Pergamon

S0098-1354(96)00170-6

INFERRING CONCENTRATIONS ON-LINE FROM NEAR-INFRARED SPECTRA: NONLINEAR CALIBRATION *VIA* MID-INFRARED MEASUREMENTS

M. AMRHEIN, B. SRINIVASAN, D. BONVIN and M.M. SCHUMACHER†

E-mail: name@ia.epfl.ch

Institut d'Automatique, École Polytechnique Fédérale de Lausanne
CH-1015 Lausanne, Switzerland
†Lonza AG, CH-3930 Visp, Switzerland

Abstract

Near-Infrared (NIR) spectroscopic methods for on-line concentration estimation are gaining popularity in chemical production. The problem with NIR data, however, is its nonlinear nature rendering standard factor-analytical (FA) techniques inapplicable to estimate the concentrations. To infer concentrations on-line from NIR spectra, the following steps are proposed: (i) measure both the NIR and Mid-Infrared (MIR) spectra of unknown reaction mixtures in the laboratory (MIR spectroscopy has the desirable property of linearity with respect to concentrations), (ii) apply FA techniques to MIR data and infer concentrations, (iii) calibrate NIR against the estimated concentrations by using standard nonlinear regression methods and (iv) use the calibration model in subsequent production runs. Experimental results are presented to illustrate the efficiency of this method.

1 INTRODUCTION

Devices used for the analysis of chemical reaction systems can be classified into two categories: (a) destructive and (b) non-destructive methods. Destructive methods, such as Flow-Injection Analysis (FIA), are often used on-line to analyze samples taken from an on-going reaction. Such a method can modify the sample during the course of analysis and, furthermore, is often slow and prohibitively expensive. Hence, the number of measurements that can be obtained is limited. In contrast, non-destructive methods (such as measurement of pressure, temperature, conductivity or absorbance) offer the advantage of analyzing the product on-line at the process level *in-situ*. Also, a large number of measurements can be taken without any additional cost. The disadvantage, however, is the non-specific nature of the measurements obtained. Thus, in the recent past, measurements from various devices have been combined or calibrated against each others for more efficient usage.

On the one hand, various second-order calibration methods have been proposed (Kiers and Smilde 1995) to deal with different types of 'second-order instruments'. A good example is the combined use of High Pressure Liquid Chromatography (HPLC) and spectra in the Ultra-Violet/Visual (UV/VIS) region for the quantitative kinetic analysis of photo-induced reactions (Weiss 1991). On the other hand, correlation between Mid-Infrared (MIR) and Near-Infrared (NIR) spectra is addressed in Barton II *et al.* (1992) and Windig *et al.* (1995). Also, prediction of MIR spectra from on-line NIR spectra for a better interpretation of the results in a polymer production process is proposed in Lew and Balke (1993). However, all these methods (second-order calibration and correlation spectroscopy) assume linearity between the two data sets involved.

Among the various spectroscopic measurements, the technique best suited in the production environment is NIR spectroscopy since it uses fiber-optic technology. The problem with NIR data, however, is its nonlinear nature, i.e. Beer's law not being valid (Burns and Ciurczak 1992). Hence, standard factor-analytical techniques cannot be used to estimate the concentrations. Also, the nonlinearity renders correlation and second-order calibration techniques unsuitable for *combining* NIR with other indirect concentration measurements. Furthermore, calibration methods for obtaining the nonlinear relationships between NIR spectroscopic measurements and concentrations require a large number of calibration points. Hence, NIR data can only be combined with non-destructive measurements. The method of choice for cross-calibration with NIR seems to be MIR spectroscopy. Contrary to NIR spectra, pretreatment techniques such as using integrated areas and corrections for peak shift or peak broadening can be applied to MIR spectra. With this, MIR spectroscopy has the desirable property of linearity with respect to the concentrations (Griffiths and de Haseth 1987). However, it necessitates a rigid optical mirror system, and it is therefore limited to laboratory use.

The aim of this paper is to obtain a model that predicts concentrations of absorbing species on-line from NIR spectroscopic measurements. The concentrations can then be used for modeling, process monitoring, control and on-line optimization purposes. The calibration is performed with respect to the concentrations estimated using MIR measurements.

Section 2 presents the various spectral models that are used, while Section 3 outlines the proposed approach. Factor analysis of MIR measurements and the nonlinear calibration are dealt with in Sections 4 and 5, respectively. Section 6 studies a real-life example: the esterification of acetic acid is investigated via both NIR and MIR spectroscopic measurements. Conclusions and significance are given in Section 7.

2 SPECTRAL MEASUREMENT MODELS

Let $a_N(k)$ be an L_N–dimensional vector representing the electromagnetic spectrum in the NIR region at L_N wavelengths at the observation instant k. Similarly, let $a_M(k)$ be an L_M–dimensional vector representing the electromagnetic spectrum in the MIR region at L_M wavenumbers at the observation instant k. The corresponding spectroscopic models are given next.

1. **MIR spectroscopy:**

   For a unit pathlength and Beer's law being valid, $a_M(k)$ can be expressed as:

   $$a_M^T(k) = c^T(k)\, E_M.$$ (1)

   where $c(k)$ is an S–dimensional concentration vector, E_M is an $S \times L_M$ matrix of pure–component spectra (or extinction coefficients), and S is the number of absorbing species.

   For K observations, (1) can be written in matrix form as:

   $$A_M = C\, E_M$$ (2)
   $$C = A_M\, E_M^+$$ (3)

   with A_M being the $K \times L_M$ data matrix and C the $K \times S$ concentration matrix. E_M^+ represents the pseudo–inverse of the matrix E_M.

2. **NIR spectroscopy:**

   It is assumed that instead of Lambert–Beer's law (1), $a_N(k)$ is an L_N–dimensional vector with each element being a nonlinear function of c. The underlying physical functional relationship is described by an unknown nonlinear function f_p parameterized by Θ:

   $$a_N(k) = f_p(c(k), \Theta).$$ (4)

   A spectroscopic sensor model f_s is the inverse of the physical functional relationship and is defined as:

   $$c(k) = f_s(a_N(k), \Theta).$$ (5)

   The constants obey the inequality $L_N, L_M, K > S$, which will be used in the remaining part of this paper. It is furthermore assumed that the species absorbing in the MIR and NIR range are the same. This can be justified since bands in the NIR range are overtones or linear combinations of the ones in the MIR range.

In practical applications, the spectroscopic data need to be pretreated to minimize the effect of noise. Data–pretreatment includes mean–centering, differentiation, normalization, projection on certain subspaces to name a few. Care should be taken to ensure that the calibration data and the data used for prediction undergo the same data–pretreatment steps. For example, when mean–centering is performed, the mean of the calibration data has to be subtracted and not the mean of the prediction data.

3 THE PROPOSED APPROACH

Due to the fact that one sensor is linear and the other one nonlinear, care must be used when trying to combine the various measurements. The two statistical techniques that will be used to analyze the data are: (i) factor analysis (FA) for the estimation of concentrations from MIR data and (ii) nonlinear PCR for the calibration of NIR with respect to the estimated concentrations. The major assumption made in this context is that the calibration model, which is derived from a laboratory–scale experiment, is also valid during the production runs.

The final goal is to use NIR measurements on–line during production runs. The steps that will be taken towards that goal are : (i) measure in parallel both the NIR and MIR spectra of unknown reaction mixtures from laboratory–scale experiments, (ii) apply FA techniques to MIR data and infer the concentration profiles, (iii) calibrate NIR with respect to the estimated concentrations by using standard nonlinear regression methods, and (iv) use the calibration model in subsequent production runs. The FA part and the nonlinear calibration part will be discussed in some detail in the following sections.

4 FACTOR ANALYSIS OF MIR SPECTROSCOPIC DATA

In factor analysis, the following steps are pursued:
   1. determination of S
   2. determination and rotation of the abstract spaces to obtain the physical spaces.

As a first step, one seeks to obtain the number of absorbing species by analyzing the rank of the data matrix. If a data matrix is noisy, the pseudo–rank (essential rank) has to be used for further analysis. A comprehensive survey of the methods available for pseudo–rank determination is given in Malinowski (1991). However, the major problem in this context is that, in many practical applications, the pseudo–rank does not reveal the number of absorbing species. Such a rank deficiency arises in reacting mixtures and/or results from certain types of data pretreatment (see Amrhein et al. (1995) for a comprehensive analysis). The same reference also illustrates how the rank can be augmented by using multiple process runs or addition of mixtures of known concentrations. In this work, it will be assumed that rank augmentation techniques have been applied so that the data matrix has a (pseudo–)rank equal to S.

To determine the abstract spaces, a popular method is to use the singular value decomposition (SVD) and neglect the non–significant singular values and their corresponding singular vectors:

$$\mathbf{A}_M \overset{SVD}{=} (\mathbf{V}_1 \Sigma_1) \mathbf{W}_1^T = \mathbf{C}_a (\mathbf{E}_M)_a \tag{6}$$

where the subscripts $(\bullet)_1$ and $(\bullet)_a$ denote a dominant quantity of the SVD and an abstract quantity, respectively. The span of the right singular vectors, $\mathbf{W}_1$, constitutes the abstract pure–component spectra space, $(\mathbf{E}_M)_a$. Also, the row span of $\mathbf{V}_1$ constitutes the abstract concentration space, $\mathbf{C}_a$. The physical spaces $\mathbf{C}$ and $\mathbf{E}_M$ are then obtained by rotation of the abstract spaces either using targets or optimization criteria (see Malinowski (1991)):

$$\mathbf{A}_M = \mathbf{C}_a \mathbf{T} \mathbf{T}^{-1} (\mathbf{E}_M)_a \quad \text{such that } \mathbf{T}^{-1} (\mathbf{E}_M)_a = \mathbf{E}_M \tag{7}$$

where $\mathbf{T}$ is an $S \times S$ rotation matrix.

To obtain the rotation matrix using targets, a pure–component spectra target is picked individually from a data–bank and checked whether it lies in the abstract space $(\mathbf{E}_M)_a$. If a match is found, subject to a certain error tolerance, the search continues till S such species are found. Two scenarios can occur: one where S targets are accepted and the other where only $S_k < S$ targets can be accepted. In the first scenario, $\mathbf{T}$ can be obtained using the least–squares technique. The concentration can then be estimated using the relation $\hat{\mathbf{C}}_M = \mathbf{C}_a \mathbf{T}$, where $\hat{\mathbf{C}}_M$ is the estimated concentration.

On the other hand, if only $S_k < S$ pure–component spectra targets can be accepted, the first S_k rows of $\mathbf{T}^{-1}$ can be determined. The remaining $S - S_k$ rows of $\mathbf{T}^{-1}$ could be calculated using optimization under constraints such as positivity and smoothness of $\mathbf{E}_M$ and $\hat{\mathbf{C}}_M$. The requirements of smoothness are typically reflected in optimization criteria such as the entropy criterion (see Malinowski (1991) and references therein). However, this procedure is likely to give non–unique solutions for $\hat{\mathbf{C}}_M$ unless some physical characteristics hold. Often, in practical applications, there are selective MIR regions, i.e. regions where only one species absorbs (orthogonality property). Hence, the concentration profile of this species can be directly computed from $\mathbf{A}_M$ apart from a normalization constant. If, in addition, the concentration of this species at one time instant is known (e.g. the initial or end concentration), then the normalization constant can be determined uniquely.

5 NONLINEAR CALIBRATION OF NIR SPECTROSCOPIC DATA

Having estimated the concentrations from MIR absorbance spectra, the next step is to calibrate a nonlinear sensor model $\mathbf{f}_s$ (equation 5) which can be used to predict unknown concentrations as a function of on–line measured NIR absorbance spectra.
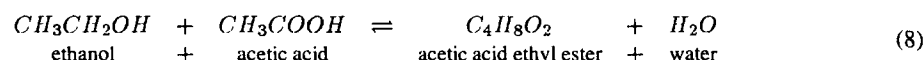
The first problem that one faces when performing nonlinear calibration with NIR spectroscopic data is that the input variables are numerous ($L_N > 500$) and highly correlated. This problem is handled here by the use of Principal Component Analysis (PCA) (Jackson 1991). In the $K \times L_N$ spectroscopic data matrix, $\mathbf{A}_N$, the L_N observed input variables can be well explained by n score variables with $n \ll L_N$. A linear combination of these n score variables is taken to model $\hat{\mathbf{C}}_M$ (Principal Component Regression — PCR). However, PCR is a linear regression method and, thus, cannot efficiently handle nonlinearities.

This problem can be overcome by combining PCA with nonlinear regression techniques. Only a few score variables are used as input variables to the nonlinear model. In the linear case, it should be sufficient to take as few score variables as there are species in the mixture. Yet, more score variables than the number of species are considered to account for the nonlinearity in equation (5). In this work, the nonlinear PCR technique is investigated following the lines of Löffler (1995). PCA is followed by a nonlinear regression step using 'Adaptive Spline Modeling of Observation Data' (ASMOD) (Kavli 1993). Since ASMOD can only handle one output at the time, a separate ASMOD model has to be built for each species concentration.

Another issue that is important in using such 'black–box' methods is the choice of the structure of the nonlinear model where the parsimony principle plays an important role (see Sjöberg et al. (1995) for a statistical interpretation). This choice is performed using a model validation step. In this work, leave–one–block–out cross–validation (CV) is used to determine the optimal structure of the calibration model. The model structure which provides the lowest prediction error is accepted. The structure parameters to be determined by CV in nonlinear PCR are (i) the number of input variables (n), (ii) the spline degree, and (iii) the number of knots or equivalently the number of degrees of freedom (DOF) in the model.

6 EXPERIMENTAL RESULTS

An esterification reaction with the following proposed stoichiometry is investigated in a batch reactor:

$$\begin{array}{ccccccc} CH_3CH_2OH & + & CH_3COOH & \rightleftharpoons & C_4H_8O_2 & + & H_2O \\ \text{ethanol} & + & \text{acetic acid} & & \text{acetic acid ethyl ester} & + & \text{water} \end{array} \tag{8}$$

The reaction is catalyzed by sulfuric acid and takes place in an excess of the solvent acetone ($10.5 \text{moles}\,\ell^{-1}$). The experiments are carried out in a 1.5ℓ vessel isothermically controlled at $50^\circ C$ in the laboratories of Lonza AG, Switzerland. For each batch run, absorbance spectra are measured in the MIR range between 500 and 4000cm^{-1} at a resolution of 1cm^{-1} and in the NIR range between 1100 and 2200 nm at a resolution of 2 nm. Each spectrum is an average of several scans (64 for MIR and 4 for NIR). The sampling interval for both MIR and NIR spectra is 3min. For verification
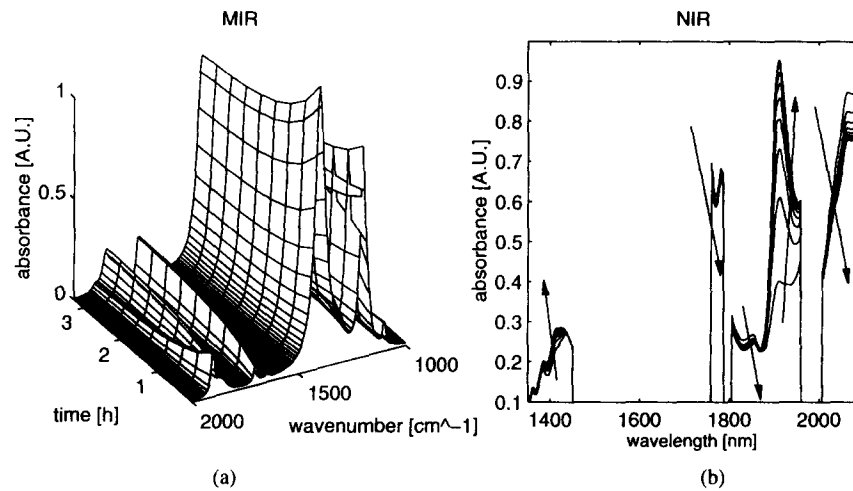
Figure 1: (a) MIR and (b) NIR spectral data for one batch run. The arrows indicate time evolution.

purposes, concentration of water at a few sampling instants is analyzed using Karl–Fischer titration. The concentrations of the other 3 reacting species are reconstructed from the water concentration using the stoichiometric model and the initial conditions.

For rank analysis, the reduced F–test (Malinowski 1991) with 95% confidence limit is chosen. Rank analysis of the MIR spectral data of a single batch run shows that rank augmentation techniques (Amrhein *et al.* 1995) have to be used before FA can be applied. It is seen that 5 data sets starting from linearly independent initial conditions are required to determine S. This confirms the fact that there are only 5 absorbing species, the 4 reacting ones and acetone. The catalyst concentration is so low that its contribution to the absorbance spectrum is negligible. Five batch runs starting from different initial concentrations are performed. Four of these data sets are used for estimation of the concentrations and nonlinear calibration (calibration set), and the remaining one for model validation (validation set).

The pretreatment procedures applied to the data sets consist of (i) selection of wavelength regions that exhibit significant change during the reaction and (ii) filtering of the spectral data matrices. For MIR data, the region $1000–2000\text{cm}^{-1}$ is selected. 2D-median filtering with a window size of $10\text{cm}^{-1} \times 9\text{min}$ is used to remove spikes. Analysis of $\mathbf{A}_M$ for each batch run exhibits a rank of 2 (see Amrhein *et al.* (1995) for details). Thus, for further filtering, PCA is applied to $\mathbf{A}_M$ retaining only the first 2 principal components (PCs). For NIR data, the regions 1350–1450, 1760–1788, 1806–1958, and 2008–2090nm are selected. A zero–phase digital filter with a cut–off frequency of 0.4rad s^{-1} is used along the time axis. From the spectral matrix of each batch run i, the initial spectrum, $\mathbf{a}_{N,0}^i$, is subtracted: $\mathbf{H}_N^i = \mathbf{A}_N^i - 1\,\left(\mathbf{a}_{N,0}^i\right)^T$ with 1 being a vector with elements of 1 and $\mathbf{H}_N^i$ the absorbance change matrix. Let $\mathbf{H}_N$ be the matrix obtained by appending the various $\mathbf{H}_N^i$. Using $\mathbf{H}_N$ for further analysis instead of $\mathbf{A}_N$ has the advantage that the problem of unknown baseline shifts from batch to batch is avoided. The MIR and NIR spectral data for one batch run after wavelength selection and filtering are shown in Fig. 1.

Due to experimental difficulties, the MIR data is reconstructed using a kinetic model and the infrequent water concentration measurements. Also, typical heteroscedastic noise (multiplicative Gaussian noise with standard deviation of 0.01 offset by 0.0025) is added to the MIR spectra. Target FA (TFA) is then applied to the MIR data corresponding to 4 batch runs with linearly–independent initial conditions. Various pure–component spectra targets are proposed among which the targets of ethanol, acetic acid, acetic acid ethyl ester, water, and acetone are accepted based on the Euclidean distance of the targets from $(\mathbf{E}_M)_a$. The rotation matrix, $\mathbf{T}$, is then obtained using the least–squares technique. The measured and estimated concentrations are shown in Fig. 2.

For each batch run, the estimated initial concentrations, $\hat{\mathbf{c}}_{M,0}^i$, are subtracted from the estimated concentrations, $\hat{\mathbf{C}}_M^i$:
$\hat{\mathbf{D}}_M^i = \hat{\mathbf{C}}_M^i - 1\,\left(\hat{\mathbf{c}}_{M,0}^i\right)^T$ with $\hat{\mathbf{D}}_M^i$ being the estimated concentration change matrix of Batch i. Let $\hat{\mathbf{D}}_M$ be the matrix composed of the various $\hat{\mathbf{D}}_M^i$. In the nonlinear regression step, $\mathbf{H}_N$ is then calibrated against $\hat{\mathbf{D}}_M$. The resulting calibration model allows one to predict the concentration change vector, $\hat{\mathbf{d}}_N$, from a newly measured NIR absorbance change spectrum, $\mathbf{h}_N$. Based on $\hat{\mathbf{d}}_N$ and the initial concentrations, $\mathbf{c}_0$, for the new batch run, the concentration vector, $\hat{\mathbf{c}}_N$, is calculated as $\hat{\mathbf{c}}_N = \hat{\mathbf{d}}_N + \mathbf{c}_0$.

The normalized root mean squared error NRMS is chosen as performance measure for both fit and prediction:

$$\text{NRMS} = \frac{\|\hat{\mathbf{d}}_M - \hat{\mathbf{d}}_N\|_2}{\|\hat{\mathbf{d}}_M\|_2}.$$

where $\| \bullet \|_2$ is the Euclidean norm. The NRMS is calculated for the entire calibration set (NRMS_{cal}), for cross vali-
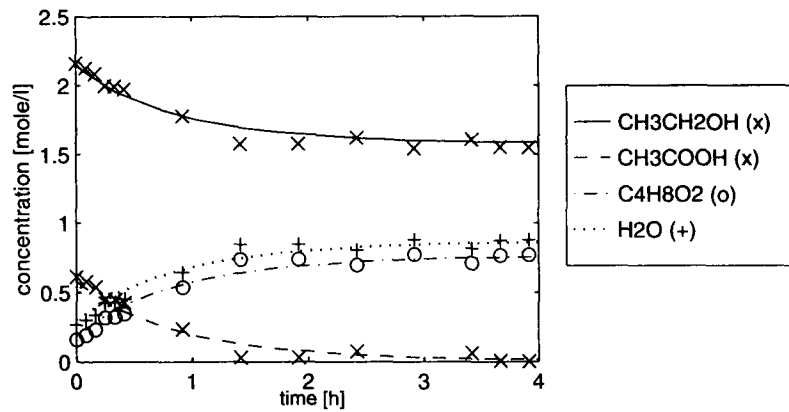
Figure 2: Concentration profiles, $\hat{C}_M$, estimated from MIR spectral data. Lines denote the estimated and marks the measured (H_2O) and reconstructed concentrations.
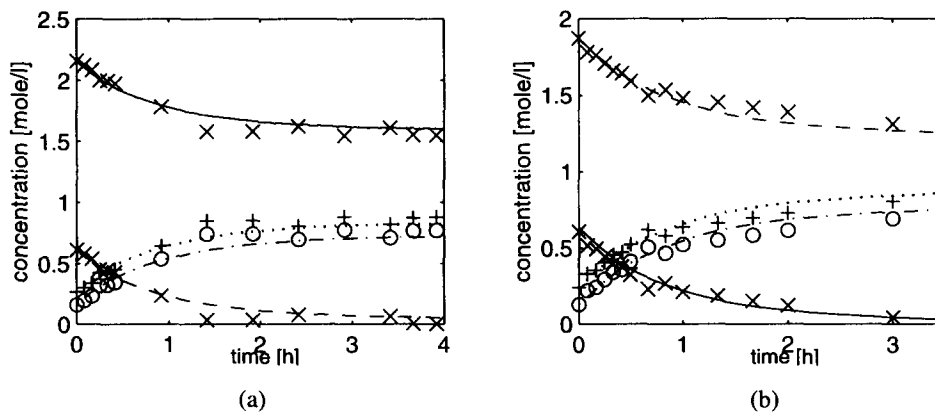

(a)   (b)

Figure 3: Predicted concentrations for (a) the NIR calibration set of one batch run and (b) the NIR validation set. Lines denote the predicted and marks the measured concentrations (see Fig. 2 for legend).

dation ($NRMS_{CV}$), and for the validation set ($NRMS_{val}$).

Based on leave–one–block–out CV with a block size of 3 spectra, the optimal structure that is chosen for the model of each species is 3 scores as inputs, three submodels with each submodel having one score as input, spline degree of 2, and DOF of 7 coefficients (knots). The values of $NRMS_{cal}$, $NRMS_{CV}$, and $NRMS_{val}$ for the four models are summarized in Table 1. The measured and predicted concentration profiles regarding both the calibration and the validation set are shown in Fig. 3.

Table 1: NRMS [%] for the entire calibration set ($NRMS_{cal}$), for cross validation ($NRMS_{CV}$), and for the validation set ($NRMS_{val}$).

NRMS	Species			
	CH_3CH_2OH	CH_3COOH	$C_4H_8O_2$	H_2O
cal	3.56	3.68	3.69	3.64
CV	3.64	3.75	3.77	3.72
val	3.99	6.06	6.71	5.62

7 CONCLUSIONS AND SIGNIFICANCE

An approach to infer concentrations in an industrial chemical reactor using on–line NIR spectroscopic measurements is presented. It uses a nonlinear calibration model based on the combined use of MIR/NIR spectroscopic measurements. This methodology can be easily extended to other types of spectroscopic measurements (Raman spectroscopy, UV/VIS, etc.) where one of the measurements is linear in the concentrations.

With such an approach, it is possible to estimate the concentration of all the absorbing species *on–line*. In isothermal cases, this often correspond to complete state information. This opens avenues hitherto unusable by the process control community. The concentrations estimated by the proposed method can be used to derive a stoichiometric model for the

reaction system by factor analysis (Bonvin and Rippin 1990) which can be used for on–line monitoring. In addition, since concentrations are estimated on–line, structure and parameters of reaction kinetics can be adapted in real time. Such models can also be used for adaptive control and/or on–line optimization.

The assumptions for the applicability of this approach is that the reaction is (i) conducted in an isothermal environment, and (ii) all reacting species are in the liquid phase. Research is under way to extend the methods beyond these assumptions.

NOTATION

Subscript:	a	= abstract quantity.
	M	= quantity corresponding to MIR spectroscopy.
	N	= quantity corresponding to NIR spectroscopy.
	0	= initial quantity.
	1	= dominant quantity.
Subscript:	T	= transpose.
	$+$	= pseudo–inverse.
Accent:	$\hat{\ }$	= estimate or prediction.

a, A	= absorbance vector, matrix.
c, C	= concentration vector, matrix.
d, D	= concentration change vector, matrix.
E	= matrix of pure–component spectra.
f_p	= physical functional relationship between c and a.
f_s	= spectroscopic sensor model: inverse of the physical functional relationship.
h, H	= absorbance change vector, matrix.
$k; K$	= time instant or observation; total number of observations.
L	= number of wavelengths, wavenumbers or integrated areas.
n	= number of scores (principal components) or loadings.
p	= number of input variables.
S	= number of absorbing species.
S_k, S_u	= number of known and unknown absorbing species.
Σ	= matrix of singular values.
T	= rotation matrix.
V, W	= left and right singular vectors.

ACKNOWLEDGMENTS

References

Kiers, H. A. and Smilde, A. K. (1995), 'Some theoretical results on second–order calibration methods for data with and without rank overlap', *Journal of Chemometrics* 9(3), 179–195.

Weiss, S. (1991), Faktorenanalyse, PhD thesis, University of Tübingen, Department of Chemistry and Pharmaceutics.

Barton II, F., Himmelsbach, D., Duckworth, J. and Smith, M. (1992), 'Two–dimensional vibrational spectroscopy: Correlation of mid– and near–infrared regions', *Applied Spectroscopy* 46(3), 420–429.

Windig, W., Margevich, D. and McKenna, W. (1995), 'A novel tool for two–dimensional correlation spectroscopy', *Chemometrics and Intelligent Laboratory Systems* 28, 109–128.

Lew, R. and Balke, S. (1993), 'Mid–infrared spectra from near–infrared spectra using partial least–squares', *Applied Spectroscopy* 47(11), 1747–1750.

Burns, D. A. and Ciurczak, E. W. (1992), *Handbook of Near–Infrared Analysis*, Marcel Dekker, Inc., New York.

Griffiths, P. R. and de Haseth, J. A. (1986), *Fourier Transform Infrared Spectrometry*, John Wiley & Sons, New York.

Malinowski, E. (1991), *Factor Analysis in Chemistry*, John Wiley & Sons, Chichester.

Amrhein, M., Srinivasan, B., Bonvin, D. and Schumacher, M. M. (1995), 'On the rank deficiency and rank augmentation of the spectral measurement matrix', *Chemometrics and Intelligent Laboratory Systems*. To appear.

Jackson, J. E. (1991), *A User's Guide to Principal Components*, John Wiley & Sons.

Wold, S. (1992), 'Nonlinear partial least squares modelling: II. spline inner relation', *Chemometrics and Intelligent Laboratory Systems* 14, 71–84.

Löffler, J. (1995), Nonlinear calibration methods for spectroscopic data: Comparison of adaptive spline modeling (AS-MOD) with radial basis function networks, Master's thesis, École Polytechnique Fédérale de Lausanne, Institut d'Automatique.

Kavli, T. (1993), 'ASMOD–an algorithm for adaptive spline modelling of observation data', *International Journal of Control* 58(4), 947–967.

Sjöberg, J., Zhang, Q., Ljung, L., Benveniste, A., Deylon, B., Glorennec, P.-Y., Hjalmarsson, H. k. and Juditsky, A. (1995), Nonlinear black–box modeling in system identification: A unified overview, Technical report, Division of Automatic Control, Linköping University, Sweden.

Bonvin, D. and Rippin, D. (1990), 'Target factor analysis for the identification of stoichiometric models', *Chemical Engineering Science* 45(12), 3417–3426.