like, the number of papers "balloons" significantly. There are many thousands of citations across these fields specifically dealing with chemometric-like data processing methods during 2006 and 2007. This should come as no surprise since disciplines related to computer science, informatics, chemoinformatics, bioinformatics, and all engineering disciplines, including bioengineering are facile in the use of multivariate analysis methods for both basic research and applied science.

During 2006 and 2007, a large number of review articles on chemometrics and applications of chemometrics to fields other than chemistry have appeared in the literature. The biannual reviews in analytical chemistry include process analytical chemistry (*2*), gas chromatography (*3*), and quantitative structure−retention relationships (*4*). A review of multivariate autofluorescence of intact food systems was also published during this period (*5*).

Specific chemometrics or chemoinformatics reviews over this same period include Chemoinformatics: Past, Present, and Future (*6*). This review gives a history of chemoinformatics from the 1940s to 2006. The review focuses on those facets of chemoinformatics that have overlap with chemometrics. This includes overall chemical database systems and structures, computer-assisted structural elucidation, computer-assisted drug and chemical synthesis design, and 3D structure configuration algorithms and systems. The review emphasizes the progress of chemoinformatics in its core areas, namely, the development of computer technology. The progress made in the design and implementation of computational systems has been a major enabler of new developments and is the power of chemoinformatics. A summary of this field and its future prospects are described in this review.

Any reader interested in the history and etymology of the term chemometrics will find the paper entitled The Past, Present, and Future of Chemometrics Worldwide: Some Etymological, Linguistic, and Bibliometric Investigations fascinating. The authors explore the history and usage of the term chemometrics in its various forms across the world and cite the significant events in the history of chemometrics in research and in publications (*7*). A few highlights delineated by the authors include their finding that a total of 82 written and 127 pronunciation forms of chemometrics were found in 48 languages worldwide. This comprehensive, historically precise, and well referenced paper states that from 1971 to 1990 the subdiscipline of chemometrics became firmly established. With citations from primary references and descriptions of historical details, they point to the foundations of chemometrics. In 1971, Professor Svante Wold (Umeå University, Sweden) coined the term kemometri in Swedish from kemo and metri and its English equivalent chemometrics. In 1972, Prof. S. Wold named his group Forskningsgruppen för Kemometri (Research Group for Chemometrics) or Kemometrigruppen (Chemometrics Group) and published the first article using the term kemometri. In 1973, the first article appeared in the literature that had the term chemometrics as part of the name of the research group (Prof. S. Wold's group), and this is recorded in the SCI database. In 1974, Prof. S. Wold gave a definition for chemometrics, "The art of extracting chemically relevant information from data produced in chemical experiments is given the name of 'chemometrics' in analogy with biometrics, econometrics, etc." Also later in 1974, The International Chemometrics Society was founded by Professor S. Wold and Professor Bruce R. Kowalski from the University of Washington, WA (June 10, 1974). In 1975, the first article that had the term chemometrics in its title and also the second article with chemometrics as part of the name of a research group, was published by Prof. B. R. Kowalski, and this is also recorded in the SCI database. Kowalski's article defined chemometrics as a new chemical discipline. A year-by-year description of the history of the field is also provided by the authors of this review as well as the origin of the fundamental review for *Analytical Chemistry*. In 1980, the first review on chemometrics appeared in the fundamental reviews issue of *Analytical Chemistry*, as a successor to previous reviews on statistical and mathematical methods in analytical chemistry (April 1980). This paper is well worth the effort to read and archive for future reference.

A review of quantitative structure−activity relationship (QSAR) methods has also been published during this period (*8*). The

* To whom correspondence should be addressed.

10.1021/ac800728t CCC: $40.75 © 2008 American Chemical Society
Published on Web 05/17/2008

authors describe and evaluate several state-of-the-art machine learning tools. Computational tools discussed include support vector machines (SVM) and decision tree methods such as boosting, bagging, and random forest. The authors compare these methods using eight data sets. By the use of comparative statistical tests, the authors conclude that all of these techniques can provide consistent improvements in predictive performance over single decision trees. The study did not identify the best-performing algorithm. The conclusion was that a more in-depth investigation into the properties of random forests would be informative. The authors also defined a set of parameters for the random forest method that provided optimal performance across all of the data sets in the study.

A basic overview of in silico methods has also been published during this period (*9*). This tutorial describes how in silico methods are a valid tool for analyzing the properties of chemical compounds and are of interest as computational modeling techniques. These methods are used to build predictive models of chemical activity given the structure of the compounds. These methods have been used to model toxicity or biological activity, such as interactions with biological receptor macromolecules. This review included examples of modeling to predict estrogen receptor (ER)-mediated effects. Nuclear receptors, such as ER, have been studied with in silico tools due to the published concern regarding endocrine disrupters, which have been implicated as interferents with hormone regulation. QSAR related methods, such as 3D-QSAR and virtual docking, have been used to investigate these phenomena and are described in this review. Molecular modeling techniques and application of the aforementioned techniques are presented and discussed.

Standard methods of analysis including chemometric and data processing are developed and maintained on a continuous basis through such organizations as the American Society for Testing and Materials (ASTM International, http://www.astm.org/), the United States Pharmacopeia (USP, http://www.usp.org/), the Association of Analytical Communities (AOAC International, http://www.aoac.org/), and the EPA (http://www.epa.gov/OSA/fem/methcollectns.htm). The most recent developments in ASTM over the review period include ASTM E2056-04 Standard Practice for Qualifying Spectrometers and Spectrophotometers for Use in Multivariate Analyses Calibrated Through Surrogate Mixtures, ASTM E1655-05 Standard Practices for Infrared Multivariate Quantitative Analysis, ASTM D6122-06e1 Standard Practice for Validation of the Performance of Multivariate Process Infrared Spectrophotometers, ASTM D7235-05 Standard Guide for Establishing a Linear Correlation Relationship Between Analyzer and Primary Test Method Results Using Relevant ASTM Standard Practices, and ASTM D3764-06e1 Standard Practice for Validation of the Performance of Process Stream Analyzer Systems. Of these methods involving chemometrics for standard laboratory and process analytical chemistry, ASTM D6122 has had its newest version released in 2006 (*10*). The scope of the document describes a standard practice for the validation of analytical measurements made by online, process near- or mid-infrared analyzers used for the calculation of physical, chemical, or quality parameters or properties. Its description applies to a broad application of these spectrometers but has special reference to mixtures of liquid hydrocarbons. Analytical standard practices for multivariate methods, describing specific algorithms and the application of such methods for calibrating instrumentation has been broadly addressed by both the D02 and E13 main committees of ASTM for over a decade.

Many of the chronic challenges to chemometric methods existing since the early 1980s still remain, such as calibration transfer, accurate spectral comparison of test samples against library spectra, improving specificity and signal-to-noise for low signal in spectroscopic first-order data, and methods to address scattering across various regimes of absorptive, scattering media. There are discussions, reviews, modeling, and algorithm comparisons published on these subjects, but the fundamental first principle derivations are still lacking. Thus, industrial practitioners are still confronted with the situation of applying a potpourri of algorithms to their data to empirically determine what "works" best for their own application. They end up searching for models with fewer factors and smaller SEPs rather than applying a thorough understanding and rigor to their analysis problems. That said there are many new and exciting studies discussing a variety of topics and providing methods that allow us to "see" more deeply into multidimensional domains and to explore and mine the information content of data measured on complex chemical systems.

National and Regional Chemometric Societies are listed within a single Web site. The list of regional or national chemometrics societies is not comprehensive but invites chemometric societies to update or newly submit their contact and general information to the Webmaster. As noted by the site listing, some of the societies are quite active while the status of activity for some is light to dormant. The links included at this Web site (http://www.namics.nysaes.cornell.edu/chem society.html) are Australian, Belgian, British, Czech, Danish, Dutch, German, Finnish, French, Italian, North American, Norwegian, Russian, South African, Spanish, and Swedish.

Umea University is the very home of the origins of chemometrics activity worldwide and Umea maintains an active research group and Web site. Their Analytical Chemistry Springboard includes a comprehensive Web site of chemometrics links located at http://www.anachem.umu.se/cgi-bin/jumpstation.exe?Chemometrics.

A special Web site with URL of http://cheminformatics.org/Cheminformatics Links contains 635 links in 90 categories, including 44 data sets in 9 distinct categories. The Wiley Chemometrics & Informatics site contains news events and conferences as well as recent publications of books and journals pertaining to chemometrics and multivariate analysis; its web address is linked to http://www.spectroscopynow.com/ (click on Chemometrics and Informatics tab). The homepage of *Chemometrics* has a URL at http://www.chemometrics.se/editorial/index.html. An introduction to chemometrics including a comprehensive overview is found at http://home.neo.rr.com/catbar/chemo/int_chem.htm

A few of the more representative papers describing important work in chemometrics over the review period include those listed herein. Chemometrics is a discipline concerned with the application of statistical and mathematical methods as well as those methods based on formal mathematical logic to chemistry. Publications concerned with development of new chemometric methods experienced modest growth during this period, no doubt

because the number of researchers actively engaged in development and publication of novel chemometrics methods has remained constant. On the other hand, the number of researchers applying chemometrics continues to grow, and the number of publications concerned with applications of chemometric methods also grew substantially. The extraction of information from chemical data continues to drive research in the field of chemometrics. Development of new methods in chemometrics and novel or important applications of these methods occurred in three major areas: calibration, resolution, and pattern recognition, which are summarized below. Topics such as feature selection, data preprocessing, signal processing, library searching, parameter estimation, and optimization are also covered in this literature survey and are treated in the context of the three major application areas that are the focus of this review.

CALIBRATION

Calibration involves relating, correlating, or modeling a measured response based on the amounts, concentrations, or other physical or chemical properties of a set of analytes. A tremendous number of papers have appeared in the last 2 years on the topic of calibration. A few of these papers focused on issues that are of interest to chemists who routinely perform calibrations in their work. For instance, is a single point calibration as good as a multiple point calibration? A comparison of bias and precision obtained in single point versus multiple point calibrations was undertaken using six data sets in a study performed by Peters and Maurer (*11*). The results obtained from a full calibration differed significantly from those obtained from a one point calibration for five of the six data sets studied. The best one point calibration results were comparable to the full calibration and occurred when the calibrator was close to the center of the full calibration range. Standards for calibrations are usually prepared by serial dilution, which often generates dilution errors. This kind of error can lead to a biased estimate. A linear calibration with correlated covariates that takes the serial dilution error into account was proposed, and an asymptotically unbiased estimator with asymptotic normality was developed as part of the study (*12*). A simulation study showed that better performance with respect to bias is obtained using this new approach. The issue of bias and its effect on measurement uncertainty in calibrations was also the subject of a review (*13*). The uncertainty of a result from a linear calibration is given by a well-known ISO-endorsed expression. Its derivation and use are explained in a paper with the approach applicable to any function that uses linear coefficients (*14*). The study provides an example of a weighted quadratic calibration for ICPAES. The author recommends repeatability of the instrumental response, rather than the standard error of the regression, be used in calculating the standard error of an estimate.

Multivariate calibration refers to the process of relating the analyte concentration or the measured value of the physical or chemical property to a measured response, e.g., near-IR spectra of multicomponent mixtures. It remains by far the fastest growing area of chemometrics as evidenced by the tremendous number of papers that have appeared in the last 2 years on partial least-squares (PLS). PLS has come to dominate the practice of multivariate calibration because of the quality of the calibration models produced and the ease of their implementation due to availability of PLS software. Latent variables in PLS are developed simultaneously along with the calibration model so that each latent variable is a linear combination of the original measurement variables rotated to ensure maximum correlation with the information provided by the property variable. Multivariate calibration has been the subject of a recent review (*15*) with special attention paid to issues that have not been thoroughly investigated by previous workers such as sample design, the number of samples necessary to obtain a reliable regression model, and the effect of noisy predictor variables on the regression model.

The effect of missing data (i.e., measurements) on the uncertainties in predictions, and residual square prediction error from PLS calibrations has been investigated (*16*). Intervals for these missing values were developed and used to assess whether or not a regression model would perform well in the presence of missing data. These intervals could also be used to determine which measurements should be recovered to achieve the greatest reduction in uncertainty. The results are demonstrated by application to process control data from a Kamyr digester. Regression based approaches are considered to be the most statistically efficient methods for estimating scores of samples with missing data. A framework has been proposed that allows a user to write these regression based methods by a unique expression function of a key matrix (*17*). From this framework, a statistical performance index is introduced as a way to predict the impact of the missing data on the estimation of the scores without requiring the use of real data. The results are shown in applications involving several industrial data sets. Another problem that often arises in PLS calibrations is confounding effects. Multilevel PLS (*18*) is proposed as a better option for analyzing data where sources of variation are confounded. The models obtained using this technique contains submodels for the different levels in the data, allowing the separation of between-run and within-run variation. Another approach to this problem is the use of experimental design techniques. A Kohonen self-organizing map has been used to select samples for training and validation, which in turn are used to operate the calibration of a multilayer feed forward neural network (*19*). The model that resulted performed significantly better than one developed using the random or Kennard−Stone selection of samples for both the training and validation sets. The effect of reference value uncertainties to a sample's specific standard error of prediction in PLS has also been investigated, and a numerical method to assess these uncertainties has been proposed (*20*). The role of cross validation in the development of reliable PLS calibration models was investigated by Westad who used PLS-2 to develop correlations between chemical structure and different regions of the electromagnetic spectrum (*21*). In the course of his study, he concluded that cross validation acted as a strong filter toward spurious correlations in the data. An inconsistency in PLS-1 has been recently reported by Ramos and Pell (*22*). Conventional PLS algorithms with orthogonal score vectors use one model space to compute the regression vector but another model space to represent the reconstructed data. The magnitude of this difference depends on the degree of truncation of the model space. By comparison, the nonorthogonal PLS-1 algorithm of Martens does not suffer from this inconsistency. Faber reported that jack-knifing, a resampling method widely used to assess uncertainty in regression coefficients, should not be used

with designed data since a basic assumption underlying all resampling methods is being violated. Faber provided support for this view through reanalysis of a literature data set (*23*).

Selecting the correct number of latent components for a PLS model continues to be a problem. Although selecting training and prediction sets using experiment design techniques is the best approach for dealing with this problem, many laboratories cannot afford to take such an approach since samples are often obtained over time in an undesigned manner. In a recent study, classical criteria such as cross validation and adjusted Wold's criterion have been compared to recently proposed alternatives such as PLS-PoLiSH and a randomization test (*24*). The recently proposed methods were shown to give reliable dimensionality predictions, which was not always the case for the so-called traditional methods which appeared to be largely affected by the largest changes in the modeling capacity of the largest PLS components. The predictive ability of principal components regression is also highly dependent on the number of latent variables selected. The popular cross validation methods, e.g., leave-one-out or Monte Carlo, are not always able to determine the proper number of latent variables, especially when so-called anomalous samples are present in the data. For data containing anomalous samples, it has been shown that ICOMP, the information complexity criterion, performs better (*25*). Overfitting in PLS has been reviewed by Faber (*26*). He has proposed a randomization test that enables the user to assess the statistical significance of each PLS component as it enters the model. When compared to cross-validation or independent test set validation, the randomization test performed as well or better and did not require the use of soft decision rules. A new approach to cross validation called moving window cross validation (MWCV) has been developed to select components rationally in a partial least-squares calibration model (*27*). This method devises a pattern to split a validation set using a number of moving windows that change synchronously along appropriate subsets of all the samples. Calculations for the mean value of all mean squares error in cross validations (MSECVs) for all splits are made for different numbers of components, and the optimal number of components for the model can then be selected. MWCV is compared to leave-one-out cross validation (LOOCV) and Monte Carlo cross validation (MCCV) for PLS models. The authors report that MWCV is useful for selecting the optimum number of factors while avoiding the usual tendency to overfit the data.

Achievement of a satisfactory multivariate calibration model is often not the final step in many practical applications. Once it is developed, it is often necessary to transfer the calibration model to other instruments or update the calibration model to ensure that the calibration can be used at the point of measurement. One way to achieve transfer of a calibration is to standardize either the instrumentation used or the calibration itself. Small (*28*) described an updating procedure to improve the robustness of multivariate calibration models for near-IR spectroscopy. With the utilization of a single blank sample, repeated spectra are acquired during the instrument warm up. These spectra are used to capture the noise profile of the instrument on the day that it is used. By augmentation of the original spectra of the training set with a group of spectra collected from the blank, an updated model can be computed that incorporates instrumental drift which may have developed. Von Stockar (*29*) addressed the problem of instrumental drift and its effect on the robustness of the calibration model by periodically injecting planned spikes of small amounts of analytes into the monitored medium. The corresponding measured difference spectra were scaled up and used as reference measurements for updating the calibration model in real time based on orthogonal projection. Barton (*30*) compared three standardization methods (direct standardization, piecewise direct standardization, and double window piecewise direct standardization) commonly used to transfer a calibration model between two instruments. The efficacy of the model transfer was evaluated based on the root-mean-square error of prediction (RMSEP) calculated using the independent prediction set samples. Results indicated that standardization using sealed reference standards was unacceptable but standardization using the prediction subset was adequate. The best results were obtained using preprocessed spectra, not raw spectra and double window piecewise direct standardization. In another study (*31*), six widely used standardization or calibration transfer methods (direct standardization, piecewise direct standardization, additive correction, multiplicative correction, slope and bias, and difference spectrum with interpolation) were evaluated for the calibration correction of a PLS model used for an online monitoring system. The results of this study showed that all of these methods required more than two samples to obtain the necessary accuracy for the nonlinearity contained in the spectral data. From the standpoint of a practical calibration in a real plant, the acceptable number of samples is one or two. Pell (*32*) investigated the transfer of calibration models between Fourier transform near-infrared (NIR) instruments using piecewise direct standardization and prediction augmented classical least-squares/partial least-squares. Although the RMSEP values for calibration transfer were within acceptable range for the two methods, the prediction augmented classical least-squares/partial least-squares method was better at preserving the outlier detection capabilities than piecewise direct standardization when a subset of samples is used to define the transfer function for the calibration. In another study, calibration transfer was used to correct for drift in a portable quadrupole mass spectrometer (*33*). In the absence of calibration transfer, quantification of test spectra was inaccurate by more than an order of magnitude. By application of a calibration transfer strategy across all measurements, errors that previously occurred in prediction were not observed until 12 months later.

During the last 2 years, fundamental work on calibration began to be refined in a number of aspects. Special focus was given to the type of preprocessing methods applied to the data prior to the development of a calibration, which can influence the performance of the model. Felizardo (*34*) investigated the effect of some commonly used preprocessing methods applied prior to PLS or principal component regression. The calibration problem involved the use of NIR to determine the amounts of methanol and water in biodiesel. The results confirm the importance of evaluating various preprocessing techniques when developing a calibration. Fernandez-Cabanas (*35*) investigated the effect of 49 different combinations of data pretreatments (first and second derivatives, autoscaling, detrending, and two version of multiplicative scatter correction) on calibration models developed from NIR spectra for ground feedlots. Although the validation statistics used to evaluate the performance of the calibration models did not

reveal a clear distribution pattern, some pretreatment combinations consistently provided better results. Light scattering effects pose a problem in the estimation of analyte concentrations from particulate systems such as blood, tissue, and pharmaceutical solids. Martens and co-workers (*36*, *37*) have proposed an extended multiplicative signal correction (EMSC) approach where light scattering effects are taken into account in an empirical manner. One of the advantages of EMSC is that it is possible to utilize causal mathematical models based on the physics of light scattering in this framework, which can lead to further improvements in the separation of absorbance and scattering effects. A new algorithm to address the effect of multiplicative light scattering on NIR spectra called optical path length estimation and correction has been described by Martin (*38*). This methodology, which has been shown to be better than extended inverse signal correction, was successfully validated using two data sets from the literature. In the past 2 years, wavelets have also been investigated as a preprocessing method for NIR data because of their ability to simultaneously accomplish two essential needs in multivariate calibration: data compression and signal correction. Data compression can lead to more parsimonious models and signal correction can improve the quality of the regression models developed. The advantages of using wavelets to preprocess NIR spectra were demonstrated in two published studies (*39*, *40*).

Applications of PLS regression dominated the literature. Use of PLS has become commonplace in analytical chemistry, and applications are appearing in very distant fields. A major part of the increase in the use of PLS can be attributed to improving commercial software for chemometrics, but better education of chemists in the use and application of multivariate calibration also appears to have a role. Vibrational spectroscopy has long been an area where chemometric methods are embraced, and it is no surprise that many of the applications appeared in analyses using near-IR spectroscopy. Analysis of glucose in various aqueous media has been the subject of attention from many groups. The lower levels of glucose found in physiological samples and the high background presents a special challenge in both spectroscopy and calibration. A spectrum simulation method is described for use in the development of calibration models for glucose using near IR spectra (*41*). Synthetic spectra are computed from background spectra collected from the spectrometer for which a calibration model is desired and from previously measured molar absorptivies and solvent displacement factors. The synthetic spectra are used with PLS to form the calibration model. This methodology is demonstrated in the analysis of physiological levels of glucose in an aqueous matrix containing variable levels of alanine, ascorbate, lactate, urea, and triacetin. Experimentally measured data from two different instruments with different noise levels are used to validate the simulation approach. With the more stable instrument, well performing calibration models are obtained, which is not the case with the less stable instrument. Simulation approaches have also been used to develop PLS calibration models for noninvasive blood glucose monitoring (*42*, *43*). A numerical simulation of light propagation in skin tissue was used to obtain simulated NIR diffuse reflectance spectra, which in turn allowed the workers to obtain parameters affecting the prediction of blood glucose levels by a calibration model. The calibration model obtained through the numerical simulation had a characteristic peak at the wavelength around 1600 nm, which corresponds to the characteristic absorption band of glucose. Through judicious design of the calibration, the workers were able to prevent the regression model from tracking chance temporal correlations that are often observed in conventional studies using NIR spectroscopy to monitor blood glucose levels. Multivariate curve resolution can be combined with PLS to develop calibration models using minimal reference data. By application of evolving factor analysis and orthogonal projection, initial estimates of the concentration and spectral profiles for the intermediates and products can be obtained. Further optimization by ALS can lead to refined estimates of the concentration profiles. The concentration profiles and the processed spectra can then be used to develop calibration models. Two studies, one published in 2006 and the other in 2007, illustrate this approach. In one study, a PLS-2 calibration model was developed from UV/attenuated total reflectance measurements to determine the end point of a chlorination reaction (*44*). In the other study, the use of fiber optics for vitro dissolution testing of gilbenclamide tablets was investigated. The application of PLS and multivariate curve resolution solved the problem of interfering absorbance of excipients and made it possible to obtain dissolution rate profiles and spectra of both the gelatin capsule and the gilbenclamide (*45*). Temperature fluctuations can have an impact on the precision of spectral measurements adversely affecting the resulting calibration model. Current methods for addressing this problem can be divided into two broad categories: calibration model based approaches and spectral standardization methodologies. A comparative study on a number of strategies reported in the literature was undertaken, and it was observed that a global modeling approach where latent variables are extracted from the spectra using PLS and then augmented with temperature as an independent variable achieved the best results (*46*).

Improving the methods for multivariate calibration themselves continues to be an active area of research in chemometrics. A hybrid multivariate calibration method called constrained regularization was developed and tested using simulated and experimental Raman spectra. In this method, a balance between model complexity and noise rejection is achieved by inclusion of a priori information as a spectral constraint, which is incorporated into the algorithm in a flexible manner (*47*). Constrained regularization when properly configured was shown to be superior to PLS and less susceptible to spurious correlations. Several groups offered improvements to existing PLS algorithms. The effects of ensemble learning methods, such as bagging and boosting, on kernel PLS regression was investigated (*48*). Using two near-IR data sets, the authors demonstrated that bagged kernel PLS and boosting kernel PLS performed better than standard PLS. Subagging, based on subsampling without replacement, has been shown to provide improvements to PLS models, without the high computational burden associated with bagging. A strategy for implementation of subagging in PLS has been proposed (*49*). The subagging member models are generated by subsampling the pool of samples available for modeling and using them to develop new calibration sets. Subagging is of value in analytical problems involving complex matrices where reproducing their compositional variability for the purposes of implementing an experimental design is not a viable alternative. Subagging was shown to improve the prediction accuracy of PLS models with improvements in accuracy varying from 16% to 35%. Support vector machines (SVM) continue to receive attention as an alternative to PLS because of

their ability to model nonlinear spectral property relationships in data. In one study, SVM was used to quantify some common adulterants in powdered milk (starch, whey, and sucrose) using near-IR spectroscopy with direct measurements by diffuse reflectance (*50*). The calibration models built using SVM were superior to PLS. In another study, the possibility to visualize and interpret the information contained in a SVM model was investigated. Analogous to the score and loading plots which make it possible to understand the driving force underlying PLS, Buydens claims that it is possible to turn a SVM regression black box into a transparent and interpretable modeling technique (*51*).

Higher order calibration continues to be explored, and analytical applications are becoming more commonplace. Ortiz who has provided a review of second order calibration techniques during this period states that techniques for calibrations involving three way signals are sufficiently developed for use in routine analyses (*52*). The ability to quantify an analyte in the presence of interfering agents and to extract the signal corresponding to the analyte of interest makes second and higher order calibration techniques useful for the identification and quantification of analytes in complex samples. A new approach to higher order calibration based on the combination of multiway partial least-squares and a procedure called residual trilinearization (N-PLS/RTL) is described (*53*). N-PLS/RTL was applied to kinetic fluorescence−excitation emission four way data to determine procaine and its metabolite, *p*-aminobenzoic acid, in equine serum. The proposed algorithm was able to predict the concentration of these two analytes in the presence of uncalibrated components despite strong linear dependencies and the loss of multilinearity in the data since one of the analytes also served as the reagent and the other was the product of the hydrolysis reaction, which was studied by fast scanning fluorescence spectroscopy. An influence function for tri PLS1 regressions was proposed by Geladi (*54*) to assess the influence of individual samples on both calibration and prediction. With the use of this influence function, a new estimate for prediction uncertainty was obtained. Available multivariate methodologies for exploiting the second order advantage of three way data are based on linear models, which are not applicable if the concentration of the analyte is related in a nonlinear manner to the spectral information. By combination of a back-propagation neural network with the technique of residual bilinearization, models that can extract analyte concentration from nonlinear data even in the presence of unsuspected components can be developed. Successful predictions of analyte concentration in samples containing unsuspected components were achieved in large data sets (*55*). The new method was also able to recover the contribution of the unsuspected component to the total test sample signal. The new method performs better than multivariate methodology based on PLS regression with second order advantage for the nonlinear data sets investigated.

Higher order calibrations often require long computation times and the handling and storage of very large data sets. Accelerating the computation time by reducing the storage requirements for multiway analyses through a data preprocessing method based on multidimensional wavelet transforms which enables a highly efficient compression of the data prior to evaluation by PARAFAC was investigated by Booksh and Vogt (*56–59*). For three-way and four-way data, computation times were reduced by a factor of 50 without loss of accuracy or interpretability of the models derived, which were in good agreement with PARAFAC models developed using the entire data set. A variety of wavelets were investigated including hybrid wavelets which use different wavelet types for different dimensions of the data. For 19 of 20 studied cases, which included hyperspectal and excitation−emission matrix fluorescence data sets, hybrid wavelets performed better than conventional wavelets.

Because of the advantage associated with higher order calibrations, accurate predictions of analyte concentrations in new samples are possible even if a sample contains components not taken into account by the calibration model. Although the calibration step in second order calibrations has been the subject of numerous studies, factors that influence the prediction quality of calibration models developed using multiway methods such as PARAFAC have not been studied in any great detail. Using both simulated and real data sets, Bro (*60*) has investigated the size of the calibration set, the number and degree of overlap of uncalibrated components, and the type and magnitude of the noise on the prediction quality of the calibration model. Synovec (*61*) describes a method for the automated selection of PARAFAC models with the appropriate number of factors for two-dimensional gas chromatography/time-of-flight mass spectrometry data. The approach taken involves increasing the number of factors in the model until the mass spectral matching of the loadings against a target analyte spectrum is indicative of overfitting. Hyperspectral imaging data have a number of errors that can be corrected by second order calibration models. Geladi has shown that multiple Spectralon calibration standards can correct for both spectral and spatial variations in the data with optimal results achieved using a two-step calibration and correction process (*62*, *63*). The advantages of using multiway methods for analysis of three-dimensional gas chromatographic data as demonstrated by Synovec (*64*) include a 10-fold improvement in signal-to-noise ratio relative to traditional integration methods and an enhancement in the effective three-dimensional peak capacity due to the deconvolution of overlapping spectral responses provided by PARAFAC. Rayleigh scattering and Raman scattering complicate the modeling of fluorescence excitation−emission matrix measurements by PARAFAC. Eliminating these scattering effects in the data ensures that satisfactory models are produced. Bro (*65*) has used interpolation in areas affected by first and second order Rayleigh and Raman scatter to remove the interfering signal. Bouveresse (*66*) has applied independent component analysis on the unfolded cubic array with the independent components related to Rayleigh and Raman scattering identified and removed prior to reconstruction of the excitation−emission fluorescence data cube. Both approaches yielded satisfactory models for the data. To address inner filter effects in fluorescence excitation−emission matrix data, Gil (*67*) has utilized a second order multivariate calibration approach called unfolded partial least-squares with residual bilinearization. Unfolded partial least-squares, multiway PLS, parallel factor analysis, self-weighted alternating trilinear decomposition, and bilinear least-squares were compared using excitation−emission fluorescence data obtained in a study to simultaneously determine mefenamic, flufenamic, and meclofenamic acids in urine samples (*68*). Parallel factor analysis, self-weighted alternating trilinear decomposition, and bilinear least-

squares, which can exploit the second order advantage inherent in a data cube performed consistently better than unfolded partial least-squares and multiway PLS for the excitation–emission fluorescence data.

RESOLUTION

This section is concerned with methods for the resolution and recovery of pure-component spectra from the overlapped spectra of mixtures. A review on the use of factor analysis and multiway methods in chromatography applied to the peak purity problem and to the resolution of overlapped chromatographic bands as well as its extension to simultaneous analysis of multiple runs to obtain both qualitative and quantitative information was published by de Juan (*69*). A simulated annealing algorithm for decomposition of linear mixtures using a Metropolis-type Monte Carlo search with mutual information from recovered components as a cost function and their nonnegativity as a hard constraint has been proposed (*70*). The performance of the simulated annealing algorithm is reported to be superior to methods based on principal component analysis such as SIMPLISMA or ALS. Maximum likelihood methods have been applied to Raman spectra of mixtures to estimate their intrinsic dimensionality and the corresponding number of chemical components in the mixture (*71*). Even when the signal-to-noise ratio of the Raman data was low, accurate estimates of the number of components in the mixture could be obtained (including minor components) by smoothing the data before applying the maximum likelihood estimator. Another advantage of using this maximum likelihood estimator was that it is computed locally at every data point, which offers the advantage of allowing the user to segment the sample specimen into homogeneous regions. Statistical heterospectroscopy, a new paradigm for the simultaneous analysis of multispectroscopic data sets, was introduced for the identification of biomarkers (*72*). The potential of this technique was established by the simultaneous analysis of proton NMR and liquid chromatography/mass spectrometry data, which demonstrated that direct cross-correlation of spectral parameters by way of chemical shifts from NMR and m/z data from mass spectra, and is readily achievable for a variety of metabolites. Band target entropy minimization, a spectral reconstruction method that elucidates both the pure component spectra and their corresponding concentration profiles without a priori information, has been applied to a variety of curve resolution problems in two-dimensional correlation spectroscopy (*73*), excitation–emission fluorescence matrix spectroscopy (*74*), and reaction monitoring using infrared emission (*75*) and FT-IR and UV–visible absorbance (*76*).

Determining the feasible region of potential solutions for multivariate curve resolution problems continues to be an active area of research. Rajko (*77*) discusses the use of computational geometry tools to draw Borgen plots of any three component system. With the use of this methodology, the highly cited and used data set of Lawton and Sylvestre is shown to consist of more than two components. The effects of noise on the rotational ambiguity in model free analyses of multivariate data were investigated by Tauler (*78*). Because of the problems associated with noise on rotational ambiguity, Tauler has proposed a new technique for self-modeling curve resolution called resolving factor analysis. Alternating least-squares continues to be the most widely used technique to solve multivariate curve resolution problems. A new nonlinear optimization algorithm based on nonlinear constraints has also been proposed by Tauler to identify appropriate rotations and perturbations of subspaces defined by solutions based on principal component analysis (*79*). The concentration and spectral profiles obtained by alternating least-squares and the new proposed algorithm are similar and are both within the boundaries of feasible solutions for this problem.

First order analytical data have mainly been resolved by principal component analysis based techniques. A modified curve resolution technique utilizing singular value decomposition was directly applied to reflectance spectra of two different soil types, each coated with di-BU phosphate (*80*). The results provided interpretable spectra for the detection and classification of organic analytes adsorbed on soil and were consistent with those previously obtained using an extended multiplicative scatter correction. Principal component analysis and alternating least-squares combined with a kinetic modeling strategy have been used to determine rate constants for a curing reaction of epoxy resins (*81*). The recovered concentration profiles are fitted to a chemical model proposed for the reaction based on information contained in the recovered profiles. A new approach for determining the shelf life of industrialized food products, the multivariate accelerated shelf life test, was proposed in which principal component analysis scores are used to estimate the multivariate rate constant, the multivariate acceleration factor, and the multivariate activation energy (*82*). The method was successfully applied to an industrialized tomato product. The structural evolution of isotactic polystryrene during the cold crystallization process was studied using FT-IR and multivariate curve resolution (*83*). Orthogonal projection, alternating least-squares, and fixed-size moving window evolving factor analysis were used to interpret the spectral changes that occurred in the investigated chemical process. Principal component analysis and alternating least-squares methodology was applied to near IR spectral data for an esterfication reaction (*84*). Spectral and concentration profiles were estimated with the pure spectra of the components and the concentration values of the acid used as soft equality constraints. Good results were obtained for the prediction of the acid value and the hydroxyl value based on the RMSE values for these constituents. The use of principal component analysis as a straightforward approach to obtain a good estimate of relative response factors for small impurity peaks in HPLC–diode array chromatograms without knowledge of the molar absorptivities and without any precalibration was demonstrated by Wiberg (*85*).

Resolution applied to higher order data (second and third order), with a special emphasis on chromatographic techniques hyphenated to spectroscopic techniques remain the main source of published papers during this period. There were too many applications of multivariate curve resolution applied to higher order data to cite here. However, representatives of those publications are summarized in the following paragraph. These were selected on the basis of a novel or important application, preprocessing, or an unusual measurement system. Two-dimensional liquid chromatography with diode array detection generates data sets on the order of 10 million data points. Using the techniques of window target testing factor analysis and parallel factor analysis/alternating least-squares with flexible constraints,

Rutan and co-workers (*86*) have shown that it is possible to resolve rank deficient data and to obtain quantitative information from 95 peaks in a maize seedling digest, focusing on compounds related to the biosynthetic pathways of indole-3-acetic acid, the primary growth regulator in plants. A novel implantation of an alternating least-squares algorithm for resolving three- and four-way data using computer simulated multiway data as well as multiway data obtained in typical kinetic experiments with liquid chromatography and diode array detection has been proposed (*87*). The new multiway ALS algorithm provides estimates of the spectral profiles for each of the components involved and simultaneously estimates the rate constants for the reaction steps at different experimental conditions. Multivariate curve resolution techniques applied to three-dimensional spectral matrices obtained from synchronous fluorescence spectra of fulvic acids at different pH value and at different concentrations of the fulvic acids were used to estimate spectral and fluorescence intensity profiles for the detected components (*88*). Some of the computed fluorescence intensity profiles exhibited a shape similar to acid−base species distribution diagrams, which allowed pK_a values to be estimated for different fulvic acids. An evolving factor analysis algorithm based on trilinearity that is intrinsic in many three-way data arrays has been developed, which allows the user to obtain a rank map of three-way chromatographic data (*89*). Selective regions in the chromatogram characteristic of a specific component can be determined using the rank map, which in turn provides the possibility of resolving profiles of individual components one at a time. A new method for the analysis of two-dimensional chromatograms based on finding a new pair of axes to represent the data has been proposed (*90*). If one of the axes captures all of the variance for a specific family of compounds, a decrease in the matrix rank will occur. The rotation necessary to achieve this desired effect is implemented by introducing retention models in both chromatographic dimensions to describe the behavior of the compounds that constitute a single family. After this transformation, a family of compounds is associated with a single source of variation which is captured by a single axis in the data. This approach has been demonstrated using the separation of families of functional poly(methyl methacrylate) polymers with different numbers of hydroxyl end groups and different degrees of polymerization.

PATTERN RECOGNITION

The overall goal of pattern recognition is classification. Developing a classifier from spectral, chromatographic, or compositional data may be desirable for any number of purposes including source identification, detection of odorants, presence or absence of disease in a patient or animal from which a sample has been taken, and food quality testing to name a few. The classification step is often accomplished using one of several techniques that are now fairly well established including principal component analysis, hierarchical clustering, k-nearest neighbor, statistical discriminant analysis, and soft independent modeling by class analogy (SIMCA). Few novel pattern recognition methods were published during the past 2 years. Instead the chemical literature on pattern recognition focused on novel and not so novel applications. Nevertheless, classification of data remains an important subject in chemometrics as evidenced by the large number of citations in the *Chemical Abstract* database on pattern recognition applications during this recent review period which were only rivaled by calibration. Hence, most of the references in this section are organized according to the type of application. However, there were papers published by research groups in this period that focused on improvements in the methods used for classification. During the last 2 years, fundamental work on classification began to be refined in a number of aspects with special focus given to data preprocessing.

A large number of publications have appeared in the chemical literature on the practical aspects and implications of preprocessing chromatographic and spectroscopic data to correct for undesirable time-shifts. For a successful pattern recognition study, it is essential that features encode the same information for all samples or objects in the data set. If variable 3 is the area of a gas chromatographic peak for acetaldehyde in sample 1, it must also be the area of the GC peak for acetaldehyde in the other samples that comprise the data set. Hence peak matching is crucial when chromatograms or spectra are translated into data vectors. Kaliszan (*91*) compared the performance of three peak alignment procedures for preprocessing electrophoretic nucleoside profiles. The three warping procedures investigated, dynamic time warping (DTW), correlation optimized warping (COW), and parametric time warping (PTW), were examined using two sets of electrophoretic data. The warping methods investigated enabled differentiation between electropherograms of healthy and cancer patients by principal component analysis. In another study, PTW was shown to be the easier and faster to use than DTW and COW (*92*). When COW and DTW are used, peak shifts in both directions can be corrected but optimization of two input parameters is required which is not an easy task in COW. However, the improvement in precision of peak retention times after alignment justifies the use of these more complex algorithms. An extension of DTW termed ordered bijective interpolated warping (OBI-Warp), which attempts to address the deficiencies of DTW by coupling it to a piecewise cubic hermite interpolation to produce a smooth warping function has been proposed (*93*). The performance of the algorithm was demonstrated using LC−MS data. From their study, the authors concluded that Pearson's correlation coefficient as a measure of similarity outperformed covariance, dot product, and Euclidean distance in its ability to produce correct alignments when optimal or suboptimal alignment parameters are used. Piecewise linear functions can be used to correct for variations in mass and elution time measurements in LC−MS or GC/MS data (*94*, *95*). A method for aligning chromatographic data by identifying reference peaks present in all of the chromatograms, calculating the amount of retention time shift between chromatograms needed to align these peaks using a cubic spline interpolation, and the subsequent shifting of the other peaks by the application of this function has been proposed (*96*). An approach to automate the alignment of chromatographic data using a discrete coordinate simplexlike optimization routine is presented with emphasis on the practical aspects including the selection of the required parameters and the reference samples used for the matching (*97*). The advantages of this approach are demonstrated using simulated GC and HPLC data. A two-dimensional peak detection algorithm is described by Jacobsson, which exploits the fact that in high resolution LC−MS data, peaks emerge flanked by data voids in the corresponding mass axis (*98*).

The method is simple and only requires a priori knowledge of the minimum chromatographic peak width, which is a system dependent parameter.

Scaling of spectral data is often crucial in the development of a classifier. However, the procedures that should be employed in a particular study are highly dependent upon the nature of the problem under investigation and the goals of the analysis. McNaughton has compared the performance of six spectral preprocessing techniques on the classification of Raman spectra of microalgal cells by their nutrient status (*99*). The preprocessing techniques investigated included baseline correction with vector normalization, multiplicative scatter correction, extended multiplicative scatter correction, standard normal variate, and vector normalized first and second derivative spectra. With the use of PLS, all of these preprocessing techniques allowed differentiation of Raman spectra of nutrient replete and nutrient starved algal cells with derivatives and extended multiplicative scattering correction performing the best. However, SIMCA was not able to differentiate the Raman spectra by nutrient class due to the small model distances involved. On the other hand, similar results were obtained when a comparison of methods for baseline correction of Raman spectra of solid samples of illegal narcotics diluted with various materials using principal component analysis to assess the performance of the preprocessing techniques investigated (derivative preprocessing and a new polynomial method described by Lieber and Mahadevan-Jansen) was undertaken (*100*). An algorithm for the removal of the fluorescence background from a Raman spectrum based on a modified multipolynomial fitting of the data with a peak removal procedure during the first iteration was proposed (*101*). The algorithm was validated using real time Raman spectroscopy and in vivo measurements that are characterized by low signal-to-noise. An algorithm for removing cosmic spikes from Raman spectra obtained from dispersive instruments equipped with charged coupled detectors was described (*102*). The algorithm is easy to implement, computationally efficient and is shown to generate recovered spectra with negative distortion. The performance of four data preprocessing methods in the presence of different types of noise was evaluated using LC−MS data obtained from the analysis of a pharmaceutical drug and its degradation product (*103*). The preprocessing methods investigated were the component detection algorithm (CODA) and three types of digital filters: matched filtration, Gaussian second derivative, and Savitzky−Golay. In the case of random white noise, matched filtering performed the best in accordance with theory. By comparison, no improvement was observed with any of the preprocessing methods studied when heteroscedastic noise was present in the data.

There appears to confusion as to the need for and the role of preprocessing acquired spectra in both the proteomics and metabolomics literature. A number of studies have been performed in the past 2 years to better understand the role of data preprocessing in the context of the particular analysis problem undertaken. Two studies on this subject are described here in. In one study, Kvalheim used factorial design with the ratio of intergroup to within group differences as the response variable to investigate the effects of smoothing, binning, noise structure, and normalization on the interpretation and classification of mass spectral data (*104*). From this study, Kvalheim concluded that mass spectral profiles have to be corrected for heteroscedastic noise prior to normalization and strong interactions exist between several of the pretreatment steps, e.g., noise reduction and normalization, which means that different pretreatment steps cannot be considered in isolation. Box car averaging (or binning) can serve as a substitute for smoothing of data by Savitsky−Golay filtering while at the same time reducing the dimensionality of the data. In the absence of smoothing, box car averaging should be performed before peak alignment. If box car averaging is not performed, the order of data pretreatment should be smoothing, peak alignment, *n*th root transform of the data to eliminate heteroscedastic noise, and normalization. In another study, problems that are associated with bucketing 1H NMR spectra derived from urine are discussed (*105*). The authors suggest definitions for and discuss preprocessing operations crucial to extracting information from NMR data. They conclude that care must be taken when bucketed data is normalized to constant sum.

Improving the methods used for pattern recognition continues to be an active area of research in chemometrics. Several groups offered new algorithms for visualization of multivariate chemical data. Melssen and Wehrens (*106*) introduced two alternative Kohonen neural network configurations which combine the transparency of counter propagation neural networks with the modeling power of a Kohonen neural network. Both network configurations performed better than a traditional Kohonen neural network or a counter propagation neural network. Smilde (*107*) introduced Grey component analysis (GCA) as a new exploratory data analysis method, which uses a soft penalty approach to push the eigenvector decomposition in the direction of previous available information about the data analysis problem. GCA can be used to confirm the validity of prior information or can work in an exploratory mode to study new phenomena in great detail. Rousseeuw (*108*) describes an algorithm based on his FASTICA method to robustify independent component analysis. This modification appears to be efficacious based on its performance with simulated and real data sets. Two way moving window principal component analysis, which considers all possible spectral regions, by using variable and sample moving windows is proposed as a new spectral data classification method (*109*). An advantage of the proposed method is its ability to identify the optimal spectral region for classification. The moving window method has performed better than principal component analysis or SIMCA on the classification problem involving visible-near IR spectra of mastitic and healthy udder quarters of cows. Detailed work comparing PLS with canonical correlation (CA) and the differentiation of these two methods from OPLS was undertaken by Trygg (*110*). He has reported that PLS is unidirectional, while canonical correlation is bidirectional. Canonical correlation can be used for prediction in a bidirectional fashion. That is, it can be used to predict Y from X and X from Y. PLS, on the other hand, predicts Y from X and is not intended to predict X from Y because PLS only models X from X and not Y from Y. However, canonical correlation, which is a regression based method, suffers from well-known regression based problems. OPLS, on the other hand, has "the best of both worlds." The idea of OPLS is to separate the systematic variation in X into two parts: a Y-related part and a Y-orthogonal part. The Y-orthogonal part may be part of the model but is useless for prediction of Y. Its main benefit is model

interpretation and pure profile estimation, while maintaining the predictive ability of the PLS method. Two papers have been published describing the application of this method to spectroscopy in process monitoring of pharmaceuticals (*111*, *112*).

Applications of pattern recognition methods dominated the literature. Lavine (*113*) used the wavelet packet tree and a genetic algorithm for pattern recognition to analyze differential mobility spectra for structural content by chemical family. The wavelet packet transform denoised and deconvoluted the mobility spectra by decomposing each spectrum into wavelet coefficients, which represents the samples constituent frequency. Wavelet coefficients characteristic of the compounds structural class were identified by a genetic algorithm (GA) for pattern recognition. The pattern recognition GA used both supervised and unsupervised learning to identify coefficients which optimize clustering of the spectra in a plot of the two or three largest principal components of the data. Because principal components maximize variance, the bulk of the information encoded by the selected coefficients is about differences between chemical families in the data set. The principal component analysis routine embedded in the fitness function of the pattern recognition GA acts as an information filter significantly reducing the size of the search space since it restricts the search to coefficients whose principal component plots show clustering on the basis of chemical family. Brereton (*114*) used tandem mass spectrometry and pattern recognition methods to characterize the contamination of banknotes by cocaine. Banknotes obtained from defendants who are in close proximity to cocaine will differ from banknotes obtained from the general population in terms of cocaine contamination. With the use of principal component analysis and class modeling via the Mahalanobis distance to analyze mass spectral data, it was possible to discriminate between the two classes of bank notes. Principle coordinate analysis and gas chromatography/mass spectrometry was used to characterize human sweat obtained from an isolated population of 200 individuals in Carinthia for the purpose of identifying individual fingerprint patterns in the data indicative of the family of the donor (*115*). A unique facet of this study is that it was necessary to study a large number of small groups instead of small number of large groups, and different approaches were required to analyze the data because of the attendant problems associated with the analysis of sparse data matrices. Canonical correlation analysis of volatile organic compounds analyzed by thermal desorption gas chromatography/mass spectrometry has been used to differentiate expired air samples from entrapped people in a collapsed building from other sources of volatile organic compounds including healthy or fasting humans, waste disposal bins, air from bags with decaying humans, and urban air (*116*). Classification of water as to its port of origin using PARAFAC-SIMCA analysis of excitation−emission matrix measurements of color dissolved organic matter provides a forensic tool to enforce ballast water exchange regulations which prevent introduction of nonindigenous aquatic nuisance species in the United States (*117*).

The combination of sensors and pattern recognition continues to be an active area of research. Principal component analysis of a library of digital images of a colorometric sensor array prepared from 25 chemically responsive dyes printed on a hydrophobic membrane and immersed in different soft drinks provided facile identification of all the soft drinks with a misclassification rate of less than 2% (*118*). A novel preprocessing method for an array of differentially tuned chemical sensors exposed to analytes dispersed in naturally turbulent chemical plumes from a point source provided a stable response fingerprint representative of the analyte generating the response (*119*). Rose-Pehrsson (*120*) used a probabilistic neural network to classify cyclic voltammograms of gaseous mixtures for sulfur compounds using a sensor array containing four cermet electrodes. Prior to pattern recognition analysis, each raw voltammogram was background subtracted using clean air and then concatenated to form a vector of points which were then compressed by wavelet transformation. Partial least-squares discriminant analysis and SIMCA pattern recognition were used to classify surface enhanced Raman spectra of chemical and biological simulants collected on multiple substrates fabricated from colloidal gold adsorbed onto a silanized silica quartz surface (*121*). The use of the pattern recognition methods demonstrated both the feasibility and limitations of this technique for the identification of known but previously unclassified spectra.

A few papers focused on the use of pattern recognition methods in the medical and biological sciences. Hierarchical clustering and linear discriminant analysis was used to classify Raman images of nontumor and tumor bladder tissue collected from 15 patients (*122*). Colored Raman maps of the data were created in which each color represented a cluster of spectra measured on tissue areas of similar composition. For each cluster, the cluster average spectrum was calculated and labeled as tumor or nontumor in accordance to pathohistology. Classification of these spectra using pattern recognition techniques yielded 93% correct classification with 94% sensitivity and 92% specificity. Principal component analysis of Raman spectra from fingernails and toenails provided discrimination between these two very similar biomaterials (*123*). Linear discriminant analysis provided a classification accuracy of 95% for differentiating toenails from fingernails. Linear discriminant analysis was used to differentiate infrared attenuated total reflectance spectra obtained from atherosclerotic and normal rabbit aorta samples (*124*). The successful classification of this data reveals the potential of using spectroscopy combined with multivariate classification for the identification of normal and atherosclerotic aorta tissue for in vitro and in vivo applications. Discriminant analysis and principal component analysis were used to classify Raman spectra of bacterial cells (*Escherichia coli* and *Staphylococcus epidermidis*) as viable or nonviable (*125*). Nonviable bacterial cells were inactivated by different chemical and stress conditions including starvation, EDTA, and high temperature. The overall classification success rate obtained for the data was 87%, which is impressive in view of the different treatment types and viable cell lines used. With the appropriate modification, Geladi et al. have shown that principal component analysis can be applied to electrical impedance data and provide more information then simple juxtaposition or Argand diagrams (*126*). They demonstrated this in a study in which different concentrations of Magainin and Gramicidn A gave different responses on lipid monolayers in different electrolyte solutions.

Another active area of research in pattern recognition is data fusion. In the analysis of complex mixtures, data from one method might not suffice for classification when using only a single

analytical technique. A comparison of three different methods of data fusion on two data sets was performed, and the possibilities and difficulties associated with these three methods were discussed (*127*). The methods in question were concatenated data, full hierarchical modeling, and batch modeling. The first two approaches were found to be the most successful. A new midlevel data fusion method based on the wavelet transform was proposed to process Raman and X-ray fluorescence spectra (*128*, *129*). Each spectrum set is divided into blocks according to frequency with the blocks that contain background or noise removed. Variable selection is then performed on the remaining blocks to extract the most informative variables, which are then concatenated to form a meta-signal ensemble. Classification of the ensemble data vectors is then performed using partial least-squares discriminant analysis. This system, which has been applied to real spectra of ancient pigments obtained from art work, can be automated. Brown (*130*) has proposed a method to improve classification by taking advantage of the fact that wavelets can transform chemical data into components called scales which differ in their underlying frequency or signal content. These scales will contain different amounts of information about the classification problem, and the scales can be selectively combined to yield a classifier with better performance than that of a classifier trained on the original data or developed on any individual scale.

There were a large number of publications in the past 2 years on the classification of second and higher order data. Using PARAFAC2 to analyze GC/MS data, Hibbert (*131*) has shown that it is possible to classify weathered diesel oils as to type. The retention time shift in chromatographic peaks between runs was tackled using the specific structure of the PARAFAC2 model. Daszykowski (*132*) applied wavelets to denoise gel electrophoretic images followed by fuzzy warping of features extracted from the electrophoretic data to align the gel images. Kernel PLS methods were then used to classify the data. Discrimination of wines based on 2D-NMR data was performed using learning vector quantization neural networks (*133*). Orthogonal signal correction was used as a data preprocessing method to remove from the data matrix information not correlated to the sample class label. Trygg (*134*) investigated the multivariate analysis of congruent images. Each image represents one object, and the data set contains a set of congruent images. Alternatively, the set of images may not be fully congruent but can be made so through the use of wavelet analysis and the distribution of the wavelet coefficients which form a set of congruent vectors amenable to multivariate data analysis. Naumann (*135*) has proposed an improved method of image segmentation of IR microspectroscopic data obtained from histological specimens provided by diseased and presumed normal donors. The database of colon microhyperspectra was used to train and validate a multilayer perceptron artificial neural network model. A differential wavelet based data smoothing algorithm has been combined with a fuzzy clustering algorithm for the classification of Raman spectral images (*136*). The preprocessing of the data is facilitated by decomposing the images in the wavelet domain where the discrimination of signal from noise is readily performed. Practical approaches for the application of pattern recognition methods to 2-DE protein patterns are demonstrated using three independent strategies for image analysis for alignment and matching of images, which allowed the successful application of multivariate analysis techniques to the data (*137*). Alignment of features between images is a problem that has limited the application of pattern recognition techniques to higher order data. Potential solutions to this problem have been recently proposed in the literature (*138*, *139*).

Several articles appeared in the literature during this period that did not fit in any of the categories or divisions within this section but are noteworthy for their methodology. Brown (*140*) investigated the transfer of multivariate classification models between laboratory and process near-IR instruments for the discrimination of whole coffee beans. A modified version of slope/bias correction, orthogonal signal correction trained on a vector of class identities, and model updating performed well in the preprocessing of the data to permit the transfer of a classification model developed from one data from one instrument to be used on another instrument. Classification of photochemical and metabolic reactions by random forest and Kohonen self-organizing maps was achieved with up to an 84% accuracy for a data set of metabolic reactions catalyzed by transferases and 93% for a data set of photochemical reactions by taking as input the difference between the ^{1}H NMR spectra of the products and the reactants (*141*). The covariance matrix computed from the retention time−ion abundance data matrix from GC/MS data of ignitable liquids has been shown to be a useful tool for automated searching of ignitable liquid databases (*142*). The correlation coefficient, which has been used by workers to match an unknown to a spectrum in a library, is not amenable to a Student's *t* test which can calculate the probability that two spectra are not related because the spectra are not normally distributed variables and as such do not fulfill the requirements for a statistical interpretation of the correlation coefficient (*143*). An intensity based probability function to identify peptides from mass spectra and amino acid sequence databases based on the central limit theorem and explicitly dependent on the cumulative product ion intensities, number of product ions of a peptide, and the expectation value of the cumulative intensity has been proposed (*144*). The central limit model performs better than other library metrics when matching low-quality tandem mass spectra, where the number of shared peaks is insufficient to correctly identify a peptide. For high-quality mass spectra, Poisson and hypergeometric models performed better. An adaptive multivariate statistical process monitoring (MSPC) approach has been proposed to routinely monitor a process subject to operating condition changes (*145*). Samplewise and blockwise recursive methods were used for updating a weighted mean and covariance matrix. A new model was derived recursively by utilizing the updated mean and covariance structures with the current model. Two monitoring metrics were described: the Hotelling's T^2 and the Q-statistic. The paper demonstrates the effective use of these metrics as well as a method for calculating and updating their control limits. The updating scheme proposed is robust in reducing the false alarm rate encountered when using monitoring charts, and it reduces the model sensitivity to the effects of outliers. The adaptive MSPC approach has been shown to be effective for monitoring processes where changes are either fast or slow.

Barry K. Lavine is an Associate Professor of Chemistry at Oklahoma State University in Stillwater, OK. He has published approximately 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.

Jerome (Jerry) Workman, Jr. is Director of Measurement Systems at Luminous Medical Inc., in Carlsbad, California. This Chemometrics review article constitutes the fourth in this series he has coauthored. In his career, Workman has focused on molecular and electronic spectroscopy, process analysis, and chemometrics. He has published widely, including numerous tutorials, scientific papers and book chapters, individual textbook volumes, software programs, and inventions; he has received several key awards for his work.

LITERATURE CITED

(1) Lavine, B.; Workman, J. *Anal. Chem.* **2006**, *78* (12), 4137–4145.
(2) Workman, J., Jr.; Koch, M.; Veltkamp, D. *Anal. Chem.* **2007**, *79* (12), 4345–4364.
(3) Eiceman, G. A.; Gardea-Torresdey, J.; Dorman, F.; Overton, E.; Bhushan, A.; Dharmasena, H. P. *Anal. Chem.* **2006**, *78* (12), 3985–3996.
(4) Kaliszan, R. *Chem. Rev.* **2007**, *107* (7), 3212–3246.
(5) Christensen, J.; Norgaard, L.; Bro, R.; Engelsen, S. B. *Chem. Rev.* **2006**, *106* (6), 1979–1994.
(6) Chen, W. L. *J. Chem. Inf. Modell.* **2006**, *46* (6), 2230–2255.
(7) Kiralj, R.; Márcia, M. C.; Ferreira., *J. Chemom.* **2006**, *20*, 247–272.
(8) Bruce, C. L.; Melville, J. L.; Pickett, S. D.; Hirst, J. D. *J. Chem. Inf. Modell.* **2007**, *47* (1), 219–227.
(9) Roncaglioni, A.; Benfenati, E. *Chem. Soc. Rev.* **2008**, *37* (1), 441–450.
(10) ASTM International, 100 Barr Harbor Drive, West Conshohocken, Pennsylvania. http://www.astm.org/.

CALIBRATION

(11) Peters, F. T.; Maurer, H. *Anal. Chem.* **2007**, *79* (13), 4967–4976.
(12) Liao, J.; Duan, F. *J. Chem.* **2006**, *20* (6−7), 294–301.
(13) Hibbert, D. B. *J. Chromatogr., A* **2007**, *1158* (1−2), 25–32.
(14) Hibbert, D. B. *Analyst* **2006**, *131*, 1273–1278.
(15) Forina, M.; Lanteri, S.; Casale, M. *J. Chromatogr., A* **2007**, *1158* (1−2), 61–93.
(16) Nelson, P. R. C.; MacGregor, J. F.; Taylor, P. A. *Chemolab.* **2006**, *80* (1), 1–12.
(17) Arteaga, F.; Ferrer, A. *J. Chem.* **2006**, *19* (8), 439–447.
(18) de Noord, O. E.; Theobald, E. H. *J. Chem.* **2006**, *19* (5−7), 301–307.
(19) Marini, F.; Magri, A. L.; Bucci, R.; Magri, A. D. *Anal. Chim. Acta* **2007**, *599* (2), 232–240.
(20) Griffiths, M. L.; Ellison, S. L. R. *Chemolab.* **2006**, *83* (2), 133–138.
(21) Westad, F.; Afseth, N. K.; Bro, R. *Anal. Chim. Acta* **2007**, *595* (1−2), 323–327.
(22) Pell, R. J.; Ramos, S. L.; Manne, R. *J. Chem.* **2007**, *21* (3−4), 165–172.
(23) Faber, N. M. *Anal. Chim. Acta* **2007**, *603* (2), 155–158.
(24) Gomez-Carracedo, M. P.; Andrade, J. M.; Rutledge, D. N.; Faber, N. M. *Anal. Chim. Acta* **2007**, *585* (2), 253–265.
(25) Capron, X.; Walczak, B.; de Noord, O. E.; Massart, D. L. *J. Chem.* **2006**, *19* (5−7), 308–316.
(26) Faber, N. M.; Rajko, R. *Anal. Chim. Acta* **2007**, *595* (1−2), 98–106.
(27) Kasemsumran, S.; Li, K. M.; Ozaki, Y. *Analyst* **2006**, *131*, 529–537.
(28) Kramer, K. E.; Small, G. W. *Appl. Spectrosc.* **2007**, *61* (5), 497–506.
(29) Dabros, M.; Amrhein, M.; Gujral, P.; von Stockar, U. *Appl. Spectrosc.* **2007**, *61* (5), 507–513.
(30) Sohn, M.; Barton, F. E.; Himmelsback, D. S. *Appl. Spectrosc.* **2007**, *61* (4), 414–418.
(31) Watari, M.; Ozaki, Y. *Appl. Spectrosc.* **2006**, *60* (5), 529–538.
(32) Guenard, R. D.; Wehlburg, C. M.; Pell, R. J.; Haaland, D. M. *Appl. Spectrosc.* **2007**, *61* (7), 747–754.
(33) Coleman, M. D.; Brewer, P. J.; Smith, I. M.; Harris, P. M.; Clift, M. G.; Milton, M. J. T. *Anal. Chim. Acta* **2007**, *601* (2), 189–195.
(34) Felizardo, P.; Baptista, P.; Menezes, J. C.; Correia, M. *Anal. Chim. Acta* **2007**, *595* (1−2), 107–113.
(35) Fernandez-Carbanas, V. M.; Garrido-Varo, A.; Perez-Marin, D.; Dardenne, P. *Appl. Spectrosc.* **2006**, *60* (1), 17–23.
(36) Thennadil, S. H.; Martens, H.; Kohler, A. *Appl. Spectrosc.* **2006**, *60* (3), 315–321.
(37) Martens, H.; Bruun, S. W.; Adt, I.; Sockalingum, G. D.; Kohler, A. *J. Chem.* **2007**, *20* (8−10), 402–417.
(38) Chen, Z. P.; Morris, J.; Martin, E. *Anal. Chem.* **2006**, *78* (22), 7674–7681.
(39) Gributs, C. E. W.; Claudia, E. W.; Burns, D. H. *Chemolab.* **2006**, *83* (1), 44–53.
(40) Esteban-Diez, I.; Gonzalez-Saiz, J. M.; Gomez-Camara, D.; Millan, C. *Anal. Chim. Acta* **2006**, *555* (1), 84–95.
(41) Sulub, Y.; Small, G. W. *Appl. Spectrosc.* **2007**, *61* (4), 406–413.
(42) Maruo, K.; Oota, T.; Tsurugi, M.; Nakagawa, T.; Arimoto, H.; Tamura, M.; Ozaki, Y.; Yamada, Y. *Appl. Spectrosc.* **2006**, *60* (4), 441–449.
(43) Maruo, K.; Oota, T.; Tsurugi, M.; Nakagawa, T.; Arimoto, H.; Hayakawa, M.; Tamura, M.; Ozaki, Y.; Yamada, Y. *Appl. Spectrosc.* **2006**, *60* (12), 1423–1431.
(44) Pedge, N. I.; Walmsley, A. D. *Appl. Spectrosc.* **2007**, *61* (9), 940–949.
(45) Wiberg, K. H.; Hultin, U.-K. *Anal. Chem.* **2006**, *78* (14), 5076–5085.
(46) Chen, T.; Martin, E. *J. Chem.* **2007**, *21* (5−6), 198–207.
(47) Shih, W.-C.; Bechtel, K.; Feld, M. S. *Anal. Chem.* **2007**, *79* (1), 234–239.
(48) Shinzawa, H.; Jiang, J.-H.; Ritthiruangdej, P.; Ozaki, Y. *J. Chem.* **2007**, *20* (8−10), 436–444.
(49) Galvao, R. K. H.; Araujo, M. C. U.; Martins, M. do N.; Jose, G. E.; Pontes, M. J. C.; Silva, E. C.; Saldanha, T. C. B. *Chemolab.* **2006**, *81* (1), 60–67.
(50) Borin, A. F.; Marco Flores, M.; Cesar, M.; Danilo, A.; Poppi, R. J. *Anal. Chim. Acta* **2006**, *579* (1), 25–32.
(51) Uestuen, B.; Melssen, W. J.; Buydens, L. M. C. *Anal. Chim. Acta* **2007**, *595* (1−2), 299–309.
(52) Ortiz, M. C.; Sarabia, L. *J. Chromatogr., A* **2007**, *1158* (1−2), 94–110.
(53) Damiani, P. C.; Duran-Meras, I.; Garcia-Reiriz, A.; Jimenez-Giron, A.; Munoz de la Pena, A.; Olivieri, A. C. *Anal. Chem.* **2007**, *79* (18), 6949–6958.
(54) Serneels, S.; Geladi, P.; Moens, M.; Blockhuys, F.; Van Espen, P. J. *J. Chem.* **2006**, *19* (8), 405–411.
(55) Olivieri, A. C. *J. Chem.* **2006**, *19* (11−12), 615–624.
(56) Cramer, J.; Kim, Y.-C.; Vogt, F.; Booksh, K. *J. Chem.* **2006**, *19* (11−12), 593–606.
(57) Luttrell, R. D.; Gilbert, M. K.; Vogt, F. *J. Chem.* **2007**, *21* (1−2), 65–75.
(58) Vogt, F.; Cramer, J.; Booksh, F. *J. Chem.* **2006**, *19* (9), 510–520.
(59) Vogt, F.; Booksh, K. S. *J. Chem.* **2006**, *19* (10), 575–581.
(60) Rinnan, A.; Riu, J.; Bro, R. *J. Chem.* **2007**, *21* (1−2), 76–86.
(61) Hoggard, J.; Synovec, R. E. *Anal. Chem.* **2007**, *79* (4), 1611–1619.
(62) Burger, J.; Geladi, P. *J. Chem.* **2006**, *19* (5−7), 355–363.
(63) Burger, J.; Geladi, P. *J. Chem.* **2007**, *20* (3−4), 106–119.
(64) Watson, N. E.; Siegler, W.; Hoggard, J.; Synovec, R. E. *Anal. Chem.* **2007**, *79* (21), 8270–8280.
(65) Bahram, M.; Bro, R.; Stedmon, C.; Afkhami, A. *J. Chem.* **2007**, *20* (3−4), 99–105.
(66) Bouveresse, D. J.-R.; Benabid, H.; Rutedge, D. N. *Anal. Chim. Acta* **2007**, *589* (2), 216–224.
(67) Gil, D. B.; de la Pena, A. M.; Arancibia, J. A.; Escandar, G. M.; Olivieri, A. C. *Anal. Chem.* **2006**, *78* (23), 8051–8058.
(68) Munoz de la Pena, A.; Mansilla, A. E.; Diez, N. M.; Gil, D. B.; Olivieri, A. C.; Escandar, G. M. *Appl. Spectrosc.* **2006**, *60* (3), 330–338.

RESOLUTION

(69) de Juan, A.; Tauler, R. *J. Chromatogr., A* **2007**, *1158* (1−2), 184–195.
(70) Astakhov, S. A.; Stogbauer, H.; Kraskov, A.; Grassberger, P. *Anal. Chem.* **2006**, *78* (5), 1620–1627.
(71) Levina, E.; Wagaman, A. S.; Callender, A. F.; Mandair, G. S.; Morris, M. D. *J. Chem.* **2007**, *21* (1−2), 24–34.
(72) Crockford, D. J.; Holmes, E.; Lindon, J. C.; Plumb, R. S.; Zirah, S.; Bruce, S. J.; Rainville, P.; Stumpf, C. L.; Nicholson, J. *Anal. Chem.* **2006**, *78* (2), 363–371.
(73) Widjaja, E.; Tan, B. H.; Garland, M. *Appl. Spectrosc.* **2006**, *60* (3), 294–303.
(74) Liangfeng, G.; Garland, M. *Appl. Spectrosc.* **2007**, *61* (2), 148–156.
(75) Cheng, S.; Tjahjono, M.; Rajarathnam, D.; Li, C.; Lyapkalo, I.; Chen, D.; Garland, M. *Appl. Spectrosc.* **2007**, *6110*, 1057–1062.
(76) Zhang, H.; Chew, W.; Garland, M. *Appl. Spectrosc.* **2007**, *61* (12), 1366–1372.
(77) Rajko, R.; Istvan, K. *J. Chem.* **2006**, *19* (8), 448–463.
(78) Vosough, M.; Mason, C.; Tauler, R.; Jalai-Heravi, M.; Maeder, M. *J. Chem.* **2007**, *20* (6−7), 302–310.
(79) Tauler, R. *Anal. Chim. Acta* **2007**, *595* (1−2), 289–298.
(80) Gallagher, N. B.; Blae, T. A.; Gassman, P. L.; Shaver, J. M.; Windig, W. *Appl. Spectrosc.* **2006**, *60* (7), 713–722.
(81) Garrido, M.; Larrechi, M. S.; Rius, F. X. *Appl. Spectrosc.* **2006**, *60* (2), 174–181.
(82) Pedro, A. M. K.; Ferreira, M. M. C. *J. Chem.* **2007**, *20* (1−2), 76–83.
(83) Li, B.; Zhang, J.; Hu, Y.; Liang, Y.; Ozaki, Y. *Appl. Spectrosc.* **2006**, *60* (2), 155–161.
(84) Blanco, M.; Castillo, M.; Peinado, A.; Beneyto, R. *Appl. Spectrosc.* **2006**, *60* (6), 641–647.

(85) Wiberg, K. *J. Chem.* **2006**, *1108* (1), 50–67.
(86) Porter, S. E. G.; Stoll, D. R.; Rutan, S. C.; Carr, P. W.; Cohen, J. D. *Anal. Chem.* **2006**, *78* (15), 5559–5569.
(87) Bezemer, E.; Rutan, S. C. *Chemolab.* **2006**, *81* (1), 82–93.
(88) Esteves da Silva, J. C. G.; Tauler, R. *Appl. Spectrosc.* **2006**, *60* (11), 1315–1321.
(89) Wang, Z.-G.; Jiang, J.-H.; Ding, Y.-J.; Wu, H.-L.; Yu, R.-Q. *Anal. Chim. Acta* **2006**, *558* (1−2), 137–143.
(90) Vivo-Truyols, G.; Schoenmakers, P. *J. Chromatogr., A* **2006**, *1120* (1−2), 273–281.

PATTERN RECOGNITION

(91) Szymanska, E.; Markuszewski, M. J.; Capron, X.; van Nederkassel, A.-M.; Vander Heyden, Y.; Markuszewski, M.; Krajka, K.; Kaliszan, R. *Electrophoresis* **2007**, *28* (16), 2861–2873.
(92) van Nederkassel, A. M.; Daszykowski, M.; Eilers, P. H. C.; Vander Heyden, Y. *J. Chromatogr., A* **2006**, *1118* (2), 199–210.
(93) Prince, J. T.; Marcotte, E. M. *Anal. Chem.* **2006**, *78* (17), 6140–6152.
(94) Jaitly, N.; Monroe, M. E.; Petyuk, V. A.; Clauss, T. R. W.; Adkins, J. N.; Smith, R. D. *Anal. Chem.* **2006**, *78* (21), 7397–7409.
(95) Pierce, K. M.; Wright, B. W.; Synovec, R. E. *J. Chromatogr., A* **2007**, *1141* (1), 106–116.
(96) Krebs, M.; Tingley, R. D.; Zeskind, J. E.; Holmboe, M. E.; Kang, J.-M.; Davis, C. E. *Chemolab.* **2006**, *81* (1), 74–81.
(97) Skov, T.; van den Berg, F.; Tomasi, G.; Bro, R. *J. Chem.* **2006**, *20* (11−12), 484–497.
(98) Stolt, R.; Torgrip, R. J. O.; Lindberg, J.; Csenki, L.; Kolmert, J.; Schuppe-Koistinen, I.; Jacobsson, S. P. *Anal. Chem.* **2006**, *78* (4), 975–983.
(99) Heraud, P.; Wood, B. R.; Beardall, J.; McNaughton, D. *J. Chem.* **2007**, *20* (5), 193–197.
(100) Leger, M. N.; Ryder, A. G. *Appl. Spectrosc.* **2006**, *60* (2), 182–193.
(101) Zhao, J.; Lui, H.; McLean, D. I.; Zeng, H. *Appl. Spectrosc.* **2007**, *61* (11), 1225–1232.
(102) Zhang, L.; Henson, M. J. *Appl. Spectrosc.* **2007**, *61* (9), 1015–1020.
(103) Fredriksson, M.; Petersson, P.; Joernten-Karlsson, M.; Axelsson, B.-O.; Bylund, D. *J. Chromatogr., A* **2007**, *1172* (2), 135–150.
(104) Arneberg, R.; Rajalahti, T.; Flikka, K.; Berven, F. S.; Kroksven, A. C.; Berle, M.; Myhr, K.-M.; Vedeler, C. A.; Ulvik, R. J.; Kvalheim, O. M. *Anal. Chem.* **2007**, *79* (18), 7014–7026.
(105) Craig, A.; Cloarec, O.; Holmes, E.; Nicholson, J. K.; Lindon, J. C. *Anal. Chem.* **2006**, *78* (7), 2262–2267.
(106) Melssen, W.; Wehrens, R.; Buydens, L. *Chemolab.* **2006**, *83* (2), 99–113.
(107) Westerhuis, J. A.; Derks, E. P. P. A.; Hoefsloot, H. C.; Smilde, A. K. *J. Chem.* **2007**, *21* (7), 474–485.
(108) Brys, G.; Hubert, M.; Rousseeuw, P. J. *J. Chem.* **2006**, *19* (5−7), 364–375.
(109) Shinawa, H.; Morita, S.; Ozaki, Y.; Tsenkova, R. *Appl. Spectrosc.* **2006**, *60* (8), 884–891.
(110) Bylesjö, M.; Rantalainen, M.; Cloarec, O.; Nicholson, J. K.; Holmes, E.; Trygg, J. *J. Chem.* **2006**, *20*, 341–351.
(111) Gabrielsson, J.; Jonsson, H.; Airiau, C.; Schmidt, B.; Excott, R.; Trygg, J. *Chemolab.* **2006**, *84* (1−2), 153–158.
(112) Gabrielsson, J.; Airiau, C.; Schmidt, B.; Jonsson, H.; Escott, R.; Trygg, J. *AIChE J.* **2006**, *52* (9), 3164–3172.
(113) Eiceman, G. A.; Wang, M.; Prasad, S.; Schmidt, H.; Tadjimukhamedov, F. K.; Lavine, B. K.; Mirankar, N. *Anal. Chim. Acta* **2006**, *579* (1), 1–10.
(114) Dixon, S. J.; Brereton, R. G.; Carter, J. F.; Sleeman, R. *Anal. Chim. Acta* **2006**, *559* (1), 54–63.
(115) Xu, Y.; Gong, F.; Dixon, S. J.; Brereton, R. G.; Soina, H. A.; Novotny, M. V.; Oberzaucher, E.; Grammer, K.; Penn, D. J. *Anal. Chem.* **2007**, *79* (15), 5633–5641.
(116) Statheropoulos, M.; Mikedi, K.; Agapiou, A.; Georgiadou, A.; Karma, S. *Anal. Chim. Acta* **2006**, *566* (2), 207–216.
(117) Hall, G. J.; Kenny, J. E. *Anal. Chim. Acta* **2007**, *581* (1), 118–124.
(118) Zhang, C.; Suslick, K. S. *J. Agric. Food Chem.* **2007**, *55* (2), 237–242.
(119) Pearce, T. C.; Gu, J.; Chanie, E. *Anal. Chem.* **2007**, *79* (22), 8511–8519.
(120) Kramer, K. E.; Rose-Pehrsson, S. L. *Anal. Chim. Acta* **2007**, *584* (1), 78–88.
(121) Pearman, W. F.; Fountain, A. W. *Appl. Spectrosc.* **2006**, *60* (4), 356–365.
(122) de Jong, B. W. D.; Bakker Schut, T. C.; Maquelin, K.; van der Kwast, T.; Bangma, C. H.; Kok, D.-J.; Puppels, G. J. *Anal. Chem.* **2006**, *78* (22), 7761–7769.
(123) Widjaja, E.; Seah, R. K. H. *Appl. Spectrosc.* **2006**, *60* (3), 343–345.
(124) Wang, L.; Chapman, J.; Palmer, R. A.; Alter, T. M.; Hooper, B. A.; van Ramm, O.; Mizaikoff, B. *Appl. Spectrosc.* **2006**, *60* (10), 1121–1126.
(125) Escoriza, M. F.; VanBriesen, J. M.; Stewart, S.; Maier, J. *Appl. Spectrosc.* **2007**, *61* (8), 812–823.
(126) Geladi, P.; Nelson, A.; Lindholm-Sethson, B. *Anal. Chim. Acta* **2007**, *595* (1−2), 152–159.
(127) Forshed, J.; Idborg, H.; Jacobsson, S. P. *Chemolab.* **2007**, *85* (1), 102–109.
(128) Ramos, P. M.; Ruisanchez, I. *Anal. Chim. Acta* **2006**, *558* (1−2), 274–282.
(129) Ramos, P. M.; Callao, M. P.; Ruisanchez, I. *Anal. Chim. Acta* **2007**, *584* (2), 360–369.
(130) Woody, N. A.; Brown, S. D. *J. Chem.* **2007**, *21* (7−9), 357–363.
(131) Ebrahimi, D.; Li, J.; Hibbert, D. B. *J. Chromatogr., A* **2007**, *1166* (1−2), 163–170.
(132) Daszykowski, M.; Stanimirova, I.; Bodzon-Kulakowska, A.; Silberring, J.; Lubec, G.; Walczak, B. *J. Chromatogr., A* **2007**, *1158* (1−2), 306–317.
(133) Masoum, S.; Jouan-Rimbaud Bouveresse, D.; Vercauteren, J.; Jalali-Heravi, M.; Rutledge, D. N. *Anal. Chim. Acta* **2006**, *558* (1−2), 144–149.
(134) Eriksson, L.; Wold, S.; Trygg, J. *J. Chem.* **2006**, *19* (5−7), 393–403.
(135) Lasch, P.; Diem, M.; Haensch, W.; Naumann, D. *J. Chem.* **2007**, *20* (5), 209–220.
(136) Wang, Y.-P.; Wang, Y.; Spencer, P. *Appl. Spectrosc.* **2006**, *60* (7), 826–832.
(137) Jacobsen, S.; Grove, H.; Jensen, K. N.; Soerensen, H. A.; Jessen, F.; Hollung, K.; Uhlen, A. K.; Joergensen, B. M.; Faergestad, E. M.; Soendergaard, I. *Electrophoresis* **2007**, *28* (8), 1289–1299.
(138) Zhou, L.; Zeng, Y. *Anal. Chim. Acta* **2007**, *584* (1), 223–227.
(139) Hollingsworth, B. V.; Reichenbach, S. E.; Tao, Q.; Visvanathan, A. *J. Chromatogr., A* **2006**, *1105* (1−2), 51–58.
(140) Myles, A. J.; Zimmerman, T. A.; Brown, S. D. *Appl. Spectrosc.* **2006**, *60* (10), 1198–1203.
(141) Latino, D. A. R. S.; Aires-de Sousa, J. *Anal. Chem.* **2007**, *79* (3), 854–862.
(142) Sigman, M. E.; Williams, M. R. *Anal. Chem.* **2006**, *78* (5), 1713–1718.
(143) Li, J.; Hibbert, D. B.; Fuller, S.; Vaughn, G. *Chemolab.* **2006**, *82* (1−2), 50–58.
(144) Sadygov, R.; Wohlschlegel, J.; Pak, S. K.; Xu, T.; Yates, J. R. *Anal. Chem.* **2006**, *78* (1), 89–95.
(145) Choi, S. W.; Martin, E. B.; Morris, A. J.; Lee, I. B. *Ind. Eng. Chem. Res.* **2006**, *45* (9), 3108–3118.

AC800728T