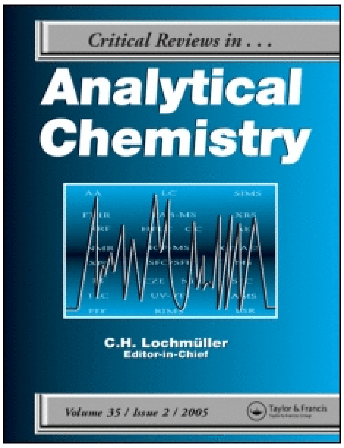
This article was downloaded by:
On: *17 January 2011*
Access details: *Access Details: Free Access*
Publisher *Taylor & Francis*
Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Critical Reviews in Analytical Chemistry

Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713400837

Recent Developments in Multivariate Calibration

Jon Gabrielsson[a]; Johan Trygg[a]
[a] Research Group for Chemometrics, Institute of Chemistry, Umeå University, Umeå, Sweden

To cite this Article Gabrielsson, Jon and Trygg, Johan(2006) 'Recent Developments in Multivariate Calibration', Critical Reviews in Analytical Chemistry, 36: 3, 243 — 255

To link to this Article: DOI: 10.1080/10408340600969924

URL: http://dx.doi.org/10.1080/10408340600969924

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.informaworld.com/terms-and-conditions-of-access.pdf

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Critical Reviews in Analytical Chemistry, 36:243–255, 2006
Copyright © Taylor and Francis Group, LLC
ISSN: 1040-8347 print / 1547-6510 online
DOI: 10.1080/10408340600969924

Recent Developments in Multivariate Calibration

Jon Gabrielsson and Johan Trygg
Research Group for Chemometrics, Institute of Chemistry, Umeå University, Umeå, Sweden

This review covers the area of multivariate calibration; from pre-processing of data prior to modeling and applications of regression methods for calibration and prediction. The importance of pre-treatment of data is highlighted with many of the recently developed methods together with traditional methods. Several articles provide comparisons between different pre-processing methods. Methods for data from coupled chromatographic methods, which have found increasing use and where data pre-processing is a prerequisite for multivariate modeling, are also included. Many of the novel chemometric methods deal with model complexity and interpretation. A diverse set of applications are also presented and references are also given to early papers, making it possible to acquire a deeper knowledge of methods of interest.

Keywords pre-processing, spectroscopy, coupled chromatographic techniques, multivariate calibration, prediction

INTRODUCTION

Multivariate calibration is used to develop a quantitative relation between the predictor variables in X and the response variables in Y. Multivariate calibration has had several enhancements/extensions in recent years (1) that have found widespread use in analytical sciences.

Often, systematic variation not relevant for the prediction of Y is present in X. Examples of such systematic variations in spectroscopy can be baseline, drift and scatter effects, as well as spectra of impurities and background. Pre-processing of data is then applied to remove systematic variation not related to the Y matrix. In this review we discuss different pre-processing methods for spectroscopic techniques commonly applied in multivariate calibration. There is an increasing use of chromatographic instruments coupled to spectroscopic detectors. Data generated using these techniques have to be processed prior to multivariate data analysis. Commonly applied processing methods for data from different instruments are reviewed.

Development of chemometric methods in the area of multivariate calibration and prediction is covered in this article. A wide range of diverse applications of multivariate calibration methods are presented that encompass commonly used analytical techniques.

Pre-Processing

The aim of data pre-processing is to remove unwanted systematic variation such as baseline shifts, scatter effects and effects from uncontrolled external factors. Methods generally applied can be divided into two categories. The first category consists of classical methods for spectral normalization and smoothing and differentiation, e.g., multiplicative scatter correction (MSC, EMSC), standard normal variate (SNV) transformation, Savitzky and Golay smoothing and first and second derivatives (2–5). The second category consists of methods for variable selection and dimensionality reduction and has received considerable attention over the past decade.

Orthogonal signal correction (OSC) filters have been developed to remove systematic variation in the descriptor matrix, **X** that is not correlated to the response matrix, **Y**.

The general, single component OSC model of **X** can be expressed by:

$$\mathbf{X} = \mathbf{t}_\mathrm{o}\mathrm{p}_\mathrm{o}^\mathrm{T} + \mathbf{E} \quad \text{where} \quad \mathbf{t}_\mathrm{o} = \mathbf{X}\mathbf{w}_\mathrm{o}, \text{ and } \mathrm{t}_\mathrm{o} \perp \mathbf{Y}$$

The OSC component should:

- contain (large) systematic variation in **X**;
- be predictive using **X** (in order to be applied to future data);
- be orthogonal to **Y**;

Wold et al. published the original work on OSC (6). Soon after, a number of papers described alternative OSC methods. Fearn and Höskuldsson modified the OSC algorithm to make it more theoretically clear, although producing slightly different results (7, 8). The DOSC method by Westerhuis et al. (9) used a generalized inverse of the X matrix in the OSC method, and in the OSC method by Trygg (10) this is replaced by a partial least squares (PLS) based approach model. Li et al. incorporated

---

Address correspondence to Johan Trygg, Research Group for Chemometrics, Institute of Chemistry, Umeå University, Umeå, SE-90187, Sweden. E-mail: johan.trygg@chem.umu.se

both the Wold OSC method and the Fearn OSC method into a hybrid OSC method (11). The piecewise orthogonal signal correction method by Feudale et al. modified the OSC method by Fearn to perform it in a piecewise manner (12). Later, Feudale also used the multi-scale property of the discrete wavelet transform to apply an OSC filter on each frequency band separately (13).

Please note that the modeling results of applying different OSC algorithms will vary, although all OSC methods have the same overall objective. Nicely written reviews can be found also in Svensson et al. (14) and Goicoechea et al. (15).

In the paper by Eriksson et al., OSC and wavelet analysis were applied on fluorescence data related to a white sugar production process (16). In the study by Blanco et al. OSC was applied to minimize the spectral differences in near infrared (NIR) spectra of two types of samples of active pharmaceutical ingredient (API): production samples and laboratory samples (17). Samp et al. applied OSC to improve a NIR calibration model for wort fermentability in brewing (18).

Hansen presented a pre-processing method called independent interference reduction (IIR) (19). IIR has similarities to OSC. In IIR two data sets are used; a principal component analysis (PCA) is performed on one data set and by projecting the second data set onto this, the interference part of this data set is obtained. This variation is then removed to yield the corrected data for multivariate calibration. An application to Fourier transform infrared (FT-IR) data for determination of acetone is presented.

Direct orthogonalization (DO) was presented by Andersson in 1999 (20). Fernández Pierna et al. investigated the effects of DO before applying principal component regression (PCR) and PLS for two polymer and two NIR data sets (21). DO did not significantly improve the predictive ability of the PLS models compared to classical pre-processing methods. However, it is suggested that DO can be used to obtain a better understanding of the variation present in the data.

Rodríguez et al. optimized and applied a method combining DOSC and classical least-squares (CLS) to the determination of the content of antioxidants and organic acids in rubber by NIR spectroscopy (22). A validation set was used to compare results obtained with PLS, including several spectral pre-processing methods, to OSC methods. The number of predictive components after applying OSC based methods decreased while the predictive ability was similar for all investigated methods.

In a study by Stordrange et al. multivariate calibration models based on NIR spectroscopy were developed for prediction of the initial reactant and an impurity in a synthesis of contrast media (23). The objective was to replace an existing off-line monitoring by high performance liquid chromatography (HPLC). Apart from conventional pre-processing methods such as normalization, differentiation and MSC, OSC and optimized scaling were tested. Optimized scaling gives a scaling vector for each sample (24). Except for optimized scaling, which uses PCR, modeling was performed by PLS. The best predictive ability was achieved using optimized scaling on differentiated data.

In nuclear magnetic resonance (NMR) based metabonomics the interpretation of biological data can be confounded by variation not related to the biological responses, such as fluctuations in either experimental conditions or in physiological status. Beckwith-Hall et al. applied OSC to 1H-NMR of biofluids in order to minimize the inherent physiological variation in the samples as well as the inter- and intra-spectrometer variation during data acquisition (25).

Moisture and protein content of alfalfa samples were analyzed by NIR diffuse reflectance spectroscopy and multivariate calibration methods by Azzouz et al. (26). Derivatives, SNV, MSC, De-trending and OSC were applied to the spectroscopic data. It was found that in general the pre-processing methods yielded similar predictions of moisture and protein content if enough PLS components were used. In a study by Fardim et al. multivariate calibration based on NIR spectroscopy was applied to kraft pulps (27). Differentiation using the first-derivative, Kubelka-Munk transformation, or a combination of both was used as pre-processing techniques. PLS models with two to seven PLS components and good predictive ability were established for determination of mechanical and optical properties of the pulp. Models were validated by using cross-validation methodology and a comparison of measurements.

Wulfert et al. proposed a method called Continuous Piecewise Direct Standardization for removing temperature effects in NIR spectra to improve predictions of ternary mixtures of alcohol and water (28). Martens et al. presented a pre-processing method for data from process spectrometers and other multichannel instruments (29). The method is based on a covariance-weighted pre-processing technique and the objective is to combine the advantages of soft modeling with prior knowledge. Spectra or estimates of the unwanted components are used in the pre-processing step to remove this variation from the subsequent calibration model. Arnold et al. (30) presented a procedure where the pure component spectra of both target analyte and possible interferences are used in the calibration model. The utility of the method was illustrated by an analysis of calibration models ternary mixtures of sugars. Martens et al. applied EMSC to diffuse transmission NIR spectra of powder mixtures of protein and starch in order to separate the effects of physical light-scattering from chemical light absorbance, thus simplifying the multivariate calibration model (31).

In the study of Roger et al., a pre-processing method called external parameter orthogonalisation (EPO) is presented (32). The objective of EPO is to remove the influence of external parameters in order to improve the robustness of a calibration model for online measurement of sugar content in fruit using NIR spectroscopy. It is an alternative to variable selection and is close to the method by Hansen (19), but with the difference that the sample size used can be small and that the response parameter does not need to be constant. Hageman et al. provide a comprehensive account of methods for dealing with temperature influences on NIR spectra (33).

A spectrophotometric method for the simultaneous determination of cobalt, nickel and palladium was developed by Arab Chamjangali et al. Due to high spectral overlapping for these components this is a difficult system to determine. A PLS model was compared to a combination of DOSC-PLS. OSC and DOSC were applied to NIR spectra of roast coffee in order to enhance the robustness and reliability of multivariate calibration models. In the paper by Pizarro et al. PLS models were developed to quantify the ash content and lipids in roasted coffee (34). The pre-processing methods based on OSC were compared to results from original data and the data corrected by derivation, SNV and MSC. It was found that for both responses, DOSC and a combination of differentiation and OSC gave PLS calibration models with improved prediction abilities. In order to enhance model performance two OSC (Wold and Höskuldsson) filters were compared as pre-processing techniques by Pöllänen et al. (35). The overall objective of the study was to develop stable PLS calibration for solute concentration measurement in batch cooling crystallization processes. The prediction errors for two external test sets were used to decide the best model.

Yu and MacGregor presented an alternative approach to OSC based on post-processing a standard PLS model with canonical correlation analysis (CCA) (36). The paper points to problems related to OSC, e.g., over-fitting and uncertainty regarding the number of OSC components to extract. Methods based on OSC + PLS and PLS + CCA were compared on an artificial data set with one part related and one part unrelated to **Y**. The authors claim little risk of over-fitting, simpler computation and a better estimate of Y-related variation using the PLS + CCA approach.

Zeaiter et al. have written two review articles on the topic of robustness of multivariate calibration methods. The first part is a review of definitions of robustness for multivariate calibration and methods for evaluating robustness (37). The focus of the second article is to present an overview of pre-processing methods commonly applied to data based on IR spectroscopic techniques in multivariate calibration (38). Vogt et al. proposed two improved PCR methods for correction of drift in spectroscopic methods (39). The methods were compared to Savitzky-Golay pre-processing followed by conventional PCR.

Woody et al. investigated transfer of PLS calibration models among four NIR spectrometers for a quantitative analysis of polymers (40). A comparative study of second derivatives, MSC, finite impulse response filtering, slope and bias correction, model updating (MU), and OSC was conducted. It was found that both OSC and MU yielded robust PLS models with enhanced predictive ability that outperformed the other studied methods. Andrew and Fearn introduced a method called transfer by orthogonal projection for deriving calibrations that are robust to between-instrument variation (41). This would enable calibration transfer without instrument adjustment. The method involves identifying directions in the spectral space in which most between-instrument variability lies and to remove this variation from the spectral data, i.e., the calibration spectra are orthogonalized with respect to this variation. Leion et al. evaluated several calibration transfer methods for NIR diffuse-reflectance data from a pharmaceutical formulation (42).

In the article by Janné et al. spectroscopic data is used in combination with hierarchical PCA and PLS (H-PCA and H-PLS) for characterization and classification of a filler in pharmaceutical tablet formulation (43). The H-PCA method gave a more accurate classification in the scores plot compared to MSC preprocessing. An H-PLS model was applied to build a calibration model between attenuated total reflectance (ATR) FT-IR spectra and measured tensile strength.

Wavelet Transform

The use of the wavelet transform in literature has primarily focused on the compression and noise reduction of separate signals and less on the analysis of multiple signals. Wavelets can be used as a pre-processing step in multivariate analysis (MVA), to build multivariate calibration models or time series models. Wavelets approximate eigenfunctions in many systems, and can produce approximate de-correlations of autocorrelated signals (44). As mentioned previously, Eriksson et al. used wavelets on process fluorescence data, in combination with other preprocessing methods, such as OSC (16). Artursson et al. applied wavelets to pre-processing and compression in multivariate calibration (45). Bos and Vrielink and Alsberg selected the wavelet scales that were optimal for classification of IR spectra (46, 47). Recently, Galvão et al. used both the X and Y matrix in a multivariate calibration setting in an attempt to optimize the selection of wavelet coefficient variables in the final calibration model (48).

The multi-scale approach is to calculate a multivariate model on each scale separately and then recombine the filtered data to allow reconstruction. No compression is achieved, but the multiscale nature of the signals is taken into account as is the ability of wavelets to separate individual features in the data. Bakshi et al., Nonou et al. and Luo et al. successfully used wavelets in multi-scale analysis and the modeling of process data for use both off-line and on-line (44, 49, 50). Björk et al. applied the multi-scale approach to acoustic process measurements as input for PLS (51). They applied the Fourier transform on each wavelet scale to make the coefficients time-invariant, which is very useful for time series and 2D images.

Tan et al. has developed a dual-domain regression method based on the discrete wavelet transform and the PLS regression method (52). First, the discrete wavelet transform is used to preprocess NIR spectral data into multiple frequency bands. Each one of these frequency bands are used as input for PLS regression. A similar wavelet transform coupled with PLS regression approach has been proposed by Cocchi et al. (53).

Coelho et al. proposed an approach that maximise wavelet compression by adjusting the quadrature mirror filter bank in combination with the newly developed successive projection algorithm (54).

In a study by Chen et al. a method based on wavelet prism and OSC for eliminating interference sources in NIR spectra of a four component aqueous sugar solutions is proposed (55).

Pre-Processing for Coupled Chromatographic Data

The use of coupled chromatographic techniques, e.g., gas chromatography-mass spectrometry (GC-MS) and liquid chromatography-mass spectrometry (LC/MS), has increased in over the last 5 years, much due to extensive use in the omics area (56–58). In contrast to a 1H-NMR spectrum, GC-MS and LC-MS data must be processed before MVA. The reason is the two-dimensional nature (chromatogram-mass spectra) of the data for each sample. For a review of earlier work on resolution of two-way data from hyphenated chromatographic instruments, see Liang and Kvalheim (59). For data, curve resolution or deconvolution methods are mainly applied for data processing, see e.g., Halket et al. and Jonsson et al. (60, 61). This gives a resolved spectral and chromatographic profile for each detected compound. The 1-D multivariate profile used to characterize each sample is made up of the integrated areas of all detected chromatographic peaks. An automated procedure for background correction and resolution of multicomponent data was published by Shen et al. (62). The performance of the method was evaluated on data from HPLC diode array detection (DAD), flow injection analysis (FIA) DAD, and. Eide et al. studied mutagenicity of mixtures of polycyclic aromatic compounds with (63). Chromatograms were resolved using an automated curve resolution procedure and then integrated in order to be used as input in a PLS model for mutagenicity. In the work by Jalali-Heravi and Borough and resolution techniques were used for determination of fatty acids in fish oil (64). In Schauer et al. the corresponding mass spectrum and retention time are used for identification purposes (65).

Curve resolution is also applied on data, see e.g. Idborg-Björkman et al. (66). In the work of van Zomeren et al. the performances of five curve resolution methods were evaluated using HPLC-DAD data (67).

Whitson and Maeder made an extension to evolving factor analysis (EFA) (68). The method, termed exhaustive EFA, connects the start and end of concentration windows by systematic rank analysis of all possible submatrices. Chen et al. presented a method for recursive evolving spectral projection as a solution to discordance problems between rank maps and elution patterns in conventional EFA when defining concentration windows (69). A number of MCR methods were applied to resolve on-flow LC-NMR data containing eight closely eluting compounds by Wasim and Brereton (70).

Jiang et al. proposed a simplex based self-modeling curve resolution (SMCR) method that was evaluated on HPLC-DAD data (71). Berbel et al. applied MCR with alternating least squares (MCR-ALS) for background correction and for resolving and quantifying coeluted analyte mixtures in HPLC coupled to voltametric detection (72). Peré-Trepat et al. used MCR-ALS for correctly resolving data for mixtures of multiple biocide compounds standard samples and environmental samples (73). MCR-ALS has also been applied for in situ monitoring of chemical reactions with FT-IR and NIR (74–76). Jiang et al. published a review on the developments in SMCR (77).

Li et al. applied MCR-ALS methods to second-order data from capillary electrophoresis DAD for the determination of dinitrotoluene isomeric compounds (78). The paper by Beltran et al. describes a method for determination of quinolones ciprofloxacin and sarafloxacin in chicken muscle samples with capillary zone electrophoresis (79).

In the article by van Zomeren et al. augmented curve resolution was applied to data matrices from HPLC and micellar electrokinetic chromatography with DAD detection for drug impurity profiling (80). A method for the alignment between wavelength scales of both detectors termed wavelength shift eigenstructure tracking was proposed.

Another alternative is to apply a peak detection algorithm that identifies all chromatographic peaks and use their integrated areas as the multivariate profile characterizing that sample, see e.g., the work by Andreev et al. (81).

Plumb et al. investigated a combination of liquid chromatography/time-of-flight mass spectrometry and PCA for detection of metabolites in biological samples (82).

Vivó-Truyols et al. presented two papers describing a program for analysis of chromatograms with minimal user interaction (83, 84). Another alternative proposed by Allen et al. is to sum the chromatographic direction to create a 1-D multivariate profile produced by the total intensity over all mass spectral channels (85).

A semi-automatic method, where chromatograms are converted into semi-quantitative variables, was developed by Christensen et al. (86). Recently partly alternative methodologies have been applied to data by Jonsson et al. (61, 87) and data (88) where all samples are processed simultaneously and a common set of descriptor variables are extracted.

Nordén et al. applied PCA and PLS discriminant analysis (PLS-DA) to data from clinical urine samples (89). The objective of the study was to assess the information contained in such data and use that information in the search for clinical biomarkers. In the work by Idborg et al. data from LC coupled to electrospray ionization mass spectrometry of rat urine samples were pre-processed with a dimensionality reduction method and two commercially available software packages for peak detection and MCR (90). The processed data were used for metabolic fingerprinting using PLS-DA.

Olivieri described how unfolded PLS and residual bilinearization, in combination, results in a second-order multivariate calibration method that is capable of achieving the second-order advantage (91). This allows concentration and spectral profiles to be extracted directly from second-order data.

Multivariate Calibration

Flåten and Walmsley presented an approach based on Design of Experiments (DoE) for selecting the calibration model (92).

Six factors were varied according to DoE; type of regression method, scaling, Box–Cox transformation, OSC pre-treatment, differentiation, and number of components.

Multivariate design (MVD) is an important tool to ensure a representative and balanced training set for multivariate modeling. Multivariate calibration in process analytical chemistry, quantitative structure activity relationships in medicinal and pharmaceutical chemistry, and data mining are application areas covered in the paper by Wold et al. (93). Designs suitable for MVD are discussed as well as how and when to apply them in multivariate calibration. Jørgensen and Næs presented a general approach based on DoE for situations where complex raw materials that lack controllable factors are used (94). A factorial design for controlled factors was combined with raw materials used as block variables. The utility of the method is exemplified with FT-IR characterization of milk samples in cheese production.

New Methods for Multivariate Calibration and Prediction

Jiang et al. proposed a new procedure for selection of wavelength intervals, moving window partial least-squares regression (MWPLSR), for multicomponent spectral analysis (95). The whole spectral region is scanned with PLS models of a moving window. Kasemsumran et al. applied MWPLSR to NIR spectroscopy of phosphate buffer solutions for simultaneous determination of human serum albumin, gamma-globulin, and glucose at 37°C (96). The most informative spectral regions were identified using MWPLSR and the predictive ability was compared to reports in the literature.

Zhang et al. presented two alternative PLS methods, averaged and weighted PLS (APLS and WPLS), that compute the (weighted) average of PLS models with different complexity (97). The resulting models were less dependent on deciding the correct number of PLS components. The predictive ability in terms of root mean squared error pf prediction of the methods was comparable to that of ordinary PLS for three real data sets. By combining PLS and boosting Zhang et al. proposed a method more resistant to over-fitting called boosting PLS (BPLS) (98).The method of boosting originated in the field of machine learning and combines a set of less complex models in one model (99). In comparisons to ordinary PLS on seven real data sets BPLS was found to have similar predictive ability.

Marbach proposed a new method for multivariate calibration that combines the best features of "classical" and "inverse" calibration (100). Provided that a response spectrum of the analyte of interest exists and the spectra of known concentrations are available, the covariance of spectral noise can be estimated and the optimum solution can be calculated. The method is demonstrated on a publicly available data set.

Continuum regression (CR) encompasses the ordinary least squares (OLS), PCR and PLS regression methods by the use of a tuning parameter in the model estimation. Qannari et al. describes a new CR approach referred to as simple CR (S-CR) (101). It is based on two variable parameters, the number of model components (latent variables) and a tuning parameter. The range of the tuning parameter between 0 and 1 that corresponds to the OLS and PLS solution respectively. Serneels has developed a robust version of the CR framework by using robust estimators based on trimmed variance (102). This combines the benefits of a robust calibration technique and the versatility of CR. For an overview of earlier CR methods see de Jong et al. (103).

Support vector machines (SVM) are an approach to regression and classification recently introduced in chemometrics from the field of machine learning (104–106). Belousov et al. give an overview of SVM and examine the performance of SVM in classification (107). Cogdill and Dardenne provide an introduction to SVM regression (SVR) for application to NIR spectroscopy (108). Least squares SVM (LS-SVM) is also compared to PLS and artificial neural networks for regression and classification on four data sets. Zomer et al. described a procedure involving active learning SVM (AL-SVM) for sample selection in classification (109). The greatest benefit from AL-SVM was found when there was intermediate separation between classes.

Trygg and Wold (10) recently developed the Orthogonal PLS method (OPLS), which is a recent modification of the PLS method that facilitates model interpretation. The main idea of OPLS is to separate the systematic variation in X into two parts, one that is linearly related (and, therefore, predictive) to Y and one that is unrelated (orthogonal) to Y. For a single Y-variable, the OPLS model will always reduce down to a single predictive component. Later Trygg proposed an extension called O2PLS (110), that provides separate models of the joint and orthogonal variation in *both* X and Y. This makes the O2PLS model predictive in both directions (X-Y).

In the review by Svensson et al. OPLS is compared to OSC filters (14). Also, some interesting properties of OPLS are discussed by Verron et al. (111). It is shown that the number of PLS components to correct is not important, and that the prediction quality for an OPLS corrected PLS model is equal to a PLS model with small number of components. For multivariate calibration the main benefit is thus reduction of the number of PLS components rather than improved predictions. Ergon presented an alternative to OPLS using ordinary PLS and similarity transformation as a post-processing method (112).

Cloarec et al. applied the OPLS method for extracting potential biomarkers in 1H-NMR metabonomic data based on an animal model of insulin resistance (113). For multivariate spectral and chromatographical data, the ability to estimate pure constituent profiles provides a valuable diagnostic tool of interpretation. Trygg demonstrated that inverse calibration methods, including OPLS, can also provide an estimate of the pure spectral profiles by a simple rotation of the regression coefficient matrix B, namely $B(B^TB)^{-1}$ (114).

Net Analyte Signal Theory in Multivariate Calibration

A generalization of the original figures of merit, e.g., sensitivity, selectivity, signal-to-noise ratio and limit of detection, which allowed the parameters to be estimated for multivariate data, was presented by Lorber (115). Later, Lorber et al. showed that estimation of the net analyte signal (NAS) was possible also for the inverse calibration models (116). Brown reviewed theoretical and practical aspects of the NAS concept in multivariate calibration and identified inconsistencies and limitations of NAS theory in the presence of e.g., measurement errors (117).

NAS is defined as the part of the signal which is orthogonal to the spectra of the other components. Improved methods for estimating NAS in inverse calibration models have been proposed by e.g., Faber (118), Ferré et al. (119), and Bro and Andersen (120).

Muñoz de la Peña et al. studied the determination of binary mixtures by stopped-flow kinetic analysis and performed a comparative study of NAS-based methods and PLS (121). In a study by Vivó-Truyols et al. multivariate selectivity (SEL) and scalar NAS were used as objective functions in the optimization of a LC method (122). Martínez Galera et al. applied a NAS-based method to determine carbendazim, fuberidazole and thiabendazole in water samples and calculated figures of merit (123). The performance of the investigated method was found to be similar to PLS. An account of NAS in multivariate applications can be found in the article by Ferré and Faber (124). The objective of the paper is to present a unified approach to multivariate calibration and prediction based on NAS methods.

A method based on NAS for product quality control, NAS statistical quality control, was proposed by Skibsted et al. (125). In line with NAS theory the method separates the systematic variation of the product, which enhances interpretability and troubleshooting.

Utilizing the concept of NAS Boschetti and Olivieri proposed a method for pre-processing spectroscopic data by orthogonal projection onto the space of the studied component, thus creating a set of net analyte signals for quantifying the desired mixture component (126). This method has been evaluated by Goicoechea and Olivieri (15, 127).

Applications to Spectroscopic Data

The feasibility of replacing conventional chemical testing in the chemical quality control of a pharmaceutical intermediate, 7-Amino-cephalosporanic acid, with NIR spectroscopy was investigated by Andre (128). Correct classification was obtained after appropriate pre-processing and scaling. Chen et al. applied NIR to seawater to model and predict ion concentrations using PCA, CCA and PLS (129). Févotte et al. used fiber-optic NIR spectroscopy to investigate features of polymorphic transitions observed during the crystallization and the filtration of an API (130). Calibration with NIR data was used to provide measures of the polymorphic composition of the solid product. Several industrial applications of the described methods are discussed. IR spectroscopy was applied for in situ monitoring of a crystallization process by Pöllänen et al. (131). ATR FT-IR spectroscopy provided valuable process information and the results demonstrate that multivariate techniques provide the powerful tool for rapid evaluation of spectral data and also enable more reliable quantification of polymorphic composition of samples. Rodrigues et al. developed a PLS calibration method based on NIR for monitoring the outlet concentration of an API in an ionic exchange elution process (132). The process analytical technology initiative by the Food and Drug Administration in the United States will increase the use of NIR spectroscopy (133).

Lestander and Rhen applied the factorization of PLS proposed by Ergon termed bi-orthogonal PLS (134) to NIR data of biofuels to model and predict moisture, ash content and gross calorific value (135).

An online application of Raman spectroscopy for monitoring of an industrial process separating aromatic compounds was investigated by Estienne et al. (136). Sivakesava et al. compared FT mid-IR and FT-Raman spectroscopy for monitoring and prediction of glucose, ethanol and optical density in a fermentative production (137). O'Grady et al. presented a pseudo-second derivative of Raman spectra to eliminate effects of intense luminescence backgrounds in spectra recorded with charge-couple device detectors (138). In a study by Johansson et al. Raman spectroscopy was evaluated for quantitative assessment of API in conventional wet granulated pharmaceutical immediate release tablets (139). Romero-Torres et al. presented a study combining Raman spectroscopy with multivariate calibration and prediction for quantitatively characterizing tablet coating variations (140).

Karlberg and Torgrip describe how FIA can be improved by applying chemometric methods when developing a FIA method and for treating raw data acquired from a FIA detection system (141). DoE has proved to be useful to optimize FIA methods.

Robust Calibration Models

A method for constructing robust calibration models by selecting robust variables was investigated by Swierenga et al. (142). The method is compared with a traditional global calibration model approach.

Thomas and Ge presented an approach to robust calibration models based on DoE that included instrumental and environmental factors in the model building (143). Despagne et al. evaluated PLS, locally weighted regression (LWR) and neural networks in the development of a robust calibration model for water content in a distillation process (144). The calibration model was based on NIR data, which contained non-linear effects due to changes in temperature and variations between the different batches. Mittermayr et al. calculated wavelet coefficients to remove background variation from spectroscopic data and then applied linear regression to obtain a robust calibration model (145). In the article by Larrechi and Callao a practical strategy for introducing NIR spectroscopy and multivariate calibration methods in industrial applications is described (146). The utility of NIR spectroscopy for online applications is discussed. Various univariate and multivariate control techniques are described together with potential applications. Sahni et al. investigated the problems that might occur in calibration transfer due to variations in distance between the transmittance fiber-optic probes (147). A number of methods were compared on data measured at five different probe distances on samples varied according to a mixture design. Zachariassen et al. developed an online process monitoring and control system based on NIR and PLS for improving prediction and dosing of ammonia in production (148).

Multivariate Calibration in Process Applications

Andersson et al. presented a method for real-time analysis of film coating on pharmaceutical pellets in an industrial process using a diffuse reflectance fiber-optic NIR probe (149). Process conditions were varied according to DoE and the good predictive ability of the calculated model also enabled end-point detection with high precision. Nordon et al. investigated a low-field NMR spectrometer for use at-line and successfully applied multivariate calibration models on a methacrylamide process (150).

In the paper by Chang et al. an approach based on LWR model with an extended library is presented for improving the performance of NIR online multivariate calibration models (151). Cho and Kim proposed a method for predicting future observations in batch modeling with library selection of the most similar past batch trajectory for a new batch (152). This trajectory is then used as the basis for predicting the unknown part of the new batch.

Bilinear and trilinear PLS models were used to model the performance and predict the resulting API concentration in an industrial fed-batch fermentation process in a study by Lopes and Menezes (153).

In a study by Skoglund et al. wood chip properties were predicted online using a visible wave length range (VIS) spectrometer (154). PLS models for pine, bark and dry content were generated in lab scale based on DoE. A NIR spectrometer, calibrated via DoE to make the same predictions as did the VIS spectrometer, was also evaluated for a shorter time period, with similar results. Jonsson et al. presented an approach based on NIR spectroscopy and multivariate data analysis for online control and monitoring of pulpwood chip properties (155). A moving average of predicted values was more suitable for process monitoring and control. Lab data were compared to the filtered on-line predictions with good agreement for moisture and pine chip contents.

Non-Linear Effects in Calibration Data

In the study by Thissen et al. the performance of SVM and PLS are compared for spectroscopic data (156). For Raman spectroscopy SVR was used successfully on low-resolution spectra. SVR was also found to give better predictions for the studied NIR system, which was affected by temperature induce non-linear effects. Rantanen et al. developed a non-linear calibration model with NIR and process data for in-line moisture measurement during fluid bed granulation (157). Thissen et al. and Chauchard et al. also investigated LS-SVM as a method for non-linear multivariate calibration (158, 159). SNV pre-processing was found to improve the model robustness for LS-SVM. In the work by Stordrange et al. regression methods suited for handling non-linear effects in data were compared to PCR and PLS on NIR spectra from a pharmaceutical process (160). Oliveira-Esquerre et al. published two papers describing an application of linear and non-linear modeling techniques for predicting biochemical oxygen demand of an aerated lagoon of a pulp and paper mill (161, 162). Li et al. presented a recursive nonlinear PLS algorithm for modeling non-linear behavior in processes (163).

Uncertainty Estimation in Multivariate Calibration Models

Fernández Pierna et al. described a number of methods useful for detecting outliers in prediction (164). They conclude that uncertainty estimation in combination with the convex hull method is a good complement to existing techniques, e.g. Mahalanobis distance and potential functions and robust techniques. In the paper by Linder and Sundberg calibration of hyphenated instruments is discussed with focus on predictions of concentrations of new samples (165).

Faber and Bro develop approximate expressions for calculating sample-specific standard errors of prediction for multi-way data (166). Simulated tri-linear data were used in the example and modeled with unfold-PLS and tri-PLS methods. Olivieri studied and described a simple approach for uncertainty propagation in multivariate calibration when data have been preprocessed by either OSC or net analyte pre-processing methods (167).

Fernández Pierna et al. and Faber et al. compare various approximate expressions for sample-specific standard error including Monte Carlo simulation methods on NIR data sets (168, 169). Faber also compared resampling methods for assessing the uncertainty in multivariate regression coefficients with an approximate variance expression method and concluded that it provides a viable alternative.

Andersen and Bro discussed the influence of different error contributions on the predictions of multivariate regression models with an example the estimation of water content of fish fillets from low-field NMR (170). In the paper by Reis and Saraiva multivariate linear regression methods that incorporate data uncertainties are compared (171).

REFERENCES

1. S. Wold, J. Trygg, A. Berglund, and H. Antti, Some recent developments in PLS modeling. *Chemometrics and Intelligent Laboratory Systems* 58 (2001):131–150.
2. R. J. Barnes, M. S. Dhanoa, and S. J. Lister, Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. *Applied Spectroscopy* 43 (1989):772–777.
3. P. Geladi, D. MacDougall, and H. Martens, Linearization and scatter-correction for near-infrared reflectance spectra of meat. *Applied Spectroscopy* 39 (1985):491–500.
4. H. Martens and E. Stark, Extended multiplicative signal correction and spectral interference subtraction: New preprocessing methods for near infrared spectroscopy. *Journal of Pharmaceutical and Biomedical Analysis* 9 (1991):625–635.

5. A. Savitzky and M. J. E. Golay, Smoothing and differentiation by simplified least squares procedures. *Analytical Chemistry* 36 (1964):1627–1632.
6. S. Wold, H. Antti, F. Lindgren, and J. Öhman, Orthogonal signal correction of near-infrared spectra. *Chemometrics and Intelligent Laboratory Systems* 44 (1998):175–185.
7. T. Fearn, On orthogonal signal correction. *Chemometrics and Intelligent Laboratory Systems* 50 (2000):47–52.
8. A. Höskuldsson, Variable and subset selection in PLS regression. *Chemometrics and Intelligent Laboratory Systems* 55 (2001):23–38.
9. J. A. Westerhuis, S. de Jong, and A. K. Smilde, Direct orthogonal signal correction. *Chemometrics and Intelligent Laboratory Systems* 56 (2001):13–25.
10. J. Trygg and S. Wold, Orthogonal projections to latent structures (O-PLS). *Journal of Chemometrics* 16 (2002):119–128.
11. B. Li, A. J. Morris, and E. B. Martin, Orthogonal signal correction: Algorithmic aspects and properties. *Journal of Chemometrics* 16 (2002):556–561.
12. R. N. Feudale, H. Tan, and S. D. Brown, Piecewise orthogonal signal correction. *Chemometrics and Intelligent Laboratory Systems* 63 (2002):129–138.
13. R. N. Feudale, Y. Liu, N. A. Woody, H. W. Tan, and S. D. Brown, Wavelet orthogonal signal correction. *Journal of Chemometrics* 19 (2005):55–63.
14. O. Svensson, T. Kourti, and J. F. MacGregor, An investigation of orthogonal signal correction algorithms and their characteristics. *Journal of Chemometrics* 16 (2002):176–188.
15. H. C. Goicoechea and A. C. Olivieri, A comparison of orthogonal signal correction and net analyte preprocessing methods. Theoretical and experimental study. *Chemometrics and Intelligent Laboratory Systems* 56 (2001):73–81.
16. L. Eriksson, J. Trygg, E. Johansson, R. Bro, and S. Wold, Orthogonal signal correction, wavelet analysis, and multivariate calibration of complicated process fluorescence data. *Analytica Chimica Acta* 420 (2000):181–195.
17. M. Blanco, J. Coello, I. Montoliu, and M. A. Romero, Orthogonal signal correction in near infrared calibration. *Analytica Chimica Acta* 434 (2001):125–132.
18. E. J. Samp, D. Sedin, and A. Foster, Enhanced NIR calibration for wort fermentability using orthogonal signal correction. *Journal of the Institute of Brewing* 109 (2003):16–26.
19. P. W. Hansen, Pre-processing method minimizing the need for reference analyses. *Journal of Chemometrics* 15 (2001):123–131.
20. C. A. Andersson, Direct orthogonalization. *Chemometrics and Intelligent Laboratory Systems* 47 (1999):51–63.
21. J. A. Fernández Pierna, D. L. Massart, O. E. de Noord, and P. Ricoux, Direct orthogonalization: Some case studies. *Chemometrics and Intelligent Laboratory Systems* 55 (2001):101–108.
22. D. Rodríguez, C. E. Boschetti, and A. C. Olivieri, Near-infrared spectroscopic determination of antioxidants and organic acids in rubbers assisted by a new multivariate calibration method based on direct orthogonalization. *Analyst* 127 (2002):304–309.
23. L. Stordrange, F. L. Libnau, D. Malthe-Sørenssen, and O. M. Kvalheim, Feasibility study of NIR for surveillance of a pharmaceutical process, including a study of different preprocessing techniques. *Journal of Chemometrics* 16 (2002):529–541.
24. T. V. Karstang and R. Manne, Optimized scaling—a novel approach to linear calibration with closed data sets. *Chemometrics and Intelligent Laboratory Systems* 14 (1992):165–173.
25. B. M. Beckwith-Hall, J. T. Brindle, R. H. Barton, M. Coen, E. Holmes, J. K. Nicholson, and H. Antti, Application of orthogonal signal correction to minimise the effects of physical and biological variation in high resolution H-1 NMR spectra of biofluids. *Analyst* 127 (2002):1283–1288.
26. T. Azzouz, A. Puigdoménech, M. Aragay, and R. Tauler, Comparison between different data pre-treatment methods in the analysis of forage samples using near-infrared diffuse reflectance spectroscopy and partial least-squares multivariate calibration method. *Analytica Chimica Acta* 484 (2003):121–134.
27. P. Fardim, M. M. C. Ferreira, and N. Durán, Determination of mechanical, and optical properties of eucalyptus kraft pulp by NIR spectrometry and multivariate calibration. *Journal of Wood Chemistry and Technology* 25 (2005):267–279.
28. F. Wulfert, W. T. Kok, O. E. de Noord, and A. K. Smilde, Correction of temperature-induced spectral variation by continuous piecewise direct standardization. *Analytical Chemistry* 72 (2000):1639–1644.
29. H. Martens, M. Høy, B. M. Wise, R. Bro, and P. B. Brockhoff, Prewhitening of data by covariance-weighted pre-processing. *Journal of Chemometrics* 17 (2003):153–165.
30. M. A. Arnold, G. W. Small, D. Xiang, J. Qui, and D. W. Murhammer, Pure component selectivity analysis of multivariate calibration models from near-infrared spectra. *Analytical Chemistry* 76 (2004):2583–2590.
31. H. Martens, J. P. Nielsen, and S. B. Engelsen, Light scattering and light absorbance separated by extended multiplicative signal correction. Application to near-infrared transmission analysis of powder mixtures. *Analytical Chemistry* 75 (2003):394–404.
32. J. M. Roger, F. Chauchard, and V. Bellon-Maurel, EPO-PLS external parameter orthogonalisation of PLS application to temperature-independent measurement of sugar content of intact fruits. *Chemometrics and Intelligent Laboratory Systems* 66 (2003):191–204.
33. J. A. Hageman, J. A. Westerhuis, and A. K. Smilde, Temperature robust multivariate calibration: An overview of methods for dealing with temperature influences on near infrared spectra. *Journal of Near Infrared Spectroscopy* 13 (2005):53–62.
34. M. Arab Chamjangali, G. Bagherian, and G. Azizi, Simultaneous determination of cobalt, nickel and palladium in micellar media using partial least square regression, and direct orthogonal signal correction. *Spectrochimica Acta Part A* 62 (2005):189–196.
35. C. Pizarro, I. Esteban-Díez, A. J. Nistal, and J.M. González-Sáiz, Influence of data pre-processing on the quantitative determination of the ash content and lipids in roasted coffee by near infrared spectroscopy. *Analytica Chimica Acta* 509 (2004):217–227.
36. K. Pöllänen, A. Häkkinen, S. P. Reinikainen, M. Louhi-Kultanen, and L. Nyström, ATR-FTIR in monitoring of crystallization processes: Comparison of indirect and direct OSC methods. *Chemometrics and Intelligent Laboratory Systems* 76 (2005):25–35.
37. H. L. Yu and J. F. MacGregor, Post processing methods (PLS-CCA): Simple alternatives to preprocessing methods (OSC-PLS). *Chemometrics and Intelligent Laboratory Systems* 73 (2004):199–205.

38. M. Zeaiter, J. M. Roger, V. Bellon-Maurel, and D. N. Rutledge, Robustness of models developed by multivariate calibration. Part I: The assessment of robustness. *Trends in Analytical Chemistry* 23 (2004):157–170.
39. M. Zeaiter, J. M. Roger, and V. Bellon-Maurel, Robustness of models developed by multivariate calibration. Part II: The influence of pre-processing methods. *Trends in Analytical Chemistry* 24 (2005):437–445.
40. F. Vogt, H. Steiner, K. Booksh, and B. Mizaikoff, Chemometric correction of drift effects in optical spectra. *Applied Spectroscopy* 58 (2004):683–692.
41. N. A. Woody, R. N. Feudale, A. J. Myles, and S. D. Brown, Transfer of multivariate calibrations between four near-infrared spectrometers using orthogonal signal correction. *Analytical Chemistry* 76 (2004):2595–2600.
42. A. Andrew and T. Fearn, Transfer by orthogonal projection: Making near-infrared calibrations robust to between-instrument variation. *Chemometrics and Intelligent Laboratory Systems* 72 (2004):51–56.
43. H. Leion, S. Folestad, M. Josefson, and A. Sparen, Evaluation of basic algorithms for transferring quantitative multivariate calibrations between scanning grating and FT NIR spectrometers. *Journal of Pharmaceutical and Biomedical Analysis* 37 (2005):47–55.
44. K. Janné, J. Pettersen, N. O. Lindberg, and T. Lundstedt, Hierarchical principal component analysis (PCA) and projection to latent structure (PLS) technique on spectroscopic data as a data pretreatment for calibration. *Journal of Chemometrics* 15 (2001):203–213.
45. B. R. Bakshi, Multiscale analysis and modeling using wavelets. *Journal of Chemometrics* 13 (1999):415–434.
46. T. Artursson, A. Hagman, S. Björk, J. Trygg, S. Wold, and S. P. Jacobsson, Study of preprocessing methods for the determination of crystalline phases in binary mixtures of drug substances by X- ray powder diffraction and multivariate calibration. *Applied Spectroscopy* 54 (2000):1222–1230.
47. B. K. Alsberg, Parsimonious multiscale classification models. *Journal of Chemometrics* 14 (2000):529–539.
48. M. Bos and J. A. M. Vrielink, The Wavelet Transform for Preprocessing Ir-Spectra in the Identification of Monosubstituted and Disubstituted Benzenes. *Chemometrics and Intelligent Laboratory Systems* 23 (1994):115–122.
49. R. K. H. Galvão, G. E. José, H. A. Dantas, M. C. U. Araujo, E. C. da Silva, H. M. Paiva, T. C. B. Saldanha, and E. S. O. N. de Souza, Optimal wavelet filter construction using X and Y data. *Chemometrics and Intelligent Laboratory Systems* 70 (2004): 1–10.
50. R. F. Luo, M. Misra, and D. M. Himmelblau, Sensor fault detection via multiscale analysis and dynamic PCA. *Industrial & Engineering Chemistry Research* 38 (1999):1489–1495.
51. M. N. Nonou and B. R. Bakshi, Multiscale linear modeling with application to inferential modeling and system identification. In *AIChE Annual Meeting*, Miami Beach, FL 1998.
52. A. Björk, and L.-G. Danielsson, Spectra of wavelet scale coefficients of process acoustic measurements as input for PLS modelling of pulp quality. In *7th Scandinavian Symposium on Chemometrics*, (Copenhagen, Denmark, 2001).
53. H. W. Tan and S. D. Brown, Multivariate calibration of spectral data using dual-domain regression analysis. *Analytica Chimica Acta* 490 (2003):291–301.
54. M. Cocchi, R. Seeber, and A. Ulrici, Multivariate calibration of analytical signals by WILMA (wavelet interface to linear modelling analysis). *Journal of Chemometrics* 17 (2003):512–527.
55. C. J. Coelho, R. K. H. Galvão, M. C. U. Araujo, M. F. Pimentel, and E. C. da Silva, A solution to the wavelet transform optimization problem in multicomponent analysis. *Chemometrics and Intelligent Laboratory Systems* 66 (2003):205–217.
56. D. Chen, B. Hu, X.G. Shao, and Q. D. Su, Removal of major interference sources in aqueous near-infrared spectroscopy techniques. *Analytical and Bioanalytical Chemistry* 379 (2004):143–148.
57. R. D. Hall, Plant metabolomics: From holistic hope, to hype, to hot topic. *The New Phytologist* 169 (2006):453–468.
58. S. Rochfort, Metabolomics reviewed: A new “Omics” platform technology for systems biology and implications for natural products research. *Journal of Natural Products* 68 (2005):1813–1820.
59. S. M. Stephens and J. Rung, Advances in systems biology: Measurement, modeling and representation. *Current Opinion in Drug Discovery & Development* 9 (2006):240–250.
60. Y. Z. Liang and O. M. Kvalheim, Resolution of two-way data: Theoretical background and practical problem-solving – Part 1: Theoretical background and methodology. *Fresenius Journal of Analytical Chemistry* 370 (2001):694–704.
61. J. M. Halket, A. Przyborowska, S. E. Stein, W. G. Mallard, S. Down, and R. A. Chalmers, Deconvolution gas chromatography mass spectrometry of urinary organic acids—Potential for pattern recognition and automated identification of metabolic disorders. *Rapid Communications in Mass Spectrometry* 13 (1999):279–284.
62. P. Jonsson, A. I. Johansson, J. Gullberg, J. Trygg, J. A. B. Grung, S. Marklund, M. Sjöström, H. Antti, and T. Moritz, High-throughput data analysis for detecting and identifying differences between samples in GC/MS-based metabolomic analyses. *Analytical Chemistry* 77 (2005):5635–5642.
63. H. L. Shen, B. Grung, O. M. Kvalheim, and I. Eide, Automated curve resolution applied to data from multi-detection instruments. *Analytica Chimica Acta* 446 (2001):313–328.
64. I. Eide, G. Neverdal, B. Thorvaldsen, H. L. Shen, B. Grung, and O. Kvalheim, Resolution of GC-MS data of complex PAC mixtures and regression modeling of mutagenicity by PLS. *Environmental Science & Technology* 35 (2001):2314–2318.
65. M. Jalali-Heravi and M. Vosough, Characterization and determination of fatty acids in fish oil using gas chromatography-mass spectrometry coupled with chemometric resolution techniques. *Journal of Chromatography A* 1024 (2004):165–176.
66. N. Schauer, D. Steinhauser, S. Strelkov, D. Schomburg, G. Allison, T. Moritz, K. Lundgren, U. Roessner-Tunali, M. G. Forbes, L. Willmitzer, A. R. Fernie, and J. Kopka, GC-MS libraries for the rapid identification of metabolites in complex biological samples. *FEBS Letters* 579 (2005):1332–1337.
67. H. Idborg-Björkman, P. O. Edlund, O. M. Kvalheim, I. Schuppe-Koistinen, and S. P. Jacobsson, Screening of biomarkers in rat urine using LC/electrospray ionization-MS and two-way data analysis. *Analytical Chemistry* 75 (2003):4784–4792.
68. P. V. van Zomeren, H. Darwinkel, P. M. J. Coenegracht, and G. J. de Jong, Comparison of several curve resolution methods for drug impurity profiling using high-performance liquid chromatography with diode array detection. *Analytica Chimica Acta* 487 (2003):155–170.

69. A. C. Whitson and M. Maeder, Exhaustive evolving factor analysis (E-EFA). *Journal of Chemometrics* 15 (2001):475–484.
70. Z. P. Chen, J. Morris, E. Martin, R. Q. Yu, Y. Z. Liang, and F. Gong, Recursive evolving spectral projection for revealing the concentration windows of overlapping peaks in two-way chromatographic experiments. *Chemometrics and Intelligent Laboratory Systems* 72 (2004):9–19.
71. M. Wasim and R. G. Brereton, Application of multivariate curve resolution methods to on-flow LC-NMR. *Journal of Chromatography A* 1096 (2005):2–15.
72. J. H. Jiang, Y. Z. Liang, and Y. Ozaki, On simplex-based method for self-modeling curve resolution of two-way data. *Chemometrics and Intelligent Laboratory Systems* 65 (2003):51–65.
73. F. Berbel, E. Kapoya, J. M. Diaz-Cruz, C. Arino, M. Esteban, and R. Tauler, Multivariate resolution of coeluted peaks in hyphenated liquid chromatography – Linear sweep voltammetry. *Electroanalysis* 15 (2003):499–508.
74. E. Peré-Trepat, S. Lacorte, and R. Tauler, Solving liquid chromatography mass spectrometry coelution problems in the analysis of environmental samples by multivariate curve resolution. *Journal of Chromatography A* 1096 (2005):111–122.
75. M. Garrido, I. Lázaro, M. S. Larrechi, and F. X. Rius, Multivariate resolution of rank-deficient near-infrared spectroscopy data from the reaction of curing epoxy resins using the rank augmentation strategy and multivariate curve resolution alternating least squares approach. *Analytica Chimica Acta* 515 (2004):65–73.
76. S. Richards, M. Ropic, D. Blackmond, and A. Walmsley, Quantitative determination of the catalysed asymmetric transfer hydrogenation of 1-methyl-6,7-dimethoxy-3,4-dihydroisoquinoline using in situ FTIR and multivariate curve resolution. *Analytica Chimica Acta* 519 (2004):1–9.
77. C. Ruckebusch, L. Duponchel, J.-P. Huvenne, and J. Saurina, Multivariate curve resolution of step-scan FTIR spectral data. *Vibrational Spectroscopy* 35 (2004):21–26.
78. T. H. Jiang, Y. Liang, and Y. Ozaki, Principles and methodologies in self-modeling curve resolution. *Chemometrics and Intelligent Laboratory Systems* 71 (2004): 1–12.
79. H. Li, F. Zhang, and J. Havel, Quantification of analytes in overlapping peaks from capillary electrophoresis using multivariate curve resolution-alternating least squares methods. *Electrophoresis* 24 (2003):3107–3115.
80. J. L. Beltrán, E. Jiménez-Lozano, D. Barron, and J. Barbosa, Determination of quinolone antimicrobial agents in strongly overlapped peaks from capillary electrophoresis using multivariate calibration methods. *Analytica Chimica Acta* 501 (2004):137–141.
81. P. V. van Zomeren, H. J. Metting, P. M. J. Coenegracht, and G. J. de Jong, Simultaneous resolution of overlapping peaks in high-performance liquid chromatography and micellar electrokinetic chromatography with diode array detection using augmented iterative target transformation factor analysis. *Journal of Chromatography A* 1096 (2005):165–176.
82. V. P. Andreev, T. Rejtar, H. S. Chen, E. V. Moskovets, A. R. Ivanov, and B. L. Karger, A universal denoising and peak picking algorithm for LC-MS based on matched filtration in the chromatographic time domain. *Analytical Chemistry* 75 (2003):6314–6326.
83. R. S. Plumb, C. L. Stumpf, J. H. Granger, J. Castro-Perez, J. N. Haselden, and G. J. Dear, Use of liquid chromatography/time-of-flight mass spectrometry and multivariate statistical analysis shows promise for the detection of drug metabolites in biological fluids. *Rapid Communications in Mass Spectrometry* 17 (2003):2632–2638.
84. G. Vivó-Truyols, J. R. Torres-Lapasió, A. M. van Nederkassel, Y. Vander Heyden, and D. L. Massart, Automatic program for peak detection and deconvolution of multi-overlapped chromatographic signals – Part I: Peak detection. *Journal of Chromatography A* 1096 (2005):133–145.
85. G. Vivó-Truyols, J. R. Torres-Lapasió, A. M. van Nederkassel, Y. Vander Heyden, and D. L. Massart, Automatic program for peak detection and deconvolution of multi-overlapped chromatographic signals – Part II: Peak model and deconvolution algorithms. *Journal of Chromatography A* 1096 (2005):146–155.
86. J. Allen, H. M. Davey, D. Broadhurst, J. K. Heald, J. J. Rowland, S. G. Oliver, and D. B. Kell, Highthroughput classification of yeast mutants for functional genomics using metabolic footprinting. *Nature Biotechnology* 21 (2003):692–696.
87. J. H. Christensen, J. Mortensen, A. B. Hansen, and O. Andersen, Chromatographic preprocessing of GC-MS data for analysis of complex chemical mixtures. *Journal of Chromatography A* 1062 (2005):113–123.
88. P. Jonsson, J. Gullberg, A. Nordström, M. Kusano, M. Kowalczyk, M. Sjöström, and T. Moritz, A strategy for identifying differences in large series of metabolomic samples analyzed by GC/MS. *Analytical Chemistry* 76 (2004):1738–1745.
89. P. Jonsson, S. J. Bruce, T. Moritz, J. Trygg, M. Sjöström, R. Plumb, J. Granger, E. Maibaum, J. K. Nicholson, E. Holmes, and H. Antti, Extraction, interpretation and validation of information for comparing samples in metabolic LC/MS data sets. *Analyst* 130 (2005):701–707.
90. B. Nordén, P. Broberg, C. Lindberg, and A. Plymoth, Analysis and understanding of high-dimensionality data by means of multivariate data analysis. *Chemistry & Biodiversity* 2 (2005):1487–1494.
91. H. Idborg, L. Zamani, P. O. Edlund, I. Schuppe-Koistinen, and S. P. Jacobsson, Metabolic fingerprinting of rat urine by LC/MS Part 2. Data pretreatment methods for handling of complex data. *Journal of Chromatography B* 828 (2005):14–20.
92. A. C. Olivieri, On a versatile second-order multivariate calibration method based on partial least-squares and residual bilinearization: Second-order advantage and precision properties. *Journal of Chemometrics* 19 (2005):253–265.
93. G. R. Flåten, and A. D. Walmsley, Using design of experiments to select optimum calibration model parameters. *Analyst* 128 (2003):935–943.
94. S. Wold, M. Josefson, J. Gottfries, and A. Linusson, The utility of multivariate design in PLS modeling. *Journal of Chemometrics* 18 (2004):156–165.
95. K. Jørgensen and T. Næs, A design and analysis strategy for situations with uncontrolled raw material variation. *Journal of Chemometrics* 18 (2004):45–52.
96. J. H. Jiang, R. J. Berry, H. W. Siesler, and Y. Ozaki, Wavelength interval selection in multicomponent spectral analysis by moving window partial least-squares regression with applications to mid-infrared and near-infrared spectroscopic data. *Analytical Chemistry* 74 (2002):3555–3565.
97. S. Kasemsumran, Y. P. Du, K. Murayama, M. Huehne, and Y. Ozaki, Simultaneous determination of human serum albumin, gamma-globulin, and glucose in a phosphate buffer solution by

near-infrared spectroscopy with moving window partial least-squares regression. *Analyst* 128 (2003):1471–1477.

98. M. H. Zhang, Q. S. Xu, and D. L. Massart, Averaged and weighted average partial least squares. *Analytica Chimica Acta* 504 (2004):279–289.
99. M. H. Zhang, Q. S. Xu, and D. L. Massart, Boosting partial least squares. *Analytical Chemistry* 77 (2005):1423–1431.
100. M. Kearns, M. Li, and L. Valiant, Cryptographic limitations on learning Boolean formulae and finite automata. *Journal of the Association for Computing Machinery* 41 (1994):67–95.
101. R. Marbach, A new method for multivariate calibration. *Journal of Near Infrared Spectroscopy* 13 (2005):241–254.
102. E. M. Qannari and M. Hanafi, A simple continuum regression approach. *Journal of Chemometrics* 19 (2005):387–392.
103. S. Serneels, P. Filzmoser, C. Croux, and P. J. Van Espen, Robust continuum regression. *Chemometrics and Intelligent Laboratory Systems* 76 (2005):197–204.
104. S. de Jong, B. M. Wise, and N. Lawrence Ricker, Canonical partial least squares and continuum power regression. *Journal of Chemometrics* 15 (2001):85–100.
105. C. Campbell, Kernel methods: A survey of current techniques. *Neurocomputing* 48 (2002):63–84.
106. V. Vapnik, *The nature of statistical learning theory* (New York: Springer, 1995).
107. V. Vapnik, *Statistical learning theory* (New York: Wiley, 1998).
108. A. I. Belousov, S. A. Verzakov, and J. von Frese, A flexible classification approach with optimal generalisation performance: Support vector machines. *Chemometrics and Intelligent Laboratory Systems* 64 (2002):15–25.
109. R. P. Cogdill and P. Dardenne, Least-squares support vector machines for chemometrics: An introduction and evaluation. *Journal of Near Infrared Spectroscopy* 12 (2004):93–100.
110. S. Zomer, M. D. N. Sanchez, R. G. Brereton, and J. L. Pérez Pavón, Active learning support vector machines for optimal sample selection in classification. *Journal of Chemometrics* 18 (2004):294–305.
111. J. Trygg, O2-PLS for qualitative and quantitative analysis in multivariate calibration. *Journal of Chemometrics* 16 (2002):283–293.
112. T. Verron, R. Sabatier, and R. Joffre, Some theoretical properties of the O-PLS method. *Journal of Chemometrics* 18 (2004):62–68.
113. R. Ergon, PLS post-processing by similarity transformation (PLS + ST): A simple alternative to OPLS. *Journal of Chemometrics* 19 (2005): 1–4.
114. O. Cloarec, M. E. Dumas, A. Craig, R. H. Barton, J. Trygg, J. Hudson, C. Blancher, D. Gauguier, J. C. Lindon, E. Holmes, and J. Nicholson, Statistical total correlation spectroscopy: An exploratory approach for latent biomarker identification from metabolic H-1 NMR data sets. *Analytical Chemistry* 77 (2005):1282–1289.
115. J. Trygg, Prediction and spectral profile estimation in multivariate calibration. *Journal of Chemometrics* 18 (2004):166–172.
116. A. Lorber, Error propagation and figures of merit for quantification by solving matrix equations. *Analytical Chemistry* 58 (1986):1167–1172.
117. A. Lorber, K. Faber, and B. R. Kowalski, Net analyte signal calculation in multivariate calibration. *Analytical Chemistry* 69 (1997):1620–1626.
118. C. D. Brown, Discordance between net analyte signal theory and practical multivariate calibration. *Analytical Chemistry* 76 (2004):4364–4373.
119. N. M. Faber, Efficient computation of net analyte signal vector in inverse multivariate calibration models. *Analytical Chemistry* 70 (1998):5108–5110.
120. J. Ferré, S. D. Brown, and F. X. Rius, Improved calculation of the net analyte signal in inverse multivariate calibration. *Journal of Chemometrics* 15 (2001):537–553.
121. R. Bro and C. M. Andersen, Theory of net analyte signal vectors in inverse regression. *Journal of Chemometrics* 17 (2003):646–652.
122. A. Muñoz de la Peña, A. Espinosa-Mansilla, M.I. Acedo Valenzuela, H. C. Goicoechea, and A. C. Olivieri, Comparative study of net analyte signal-based methods and partial least squares for the simultaneous determination of amoxycillin and clavulanic acid by stopped-flow kinetic analysis. *Analytica Chimica Acta* 463 (2002):75–88.
123. G. Vivó-Truyols, J. R. Torres-Lapasió, and M. C. García-Alvarez-Coque, Net analyte signal as a deconvolution-oriented resolution criterion in the optimisation of chromatographic techniques. *Journal of Chromatography A* 991 (2003):47–59.
124. M. Martínez Galera, D. Picón Zamora, J. L. Martínez Vidal, A. Garrido Frenich, A. Espinosa-Mansilla, A. Muñoz de la Peña, and F. Salinas López, Determination of carbendazim, thiabendazole and fuberidazole using a net analyte signal-based method. *Talanta* 59 (2003):1107–1116.
125. J. Ferré, and N. M. Faber, Net analyte signal calculation for multivariate calibration. *Chemometrics and Intelligent Laboratory Systems* 69 (2003):123–136.
126. E. T. S. Skibsted, H. F. M. Boelens, J. A. Westerhuis, A. K. Smilde, N. W. Broad, D.R. Rees, and D. T. Witte, Net analyte signal based statistical quality control. *Analytical Chemistry* 77 (2005):7103–7114.
127. C. E. Boschetti, and A. C. Olivieri, Net analyte preprocessing: A new and versatile multivariate calibration technique. Analysis of mixtures of rubber antioxidants by NIR spectroscopy. *Journal of Near Infrared Spectroscopy* 9 (2001):245–254.
128. H. C. Goicoechea and A. C. Olivieri, Sustained prediction ability of net analyte preprocessing methods using reduced calibration sets. Theoretical and experimental study involving the spectrophotometric analysis of multicomponent mixtures. *Analyst* 126 (2001):1105–1112.
129. M. Andre, Multivariate analysis and classification of the chemical quality of 7-aminocephalosporanic acid using near-infrared reflectance spectroscopy. *Analytical Chemistry* 75 (2003):3460–3467.
130. R. Y. Chen, R. Matsunaga, K. Ishikawa, and H. Zhang, Main inorganic component measurement of seawater using near-infrared spectroscopy. *Applied Spectroscopy* 57 (2003):1399–1406.
131. G. Févotte, J. Calas, F. Puel, and C. Hoff, Applications of NIR spectroscopy to monitoring and analyzing the solid state during industrial crystallization processes. *International Journal of Pharmaceutics* 273 (2004):159–169.
132. K. Pöllänen, A. Häkkinen, S. P. Reinikainen, J. Rantanen, M. Karjalainen, M. Louhi-Kultanen, and L. Nyström, IR spectroscopy together with multivariate data analysis as a process analytical tool for in-line monitoring of crystallization process

and solid-state analysis of crystalline product. *Journal of Pharmaceutical and Biomedical Analysis* 38 (2005):275–284.

133. L. O. Rodrigues, J. L. Marques, J. P. Cardoso and J. C. Menezes, From lab to plant: A scale-independent NIR calibration to monitor an ionic exchange column. *Chemometrics and Intelligent Laboratory Systems* 75 (2005):101–108.
134. R. P. Cogdill, C. A. Anderson, M. Delgado-Lopez, D. Molseed, R. Chisholm, R. Bolton, T. Herkert, A. M. Afnan, and J. K. Drennen, Process analytical technology case study part I: Feasibility studies for quantitative near-infrared method development. *AAPS PharmSciTech* 6 (2005).
135. R. Ergon, PLS score-loading correspondence and a bi-orthogonal factorization. *Journal of Chemometrics* 16 (2002):368–373.
136. T. A. Lestander and C. Rhen, Multivariate NIR spectroscopy models for moisture, ash and calorific content in biofuels using bi-orthogonal partial least squares regression. *Analyst* 130 (2005):1182–1189.
137. F. Estienne, D. L. Massart, N. Zanier-Szydlowski, and P. Marteau, Multivariate calibration with Raman spectroscopic data: A case study. *Analytica Chimica Acta* 424 (2000):185–201.
138. S. Sivakesava, J. Irudayaraj, and A. Demirci, Monitoring a bioprocess for ethanol production using FT-MIR and FT-Raman spectroscopy. *Journal of Industrial Microbiology & Biotechnology* 26 (2001):185–190.
139. A. O'Grady, A. C. Dennis, D. Denvir, J. J. McGarvey, and S. E. J. Bell, Quantitative Raman spectroscopy of highly fluorescent samples using pseudosecond derivatives and multivariate analysis. *Analytical Chemistry* 73 (2001):2058–2065.
140. J. Johansson, S. Pettersson, and S. Folestad, Characterization of different laser irradiation methods for quantitative Raman tablet assessment. *Journal of Pharmaceutical and Biomedical Analysis* 39 (2005):510–516.
141. S. Romero-Torres, J. D. Pérez-Ramos, K. R. Morris, and E. R. Grant, Raman spectroscopic measurement of tablet-to-tablet coating variability. *Journal of Pharmaceutical and Biomedical Analysis* 38 (2005):270–274.
142. B. Karlberg and R. Torgrip, Increasing the scope and power of flow-injection analysis through chemometric approaches. *Analytica Chimica Acta* 500 (2003):299–306.
143. H. Swierenga, F. Wülfert, O. E. de Noord, A. P. de Weijer, A. K. Smilde, and L. M. C. Buydens, Development of robust calibration models in near infra-red spectrometric applications. *Analytica Chimica Acta* 411 (2000):121–135.
144. E. V. Thomas and N. X. Ge, Development of robust multivariate calibration models. *Technometrics* 42 (2000):168–177.
145. F. Despagne, D. L. Massart, and P. Chabot, Development of a robust calibration model for nonlinear in-line process data. *Analytical Chemistry* 72 (2000):1657–1665.
146. C. R. Mittermayr, H. W. Tan, and S. D. Brown, Robust calibration with respect to background variation. *Applied Spectroscopy* 55 (2001):827–833.
147. M. S. Larrechi and M. P. Callao, Strategy for introducing NIR spectroscopy and multivariate calibration techniques in industry. *Trends in Analytical Chemistry* 22 (2003):634–640.
148. N. S. Sahni, T. Isaksson, and T. Næs, Comparison of methods for transfer of calibration models in near-infared spectroscopy: A case study based on correcting path length differences using fiber-optic transmittance probes in in-line near-infrared Spectroscopy. *Applied Spectroscopy* 59 (2005):487–495.
149. C. B. Zachariassen, J. Larsen, F. van den Berg, and S. B. Engelsen, Use of NIR spectroscopy and chemometrics for on-line process monitoring of ammonia in Low Methoxylated Amidated pectin production. *Chemometrics and Intelligent Laboratory Systems* 76 (2005):149–161.
150. M. Andersson, S. Folestad, J. Gottfries, M. O. Johansson, M. Josefson, and K. G. Wahlund, Quantitative analysis of film coating in a fluidized bed process by in line NIR spectrometry and multivariate batch calibration. *Analytical Chemistry* 72 (2000):2099–2108.
151. A. Nordon, C. A. McGill, and D. Littlejohn, Evaluation of low-field nuclear magnetic resonance spectrometry for at-line process analysis. *Applied Spectroscopy* 56 (2002):75–82.
152. S. Y. Chang, E. H. Baughman, and B. C. McIntosh, Implementation of locally weighted regression to maintain calibrations on FT-NIR analyzers for industrial processes. *Applied Spectroscopy* 55 (2001):1199–1206.
153. H. W. Cho and K. J. Kim, A method for predicting future observations in the monitoring of a batch process. *Journal of Quality Technology* 35 (2003):59–69.
154. J. A. Lopes and J. C. Menezes, Industrial fermentation end-product modelling with multilinear PLS. *Chemometrics and Intelligent Laboratory Systems* 68 (2003):75–81.
155. A. Skoglund, W. Kessler, R. W. Kessler, A. Brundin, and C. F. Mandenius, On-line spectroscopic measurements of wood chips before a continuous digester. *Chemometrics and Intelligent Laboratory Systems* 70 (2004):129–135.
156. P. Jonsson, M. Sjöström, L. Wallbäcks, and H. Antti, Strategies for implementation and validation of on-line models for multivariate monitoring and control of wood chip properties. *Journal of Chemometrics* 18 (2004):203–207.
157. U. Thissen, M. Pepers, B. Ustun, W. J. Melssen, and L. M. C. Buydens, Comparing support vector machines to PLS for spectral regression applications. *Chemometrics and Intelligent Laboratory Systems* 73 (2004):169–179.
158. J. Rantanen, E. Räsänen, O. Antikainen, J. P. Mannermaa, and J. Yliruusi, In-line moisture measurement during granulation with a four-wavelength near-infrared sensor: An evaluation of process-related variables and a development of non-linear calibration model. *Chemometrics and Intelligent Laboratory Systems* 56 (2001):51–58.
159. F. Chauchard, R. Cogdill, S. Roussel, J. M. Roger, and V. Bellon-Maurel, Application of LS-SVM to non-linear phenomena in NIR spectroscopy: Development of a robust and portable sensor for acidity prediction in grapes. *Chemometrics and Intelligent Laboratory Systems* 71 (2004):141–150.
160. U. Thissen, B. Ustun, W. J. Melssen, and L. M. C. Buydens, Multivariate calibration with least-squares support vector machines. *Analytical Chemistry* 76 (2004):3099–3105.
161. L. Stordrange, O. M. Kvalheim, P. A. Hassel, D. Malthe-Sorenssen, and F. O. Libnau, A comparison of techniques for modelling data with non-linear structure. *Journal of Near Infrared Spectroscopy* 11 (2003):55–70.
162. K. P. Oliveira-Esquerre, D. E. Seborg, R. E. Bruns, and M. Mori, Application of steady-state and dynamic modeling for the prediction of the BOD of an aerated lagoon at a pulp and paper mill—Part I. Linear approaches. *Chemical Engineering Journal* 104 (2004):73–81.

163. K. P. Oliveira-Esquerre, D. E. Seborg, M. Mori, and R. E. Bruns, Application of steady-state and dynamic modeling for the prediction of the BOD of an aerated lagoon at a pulp and paper mill—Part II. Nonlinear approaches. *Chemical Engineering Journal* 105 (2004):61–69.
164. C. F. Li, H. Ye, G.Z. Wang, and J. Zhang, A recursive Nonlinear PLS algorithm for adaptive nonlinear process modeling. *Chemical Engineering & Technology* 28 (2005):141–152.
165. J. A. Fernández Pierna, F. Wahl, O.E. de Noord, and D. L. Massart, Methods for outlier detection in prediction. *Chemometrics and Intelligent Laboratory Systems* 63 (2002):27–39.
166. M. Linder and R. Sundberg, Precision of prediction in second-order calibration, with focus on bilinear regression methods. *Journal of Chemometrics* 16 (2002):12–27.
167. N. M. Faber and R. Bro, Standard error of prediction for multiway PLS 1. Background and a simulation study. *Chemometrics and Intelligent Laboratory Systems* 61 (2002):133–149.
168. A. C. Olivieri, A simple approach to uncertainty propagation in preprocessed multivariate calibration. *Journal of Chemometrics* 16 (2002):207–217.
169. N. M. Faber, X.-H. Song, and P. K. Hopke, Sample-specific standard error of prediction for partial least squares regression. *Trends in Analytical Chemistry* 22 (2003):330–334.
170. J. A. Fernández Pierna, L. Jin, F. Wahl, N.M. Faber, and D. L. Massart, Estimation of partial least squares regression prediction uncertainty when the reference values carry a sizeable measurement error. *Chemometrics and Intelligent Laboratory Systems* 65 (2003):281–291.
171. C. M. Andersen and R. Bro, Quantification and handling of sampling errors in instrumental measurements: A case study. *Chemometrics and Intelligent Laboratory Systems* 72 (2004):43–50.
172. M. S. Reis and P. M. Saraiva, A comparative study of linear regression methods in noisy environments. *Journal of Chemometrics* 18 (2004):526–536.