Monitoring of Batch Processes Using Spectroscopy

Stephen P. Gurden, Johan A. Westerhuis, and Age K. Smilde
Process Analysis & Chemometrics, Dept. of Chemical Engineering, University of Amsterdam,
1018 WV Amsterdam, The Netherlands

There is an increasing need for new techniques for the understanding, monitoring and the control of batch processes. Spectroscopy is now becoming established as a means of obtaining real-time, high-quality chemical information at frequent time intervals and across a wide range of industrial applications. In this article, the role of spectroscopy for batch process monitoring is discussed in terms of both current and potential advances. The emphasis is on how to handle the measured data to extract maximum information for improved process performance and efficiency. In particular, the use of spectroscopy for statistical process monitoring is detailed and considered as complementary to the use of engineering process data. A case study of the ultraviolet-visible monitoring of a first-order biochemical conversion reaction is described, as well as the advantages of spectroscopy for process fault detection and diagnosis. Future prospects for the use of on-line spectroscopy are also discussed.

Introduction

Batch processes form a substantial part of the chemical and food industries, including the production of high-grade pharmaceuticals and polymers. There is an increasing need for improved methods for the understanding, monitoring, and control of these processes. New techniques for measuring the state of the process in real time are sought, particularly those that are easy to automate and implement *in situ*.

Spectroscopic analysis techniques are now receiving a great deal of attention in the process industries (Dahl, 2000; Hassell and Bowman, 1998). The increased robustness and reliability of modern spectrometers means that spectroscopy now provides a realistic alternative to off-line laboratory analysis as a source of chemical information. Slow-but-accurate chromatographic analyses are being replaced by fast-and-sufficiently accurate spectroscopic analyses, particularly in cases where removing the delay caused by analysis time allows for a higher overall production rate, for example, through improved batch scheduling.

A major advantage offered by spectroscopy is that sampling problems, usually the bottleneck in process analysis, are greatly reduced. Spectroscopy is nondestructive, because no physical sampling of the process is necessary and reagentless, eliminating the need for volatile solvents in the plant environment. The interface between the spectrometer and the process can take a number of forms. Sometimes, a distinction is made between “on-line” measurement, in which an automated sampling loop is used to transport the sample to the spectrometer, and “in-line” measurement, in which the sample interface sits directly in the process line/reactor. Advances in fiber-optic technology mean that in-line analysis is now becoming the norm, a typical example being a Fourier-transform infrared (FT-IR) spectrometer connected to an attenuated total reflection probe (ATR). A third scenario is that of noninvasive measurement whereby no direct contact with the sample is made, but the measurement is made through a reflectance window in the reactor wall. A more detailed discussion of the technical aspects of process sampling using spectroscopy can be found in the literature (McLennan and Kowalski, 1995; Sherman and Rhodes, 1996; Chalmers, 2000; Hassell and Bowman, 1998). Note that the term on-line is used in the following sections of this article as a general term to refer to any process measurement made in real time.

A major advantage of spectroscopy is the ability to obtain data that are rich in chemical information, in contrast to the standard process measurements of temperature, pressure, and

---

Correspondence concerning this article should be addressed to A. K. Smilde.

flow rate. Techniques such as near-infrared (NIR), FT-IR, ultraviolet-visible (UV-Vis), Raman, and fluorescence spectroscopy all provide information directly related to the molecular nature of the sample being measured. While this range of techniques makes spectroscopy applicable to a wide range of chemical processes, each of the techniques have their own characteristics. For example, NIR spectroscopy allows long path lengths (an advantage for in-line analysis) and can be used with fiber-optic cables to allow remote measurement. A disadvantage of NIR, however, is that it is not as chemically informative as FT-IR in terms of compound identification. Each technique has a similar set of strengths and weaknesses, and it is important to use both chemical and process knowledge to decide which is most appropriate to the process at hand.

The aim of this article is to discuss the role of spectroscopy for real-time process monitoring, to describe how spectroscopy fits into the framework of multivariate statistical process control of batch processes, and to propose new ways in which spectroscopy can be used to improve process performance and efficiency. In the following section, spectroscopic data are contrasted with the more typical on-line process measurements of temperature, pressure, and so on. In the third section, the general methodology on multivariate statistical process monitoring is described. In the fourth section, ways in which this methodology can be adapted to spectroscopic data are discussed. An example of on-line process monitoring using a UV-Vis spectrometer is given in the fifth section. In this example, a new type of model is used, in which process information such as known compound spectra and reaction kinetics are explicitly incorporated into an empirical model of the data. Finally, some ideas and future prospects for spectroscopic process monitoring are given in the final section.

There is currently substantial literature on the application of various spectroscopic techniques to a range of chemical processes, both batch and continuous. As a prelude to the main discussion, a selection of these are discussed below in order to demonstrate the diversity of spectroscopic monitoring applications. The reader can find a more complete summary in some of the review articles on the subject (Dallin, 1997; Dumoulin et al., 1996; Podkulski, 1997; Workman et al., 1999; Workman, 1999).

Calibration-based spectroscopy

One of the most common uses of spectroscopy is to obtain quantitative measures for process characteristics that are otherwise difficult to measure on-line, for example, water content, polymer particle size, or percentage conversion. This is usually a two-step procedure, the first being the training of a calibration model in an off-line mode, and the second being the use of the calibration model for making on-line predictions (DiFoggio, 2000; Miller, 1995). In some cases only selected regions of the spectrum rather than the entire wavelength range are used, the aim being purely to optimize the predictive accuracy of the calibration model for the one or more response properties of interest.

Fiber-optic UV-Vis spectroscopy has been used for the *in situ* analysis of pharmaceutical reactions (Gemperline et al., 1997; Quinn et al., 1999). Control charts based on estimated concentration profiles of the reacting species were used to detect disturbances in the batch reaction. Mid-infrared (MIR) and NIR spectroscopy have been used to predict water concentration during an esterification reaction (Amrhein et al., 1999). Calibration models were built using data from the reacting system in order to avoid the influence of variation from reactive intermediates. In the food industry, a calibration model has been developed for on-line measurement of crystalline sugar using fluorescence spectroscopy (Baunsgaard et al., 2000). In the field of biotechnology, fluorescence (Skibsted et al., 2001), dispersive Raman (Shaw et al., 1999), and NIR (Hall et al., 1996; Macaloney et al., 1996) have been used for on-line and at-line generation of key trend analysis variables during fermentation reactions. For a polyolefine polymerization process, FT-IR was used to monitor the principal monomer and comonomer in the feed stream, where use of gas chromatography was found to suffer from unacceptable analysis time (Kirsch et al., 1996). An evaluation of the on-line FT-IR analyzer for closed-loop control using simulated data demonstrated significant improvements in product consistency. Finally, short-wave near-infrared (SW-NIR) and laser-induced fluorescence (LIF) measurements were combined to improve on-line prediction of gasoline properties (Litani-Barzilai et al., 1997), where the ability to measure remotely using fiber-optics would enable field operation in harsh environments.

Endpoint determination of batch and blending processes

Another area of interest is the use of spectroscopy for endpoint monitoring. Many batch reactions have variable duration depending upon certain process conditions being attained. Again, spectroscopy provides a fast, real-time method of determining current process conditions in cases where chromatographic analysis would be too slow, especially considering sample handling and preparation time. One approach is to use a calibration model to determine when a property of interest reaches a desired value or to determine when an undesired side-reaction begins (Frake et al., 1997; Ward et al., 2000). Alternatively, analysis of the spectral variability across a local time window, either directly or in a reduced subspace, can be used to determine when a reaction has finished (Svensson et al., 2000).

In the pharmaceutical industry, homogenous powder blending of active ingredients and excipients is essential so as to ensure tablet-to-tablet reproducibility and content uniformity. On-line NIR has been used to determine when optimal blending has been reached (Cho et al., 1997; Hailey et al., 1996; Sekulic et al., 1996). In one application, an index of dissimilarity between the measured spectrum and an ideal blend spectrum, defined using high-performance liquid chromatography (HPLC) as a reference method, was used (de Maesschalck et al., 1998).

Process Monitoring Using Engineering vs. Spectroscopic Data

Analysis of data from chemical batch processes for the purposes of process understanding, monitoring, optimization, and control has typically focused on those process variables that are easy to measure on-line. These include temperature, pressure, flow rate, and % reagent added, and are normally recorded automatically at regular time intervals by the data-collection system. This type of data is referred to here as "engineering" process data, because it defines process conditions typically of interest to the process engineer. Engineering process data are characterized by:

1. Physical-state information: the process variables describe physical conditions relating, for example, to mass and energy balances within the process.
2. Heterogeneity: the process variables are measured in different units and some form of scaling is required prior to multivariate analysis.
3. Process noise: due to the complexity of most batch processes, the measured data contains a relatively high level of unsystematic variation. Empirical models of the data typically describe only 20–50% of the total variation from the mean trajectories.
4. Nonlinearity: the process variable trajectories are often highly nonlinear in time, although removing the mean trajectories taken across all batch runs is often used as a way of linearizing the data (Nomikos and MacGregor, 1994). However, the nonlinearity with respect to the relationships between the process variables is only partially removed by this approach.

A primary concern of batch process operation is the production of a consistently high-quality product. However, obtaining analytical information (such as performing a GC-MS run) about the state of the process midbatch is not usually viable, due to the difficulty of sampling the process and/or the time and cost involved. The standard approach is, therefore, to monitor those process variables that are easier to measure, such as temperature and pressure. If these physical-state conditions remain in control, it is inferred that the chemical process itself is in control and that a good-quality product will follow.

Spectroscopic techniques that can be implemented on-line, such as NIR, Raman, and UV-Vis, provide an important solution to the difficulty of obtaining on-line information about the chemical state of a process. Spectroscopy provides a more direct means of monitoring specific chemical properties of the process, such as concentration changes of reacting species; changes in chemical conditions, such as pH or catalyst activity; and the presence of impurities. Therefore, the need to infer the chemical state from the physical state is removed.

There are both fundamental and practical differences between the analysis of engineering and spectroscopic process data. In contrast to engineering variables, spectroscopic data are characterized by:

1. Chemical-state information: the measured spectra describe chemical conditions relating to the molecular nature and the absolute concentrations of the species present within the process.
2. Homogeneity: all wavelengths are measured in the same units (usually absorbance). This removes the need for data scaling prior to analysis, although some form of spectral preprocessing may be necessary.
3. Reduction in process noise: spectroscopic data usually contains a far higher amount of systematic variation than engineering data. This is because, despite the complexity of the actual process, not all of the unsystematic variation in the physical state influences the chemical state, and so is not measured. It should be noted, however, that some physical conditions such as temperature can affect the measured spectra.
4. Linearity: spectroscopic data consist of a linear addition of contributions from each absorbing species according to the Lambert-Beer law. This makes spectral data highly suitable for analysis using linear models.

These characteristics indicate that spectroscopy has an important role in process modeling and control within the process industries. While engineering process variables are able to characterize a reacting system to a certain extent, and are important for process safety and control, adding spectroscopic measurements could greatly improve the ability to fully characterize the process. An ideal scenario would be the combination of engineering and spectroscopic data into a complementary process monitoring strategy—this is discussed later in the subsection titled "Combining spectroscopic and engineering measurements."

Potential difficulties of using spectroscopy

One of the main challenges in the use of spectroscopy as an on-line industrial analyzer is that of model robustness, that is, obtaining accurate predictions and/or reliable control charts over a period of time. In comparison to engineering data, spectroscopic data are generally more sensitive to changes in process conditions such as variation in the background medium, seasonal variation in feedstock, or temperature fluctuations. Spectroscopic effects such as light scattering and artifacts of the spectrometer such as baseline or wavelength shift can also degrade quantitative accuracy. The use of spectroscopy to measure physical sample properties, such as octane number, is sometimes difficult, as these properties may be nonlinearly related to the chemical composition or may be affected by nonabsorbing trace compounds. In practice, these problems can often be avoided by selecting the spectroscopic techniques most suited to the chemical process under consideration.

The issue of model robustness is an ongoing topic of research in the literature (Despagne et al., 2000), along with methods for identifying sources of variation in measured spectra and evaluating their effect on quantitative predictions (Rutan et al., 1998). Wavelength selection has been commonly used as a means of focusing on regions of the spectrum immune to interferent variation (Swierenga et al., 2000). Mathematical preprocessing techniques, such as taking derivatives or orthogonal signal correction (Wold et al., 1998a), can also be used to remove an undesired variation. In cases where spectral variation is due to known causes, such as instrumental laser drift or temperature influence, this can be corrected for explicitly using direct standardization techniques (Wang and Kowalski, 1993; Wülfert et al., 2000). Alternatively, experimental design can be used to ensure that all known sources of external variation are present in the calibration training set (Swierenga et al., 1999), although this can sometimes lead to more complex models.

Multivariate Statistical Process Monitoring for Batch Processes

In the previous section we discussed spectroscopy generally as a way of more completely characterizing a chemical process. In this and the following two sections, we focus on spectroscopy as a tool for statistical process monitoring, that is, the use of on-line measurements to (1) detect if and when an abnormal process event occurs, and (2) diagnose the cause of the abnormal event.

The main feature of statistical process monitoring, going back to the univariate approach of Shewhart (1931), is the use of one or more control charts that describe the state of the process in real time. If the process moves significantly away from normal operating behavior, the process operator is alerted so as to allow corrective action to be taken. In recent years, attention has been focused on the application of subspace methods such as principal component analysis (PCA) and partial least squares (PLS) to engineering process data. It is well known that while it is possible to measure a large number of process variables on-line at frequent time intervals, thus acquiring a large amount of information, these variables are often highly correlated. Subspace (or latent variable) techniques provide a way of summarizing the data in terms of a few multivariate factors that describe the real, underlying causes driving the chemical process.

Nomikos and MacGregor (1994, 1995a), developed a strategy for the multivariate statistical process control (MSPC) of batch processes by extending an approach applied previously to continuous process data (Kresta et al., 1991; Wise et al., 1990). The MSPC approach is based on the collection of an historical set of process data "obtained when the process is operating well and is in a state of control" (Nomikos and MacGregor, 1994). This data consist only of "common-cause" variation, that is, process variation that occurs under normal operating conditions (NOC), and is used as a reference data set against which newly measured batches can be compared. It is assumed that the NOC batch runs are normally distributed around a theoretical "average" batch run. An empirical subspace model is built on the data and used as a statistical basis for the construction of multivariate control charts. New batches are then monitored and batch runs for which a process disturbance causes variation outside that expected for NOC data are flagged. Diagnostic tools can be used to identify the reason for this "special cause" variation.

MSPC of batch processes can be applied both off-line and on-line. Implemented off-line, previously measured batches are analyzed in order to check whether they ran according to normal operating behavior and to monitor the general operating efficiency of the batch process with a view to increased process understanding and optimization. Implemented on-line, a batch run is monitored in real time, and early warning of abnormal operation allows the operator to either adjust process conditions accordingly or, if necessary, abort the batch run prematurely.

It is worth noting that the term "control" in MSPC is used in the sense of process monitoring with the aim of rapid detection and diagnosis of process faults. Preventative measures can then be taken to ensure that these faults do not occur during future batch runs, but few examples of automatic feedback control have been published to date. For process control and optimization strategies, carefully designed data are usually necessary, rather than the historical NOC data used to build process monitoring models. Therefore, in this article the more general terms "statistical process monitoring" or just "process monitoring" are used.

A summary of the multivariate statistical process monitoring approach is now given in the following three subsections. For further information on the mathematics and statistics involved, the reader is referred to the literature (Boqué and Smilde, 1999; Kosanovich et al., 1996; Louwerse and Smilde, 2000; Neogi and Schlags, 1998; Nomikos and MacGregor, 1994; Nomikos and MacGregor, 1995a).

Modeling of the NOC data

Batch data are characterized by a three-way array of process measurements, $\boldsymbol{X}$. If J process variables are measured at K time points for each of I NOC batch runs, then an $I \times J \times K$ array of NOC data is produced. Typically, around 30–60 batches are used, which should, ideally, span the operating range of process and be normally distributed around the average batch. Empirical, latent-variable-based methods, such as PCA (Jackson, 1991; Jolliffe, 1986; Wold et al., 1987) and, more recently, parallel factor analysis (PARAFAC) (Bro, 1997; Dahl et al., 1999; Louwerse and Smilde, 2000), have been used to model the NOC process data. These methods are particularly suitable for data where the dimensions of the data arrays are much larger than the true dimensionality of the data themselves, due to a high correlation between variables. Methods such as cross-validation exist for estimating the true dimensionality of the data (Eastment and Krzanowski, 1982; Louwerse et al., 1999; Wold, 1977).

Prior to modeling of the NOC data, the mean of each column of $\boldsymbol{X}$ is subtracted. This has an important implication for the subsequent model of the data, because it implies that the aim is to model the *differences* of each batch run from a theoretical, average NOC batch run. Scaling of process variables is also normally performed, although for solely spectroscopic data, where all spectral intensities are measured in the same units, this is not usually necessary. The model of the data can be expressed in generic form by

$$\boldsymbol{X} = \boldsymbol{T}\boldsymbol{P}^T + \boldsymbol{E} \tag{1}$$

where $\boldsymbol{X}$ $(I \times JK)$ is the matricized, mean-centered NOC process data, as shown in Figure 1; $\boldsymbol{T}$ $(I \times R)$ is a reduced-dimension matrix (sometimes called a "scores" matrix) that describes the difference between the NOC batch runs; $\boldsymbol{P}$ $(JK \times R)$ is a reduced-dimension matrix (sometimes called a "loadings" matrix) that defines the model itself in terms of multivariate relationships between the measured process variables; and $\boldsymbol{E}$ $(I \times JK)$ is a matrix of model residuals. The number of model components is typically small (that is, $R = 2$ or 3). The nature of $\boldsymbol{P}$ determines the type of latent variable model imposed on the data. For a PCA model, a constraint of orthogonality is imposed on the columns of $\boldsymbol{P}$, that is, $\boldsymbol{P}^T\boldsymbol{P} = \boldsymbol{I}$. Other types of models are described in the next subsection.

Calculation of control statistics and limits

Once a model of the NOC data has been calculated, new batches can be monitored. For off-line (or postbatch) monitoring, a new completed set of batch measurements, $\boldsymbol{x}_{\text{new}}$ $(JK \times 1)$, is first mean-centered and scaled using the same coefficients as used for the NOC data. The new batch is then projected onto the model in order to test whether or not it corre-

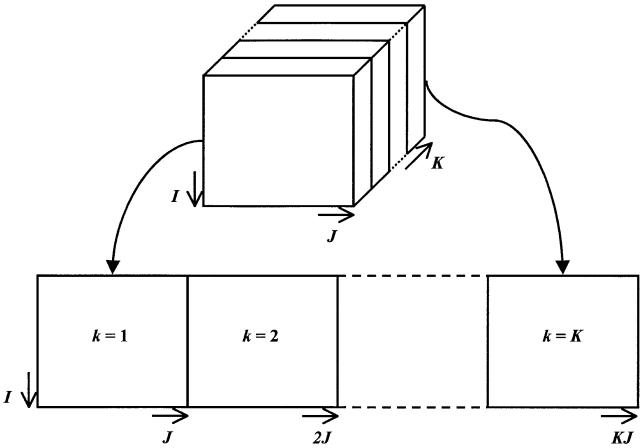


Figure 1. Matricizing (or "unfolding') used to transform the $I \times J \times K$ array of batch data into an $I \times JK$ matrix.

sponds to normal operating data. This projection yields a new set of scores and residuals as follows

$$t_{\text{new}} = (P^T P)^{-1} P^T x_{\text{new}} \tag{2}$$

$$e_{\text{new}} = x_{\text{new}} - P t_{\text{new}} \tag{3}$$

From the new scores and residuals, two informative statistics are calculated. The D-statistic is a Mahalanobis distance in the scores space given by

$$D = t_{\text{new}}^T S^{-1} t_{\text{new}} \tag{4}$$

where S is the covariance matrix of the NOC scores $1/(I-1)$, $T^T T$. The Q-statistic is simply the sum of squared residuals, given by

$$Q = \sum_{(jk)=1}^{JK} e_{(jk)}^2 \tag{5}$$

These two statistics describe different aspects of the new batch. The D-statistic defines variation *within* the model (that is, how different the projected new batch is from the average batch), and the Q-statistic defines variation *from* the model (that is, how different the correlation structure in the new batch is from that of the NOC data).

For both statistics, corresponding control limits (95% and 99%) are calculated using an F distribution for D (Tracy et al., 1992) and a fitted χ^2 distribution for Q (Jackson and Mudholkar, 1979) using the NOC data as a reference. If one or both statistics fall above the control limits, then the new batch is said to be "out of statistical control." Generally speaking, an unexpected process disturbance (such as a temperature controller failure) will be signaled by the Q-statistic, because the correlation structure defined by the model is broken. However, a change in process conditions for which the correlation structure is maintained (such as a significant, but safe, increase in the initial dosage level) may be signaled only by the D-statistic.

On-line monitoring involves one important difference from the off-line case described earlier. At time point k, only the process variables measured up to that point are available for projection onto the model, because the future state of the process from time $k + 1$ to K is not yet known. A number of methods have been proposed to handle this problem, the method used here being to fill in all future measurements with their current deviation from the average batch (Nomikos and MacGregor, 1995a). At each time point the D-statistic and the SPE, an "instantaneous" form of the Q-statistic using only the residuals at the current time point

$$SPE_k = \sum_{j=1}^{J} e_{jk}^2 \tag{6}$$

are calculated and, again, appropriate control limits are determined.

Fault diagnosis

If an out-of-control situation is signaled during an on-line or off-line analysis, diagnostic tools are available in order to determine which process variables are responsible. A commonly used method is that of contribution plots, which provide a measure of how much each process variable contributes to the overall level of the D- and Q-statistics or the SPE (Miller et al., 1998; Westerhuis et al., 2000a). For the Q-statistic and the SPE, this is tantamount to a plot of the residuals, but for the D-statistic, the contribution of each process variable, both to the overall model and, in some cases, to individual scores, is possible (Kourti and MacGregor, 1996; Nomikos, 1996; Westerhuis et al., 2000a).

Other approaches to MSPC

In the last five years, refinements and alternative approaches to that just described have been proposed. These include matricizing the X array in a different way to that shown in Figure 1 for a case where the batch runs were of varying duration (Wold et al., 1998b) [the problem of unequal batch lengths is also addressed elsewhere (Kassidas et al., 1998)]; an adaptive monitoring approach based upon hierarchical PCA (Ränner et al., 1998); and methods for improving MSPC model robustness through model updating (Gallagher et al., 1997).

How to Adapt Batch Process Monitoring to Work for Spectroscopic Data

Spectroscopic data can be considered as a particular form of process data for which the process variables are a series of spectral absorbances over a range of wavelengths. Some implications of applying the batch MSPC approach to this type of data are discussed in this section.

Use of different models

The choice of model to use for the X data has been the subject of a number of recent publications (Dahl et al., 1999;

Louwerse and Smilde, 2000; Westerhuis et al., 1999). The original approach was to matricize the three-way array of batch data (see Figure 1) and apply a two-way PCA decomposition, given here in element form

$$x_{ijk} = \sum_{r=1}^{R} t_{ir} p_{jkr} + e_{ijk} \tag{7}$$

For applications involving process engineering variables, the nonlinear nature of the process interactions does not suggest that the relationships between the process variables is consistent in time and, therefore, this approach, which treats each process variable measured at each time point as a separate variable, proved sufficient. However, for spectroscopic process measurements, which are highly structured and often exhibit strong linear characteristics (that is, the property of linear additivity prescribed by the Lambert-Beer law), trilinear models may be appropriate.

The PARAFAC model is a generalization of the bilinear PCA to data of higher orders (Bro, 1997; Carroll and Chang, 1970; Harshman, 1970) given by

$$x_{ijk} = \sum_{r=1}^{R} t_{ir} b_{jr} c_{kr} + e_{ijk} \tag{8}$$

where the expression of the model in element form avoids the need for complex tensor or matrix notation. $\boldsymbol{X}$ is decomposed into three reduced-dimension matrices, $\boldsymbol{T}$ $(I \times R)$, $\boldsymbol{B}$ $(J \times R)$, and $\boldsymbol{C}$ $(K \times R)$, that describe the variation in each mode (that is, dimension) of the three-way array. For batch process data, $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{C}$ describe variation in the batch, process variable, and time modes, respectively. Comparison of this model with the PCA model (see Eqs. 1 and 7) show that the loadings matrix $\boldsymbol{P}$ is now structured so as to describe variation in the process variable and time modes separately, using $\boldsymbol{B}$ and $\boldsymbol{C}$.

PARAFAC models generally use far fewer parameters and describe less variation in the data than corresponding PCA models. However, PARAFAC offers the advantages of increased structural simplicity and improved interpretative power. For PCA, $\boldsymbol{P}$ consists of R components describing JK different interactions between the process variable and time modes. These can be highly complex and difficult to interpret; consequently, loadings plots for batch data are rarely shown in the literature. For trilinear models, the interactions are structured so as to allow a separation of sources of variation into the process variable (spectral wavelength) and time directions. Plotting $\boldsymbol{B}$ and $\boldsymbol{C}$ gives interpretable output as the nature of the variation in the process, for example, which chemical species are present and how their concentrations vary. This can provide additional insight into the process, an example of which is shown in the subsection titled "Modeling of the UV-Vis batch data."

One potential problem with the PARAFAC model is that an equal number of components, R, are used to describe the variation in the three modes. However, it is sometimes seen that the batch mode contains less variation than the other two modes, particularly for NOC data where each batch is nominally the same. Another multilinear model is the Tucker3 model (Smilde, 1992; Tucker, 1966), for which a different number of components can be used to model each mode

$$x_{ijk} = \sum_{p=1}^{P} \sum_{q=1}^{Q} \sum_{r=1}^{R} g_{pqr} t_{ip} b_{jq} c_{kr} + e_{ijk} \tag{9}$$

where g_{pqr} is an element of $\boldsymbol{G}$ $(P \times Q \times R)$, the so-called core array; P, Q and R are the number of latent components used to describe the first, second, and third modes, respectively (that is, the number of columns in $\boldsymbol{T}$, $\boldsymbol{B}$, and $\boldsymbol{C}$). Thus, a (2,4,3) Tucker3 model means that $P = 2$, $Q = 4$, and $R = 3$. From Eq. 9 it can be seen that if a core element is zero ($g_{pqr} = 0$), this implies that the triad $t_{ip} b_{jq} c_{kr}$ does not contribute to the model. A comparison of Eqs. 8 and 9 shows that the PARAFAC model can be considered as a restricted form of the Tucker3 model for which only the superdiagonal elements of the core array (that is, $p = q = r$) are nonzero.

Incorporating prior process knowledge through model constraints

A recent trend in chemical data modeling is to maximize the interpretability and diagnostic power of the model by incorporating as much available process information into the model as possible, rather than relying on purely data-driven, empirical models such as PCA. This external process information may be about the chemical process itself, such as known first principles models or reaction mechanisms, or about the process recipe and control regime, such as setpoint trajectories or controller coupling. Even if process information is only partially available, it is possible to build "gray" or "hybrid" models that consist of a "white" part describing variation due to known sources and a "black" part that empirically models a systematic variation due to unknown sources (Gurden et al., 2001; Ramaker et al., 2002).

One advantage of using three-way models to model spectroscopic batch data as described previously is that the sources of variation in the data, $\boldsymbol{X}$, are isolated within each of the three modes. Thus, known sources of variation in the batch mode (such as differences in feedstock quality), in the wavelength mode (such as spectral absorbances), and in the time mode (such as changes in species concentration) can be explicitly included in the process model in the form of constraints upon some of the model components. Appropriate constraints to use depend upon the type and level of process information available. Thus, if some of the pure spectra of the species present in the reaction are known *a priori*, the loadings in the wavelength mode can be restricted to equal these spectra (a hard constraint). For species that are known to be present, but for which no pure spectrum is available, restrictions of nonnegativity and perhaps smoothness are appropriate (soft constraints). If the reaction kinetics are known, these can be incorporated as a functional constraint upon the time-mode loadings.

The main advantage of gray models is the improved interpretability of the model, since some model parameters now correspond to known features of the process. Furthermore, in the case of a process fault being detected by the D-statistic, it is sometimes possible to determine which part of the model detected the fault, leading to improved fault diagnostics.

Contribution plots

For the analysis of process engineering measurements, contribution plots (Miller et al., 1998; Westerhuis et al., 2000a) are able to give an indication of which variables are responsible for a break in the correlation structure defined by the model. This is also true when spectroscopic measurements are used, but in this case the plots are particularly informative because they can be interpreted in terms of spectral components relating to chemical species present within the reactor. For example, the introduction of a nonreacting, spectroscopically active impurity into the batch reactor would result in an increase in the SPE, and a subsequent examination of the contributions to the SPE may allow easy identification of the impurity. An example of this is given in the case study presented in the following section. Another situation would be where a disturbance in process conditions, such as a temperature or pH change, causes the relative concentrations of the reacting compounds to shift. In this case, the contribution plots would be expected to consist of difference spectra between the affected compounds.

Case Study: UV-Vis Monitoring of a First-Order Conversion Reaction

In order to illustrate some of the concepts already discussed, a case study of the monitoring of a chemical reaction using on-line UV-Vis spectroscopy is now presented. The data have been described previously in the literature (Westerhuis et al., 2000b). New results are presented here in which new types of process disturbances are investigated and a more advanced form of gray model is used, one which explicitly incorporates a kinetic model of the process.

A biochemical two-step conversion reaction (Bisby and Thomas, 1986; Chau and Mok, 1992) of 3-chlorophenylhydrazonopropane dinitrile (A) with 2-mercaptoethanol (B) to give 3-chlorophenylhydrazonocyanoacetamide (D) and the by-product ethylene sulphide (E), via an intermediate adduct (C), is given by

$$\mathrm{A}+\mathrm{B} \overset{k_1}{\rightarrow} \mathrm{C} \overset{k_2}{\rightarrow} \mathrm{D}+\mathrm{E} \tag{10}$$

where B is present in large excess (276:1) with respect to A, and, therefore, pseudo-first-order kinetics can be assumed. A Hewlett-Packard 8453UV-Vis spectrophotometer with diode array detection was used to monitor the reaction. From the time of reaction initiation, a spectrum with a wavelength range of 200–600 nm at a resolution of 1 nm was recorded every 10 s for a total run time of 45 min. The range 300–500 nm is analyzed here, for which only the reactant, A, intermediate, C, and product, D, are spectroscopically active. Spectra recorded from the start of the reaction up to time 2 min were discarded as they were not found to be reproducible (perhaps due to mixing effects). In order to reduce the size of the data set, every second spectrum was discarded, leaving a total of 130 spectra per batch run.

Thirty-five batch runs were carried out under normal operating conditions: temperature = 25°C; pH = 5.2; $[A]_0 = 54\times 10^{-6}\ \mathrm{mol\cdot L^{-1}}$. Three of these batches were retained for model validation, leaving 32 batches for model building. In addition to the NOC data, three extra experiments were available for which deliberate disturbances were introduced. In the first, a pH disturbance was induced by the addition of NaOH after around 20 min, leading to a rise in pH of approximately 0.1. In the second, the constant operating temperature over the entire batch run was lowered from 25 to 22°C. In the third, the presence of a spectroscopically active impurity (approximately equivalent to an impurity level of 15% by peak height) was imitated by the addition of cresol red at the start of the reaction.

Table 1. Batch Data Used for the Case Study

Batch No.	Type of Data	Comments
1–32	NOC training set	Normal operating conditions
33–35	NOC validation set	Normal operating conditions
36	Disturbed batch	pH increase of 0.1 at time = 20 min
37	Disturbed batch	Temperature level lowered to 22°C
38	Disturbed batch	Impurity present (cresol red)

The aim of the analysis was twofold. First, to further model the measured NOC data in order to increase understanding of the process. For this, the 32 × 201 × 130 (batches × wavelength × time points) array of NOC data was used. Second, to determine whether UV-Vis spectroscopy is a suitable technique for on-line fault detection and diagnosis for this particular process. For this, the 6 × 201 × 130 array of “new” data was used, consisting of three NOC batches and three faulty batches. The batch data are summarized in Table 1.

Modeling of the UV-Vis batch data

A gray model of the data was built that incorporates as much available process information as possible, as discussed previously. In the example here, three pieces of information about the process were available *a priori*:

1. The pure spectrum of the reactant was available. By allowing the reaction to proceed to full conversion, it was also possible to measure the pure spectrum of the product. The pure spectrum of the intermediate was not available, as this species could not be isolated.

2. The first-order reaction kinetics are known (although the actual values of the rate constants, k_1 and k_2, were unknown) and, assuming constant temperature and pH throughout the batches, describe the concentration profiles of the three compounds in time as follows

$$\frac{d[A]}{dt} = -k_1[A] \tag{11}$$

$$\frac{d[C]}{dt} = k_1[A] - k_2[C] \tag{12}$$

$$\frac{d[D]}{dt} = k_2[C] \tag{13}$$

3. The Lambert-Beer law of linear additivity is applicable at the low concentration levels present in the reaction (Cheng

and Young, 1986). Thus, the concentration of a particular compound is linearly proportional to its spectral intensity at a given wavelength and does not affect the spectral contributions from other compounds.

These three pieces of external internal information were incorporated into a model of the NOC batch data in the form of a gray model, made up of a white and a black part. The model is shown in Figure 2, the model parameters are given in Figure 3, and the squared model residuals summed up for each of the three modes are given in Figure 4. In the white part of the model, a (1,3,3) Tucker3 structure was used, with some of the parameters being constrained to equal the known pure species spectra and the known reaction kinetics. Two extra three-way unconstrained components were used for the black part.

The main features of the model are as follows:

- The first and third components in the wavelength mode (see Figure 3a) have been fixed to equal the known pure spectra of the reactant and product. The second component, which was left free, provides an estimate of the spectrum of the unstable intermediate. This spectrum corresponds well with independent estimates of the spectrum using alternative methodologies (Bijlsma et al., 1999).
- The first three components in the time mode have been restricted to fit the first-order kinetic model given by Eqs. 11 to 13. The estimated values of the rate constants, k_1 and k_2, were 0.2716 and 0.0256, respectively; these values also correspond to independent estimates.
- The variation in the first batch mode loading is thought to be mainly due to slight changes in the initial concentration of the reactant, resulting in linear changes in the concentration profiles in accordance with the first-order model.
- The white part of the model described approximately 97% of the total variation of the mean-centered batch data. Thus, the imposed kinetic model and spectra are shown to describe well the actual measured process data. However, if only a white model was used, the residuals exhibited clear structural information of a level significant in comparison with the spectral noise. Therefore, two extra black components were used to fully describe the data.
- The first black component (see Figure 3b) describes, in the time mode, variation at the beginning of the reaction and, in the wavelength mode, a difference spectrum between the reactant and intermediate spectra. This component is interpreted as describing batch-to-batch variation due to differences in k_1 that cause the concentration profiles for the reactant and intermediate to vary slightly from the overall average profiles.
- The second black component shows a slow drift in the time mode and a clear cyclic effect in the batch mode (the batches were measured sequentially). This component is thought to describe either a baseline effect of the spectrometer or a low absorbance signal from component B (which was originally thought to be transparent in the selected wavelength range).
- The residuals are highest at the beginning of the batch run. This is typical for unistage processes for which the beginning of the batch run usually exhibits a most unsystematic variation. In the wavelength mode, the large spike at around 365 nm is an artifact of the UV-Vis instrument used.

The gray model described 99.72% of the total variation in the mean-centered data. This high percentage of explained variance is typical for UV-Vis data and corresponds well with an independent estimate of the spectral noise.

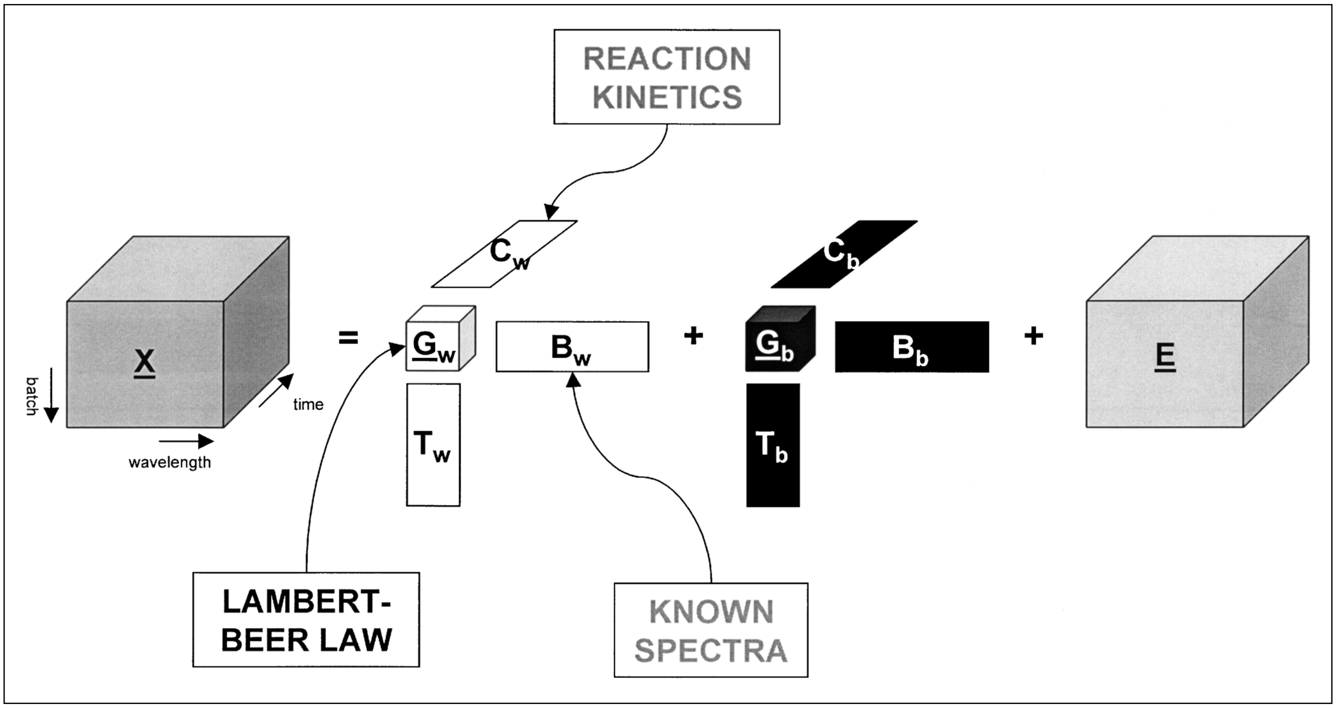


Figure 2. Gray model of the three-way batch data.

X is partitioned into a white part, which incorporates known pieces of external information, and a black part, which empirically models additional sources of variation.

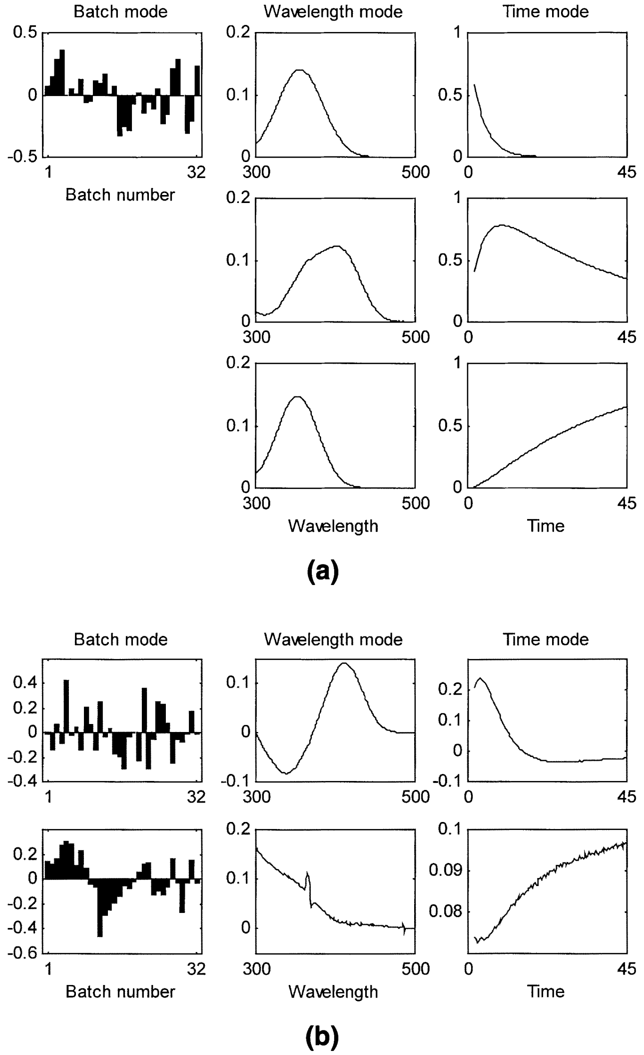


Figure 3. Loadings parameters for (a) white and (b) black parts of the gray model.

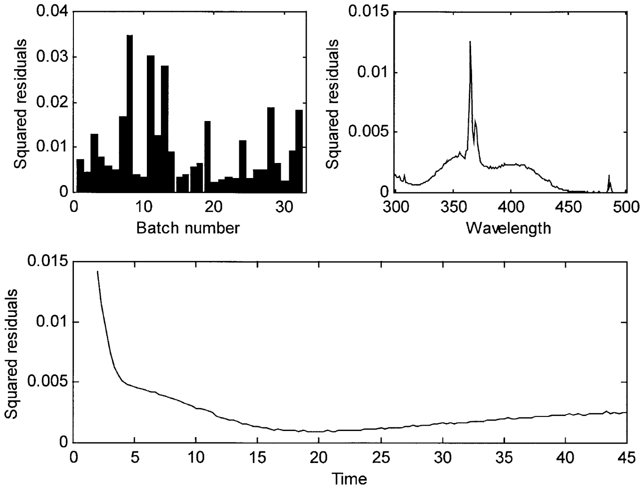


Figure 4. Sum-of-squared residuals for the gray model for the batch, wavelength, and time modes.

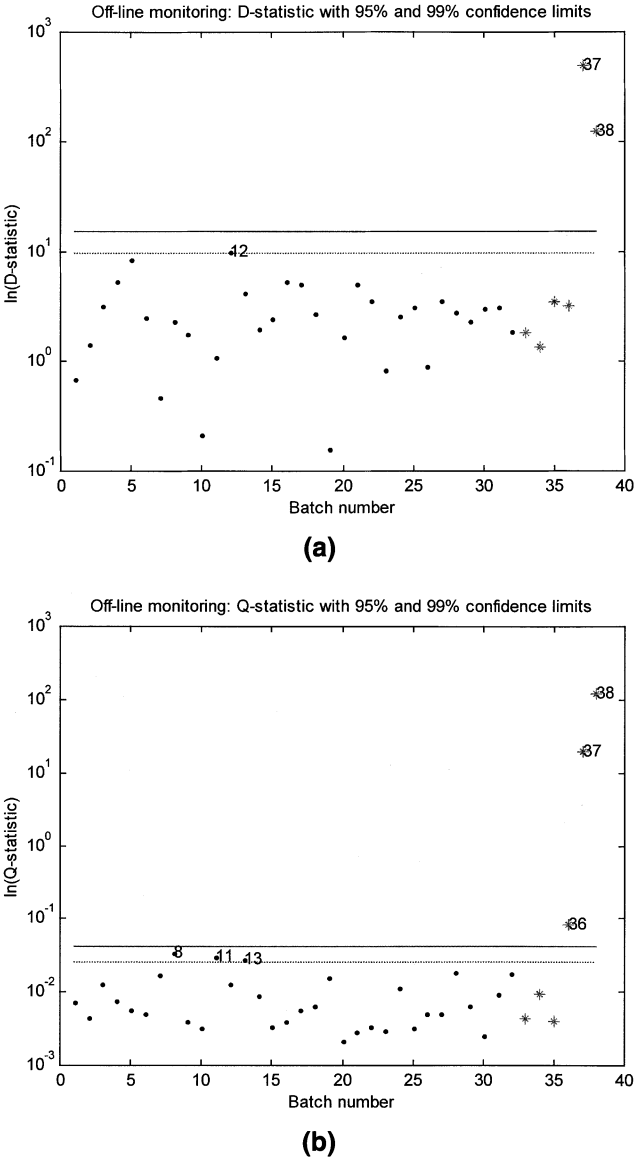


Figure 5. Off-line monitoring of (a) *D*-statistic, and (b) *Q*-statistic for the NOC (·) and "new," (∗) batches with 95% and 99% control limits.

Off-line monitoring of the UV-Vis batch data

Figure 5 shows the *D*- and *Q*-statistics for the off-line monitoring of the 32 NOC batches and the six new batches. The new NOC batches (33–35) all fall within the 95% control limits, indicating that these batch runs were in control. The three new faulty batches (36–38) all break the 99% control limits in the *Q*-statistic (batches 37 and 38 also break the *D*-statistic limits), indicating that for these batches, something abnormal occurred during the process run.

On-line monitoring of the UV-Vis batch data

The six new batches were monitored as if in an on-line situation. This can be considered either as an analysis to investigate why the batches were found to be abnormal during off-line monitoring, or as the mimicking of totally new data being measured on-line. The control charts for one of the new NOC batches (batch 33) and the three new faulty batches are shown in Figure 6.

Batch 33: Normal Batch. The process remains in control throughout the batch run, indicating that no unusual events occurred for this batch. Similar control charts were obtained for batches 34 and 35.

Batch 36: pH Disturbance. The process remains in control until around 23 min, at which point the SPE breaks the 95% control limits. The out-of-control behavior appears to begin at a point just after 20 min. Figure 7 shows the contribution plot for the SPE at 25 min. This appears as a difference spectrum between the intermediate and product spectra that indicates that the concentration of the intermediate is higher and that of the product is lower than for the NOC data. The most likely diagnosis is that something occurred at time 20 min to lower the rate constant for the second reaction, k_2. In this case, it is known that the reaction is pH dependent, and, thus, the pH disturbance at 20 min caused the abnormal situation.

Batch 37: Temperature-Level Change. The reaction is slightly out-of-control from the start of reaction monitoring and soon breaks both 99% control limits completely. A similar contribution plot to that for the previous batch, in this case, a difference spectrum between the reactant and intermediate spectra, showed that the cause was a decrease in the rate of the first reaction, k_1, in this case, caused by a lowering of the reaction temperature.

Batch 38: Presence of Impurity. The reaction is completely out-of-control from the start of reaction monitoring. Figure 8a shows the contribution plot for the SPE at 2 min, from which it appears that a new species is absorbing in the 400–500-nm range. A spectrum of cresol red is shown in Fig-

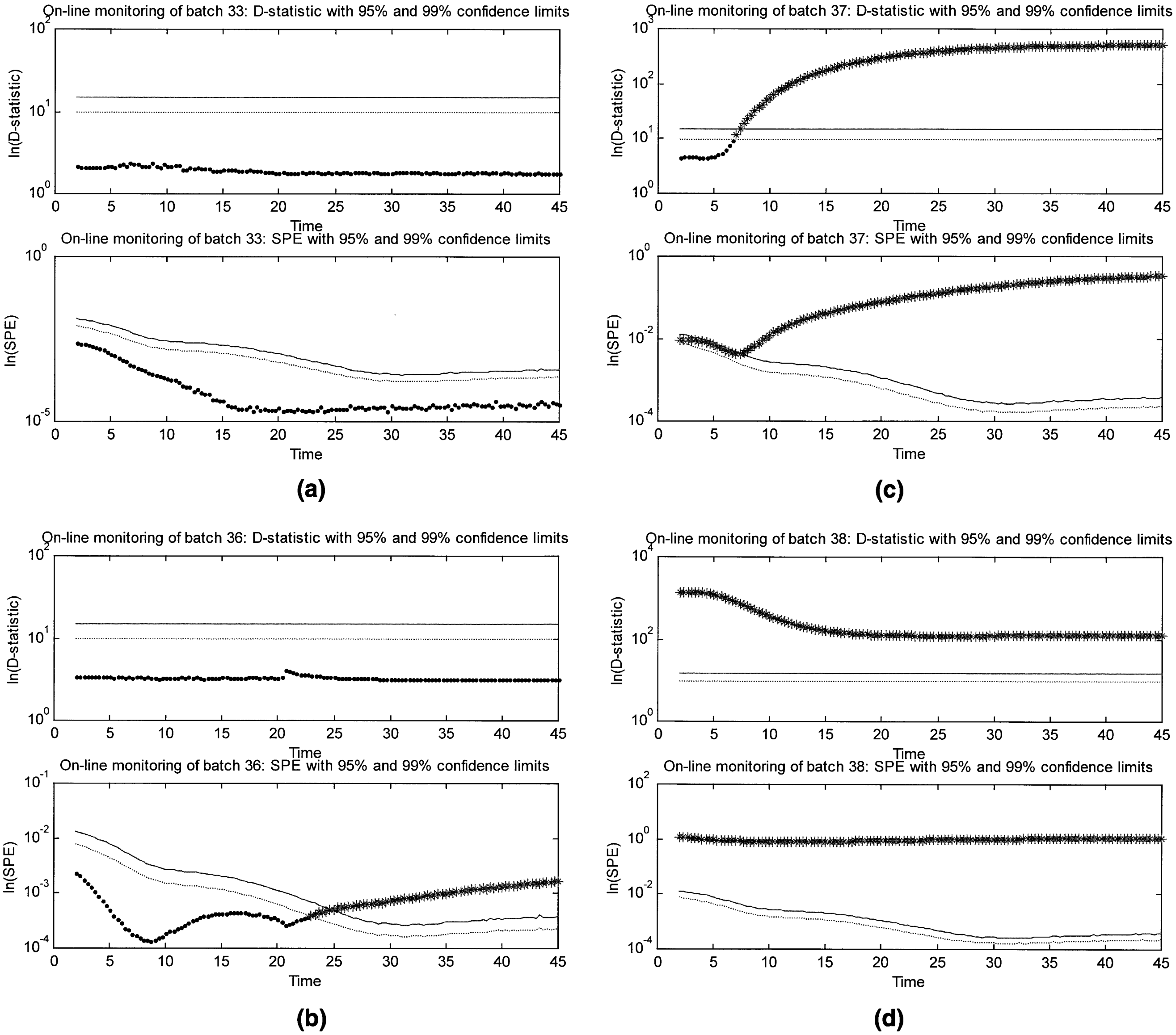


Figure 6. On-line *D*-statistic and SPE control charts with 95% and 99% control limits for (a) batch 33, (b) batch 36, (c) batch 37, and (d) batch 38.

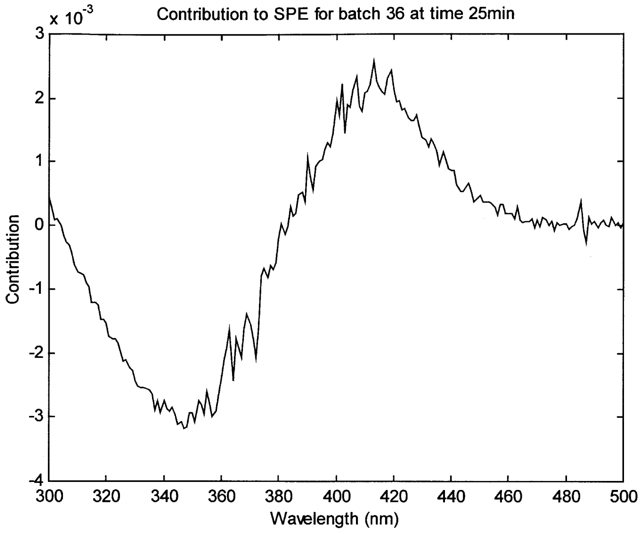
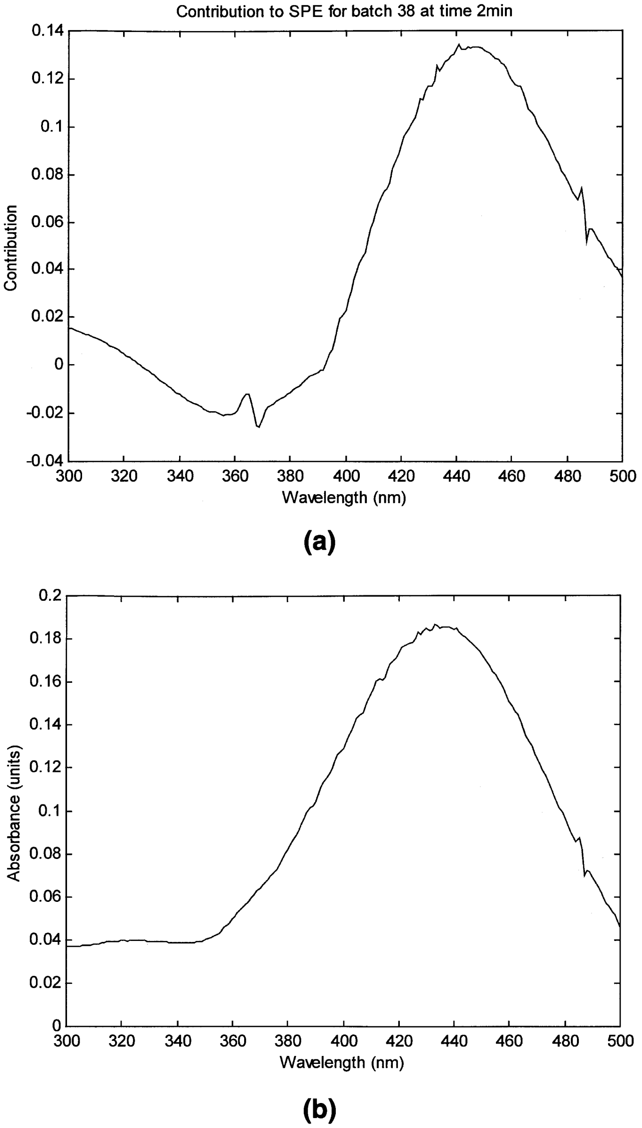


Figure 7. Contribution for the SPE of batch 36 at time 25 min.



(a)

(b)

Figure 8. (a) Contribution for the SPE of batch 38 at time 2 min; (b) spectrum of cresol red.

ure 8b, and the cause of the disturbance can be seen to correspond to this compound. In practice, the combination of this type of diagnostic and some chemical knowledge of the process could be useful in allowing the process operator to identify impurities in the feedstock or undesired side-products.

As an illustration of how spectroscopy could be used for feedstock monitoring, Figure 9 shows the spectrum of the reactant measured for each of the 32 NOC batches and for batch 38. A clear difference is seen in the spectrum of batch 38 due to the presence of the impurity. By using the NOC data to construct control limits that define an acceptable variation in the feedstock composition, it is easy to envision how spectroscopic monitoring could be used as an early check on normal batch operation.

Summary

In this section, data from the UV-Vis monitoring of a pseudo-first-order reaction was analyzed. The approach of gray modeling using restricted Tucker3 models has been described and was found to have substantial advantages over purely empirical models in terms of increasing process understanding. Using this methodology, it was possible to estimate pure component spectra of unstable intermediates and reaction rate constants. The use of a UV-Vis spectrometer for on-line batch process monitoring was found to provide a powerful method of detecting and diagnosing a variety of process faults.

Future Prospects

In this article, the use of spectroscopy as a source of high-quality chemical information for process analysis and on-line monitoring has been described. While the use of spectroscopy for improved statistical process monitoring has been discussed in detail, there are many other areas of process and quality control in which spectroscopy can play an important role. Some of these areas are discussed below.

Improved process understanding

One reason for monitoring a system spectroscopically may be to gain further understanding about the nature of the process. An advantage here is the ability to obtain frequent measurements relating to the concentration of the spectroscopically active species without disturbing the system itself. As already described in this article, the application of three-way models to spectroscopic batch data can be used to enable understanding of the different sources of batch-to-batch variation within a process. As another example, a rank analysis of spectra recorded over a current time window would indicate the number of (spectroscopically active) reacting species, enabling the detection and identification of nonstable intermediates.

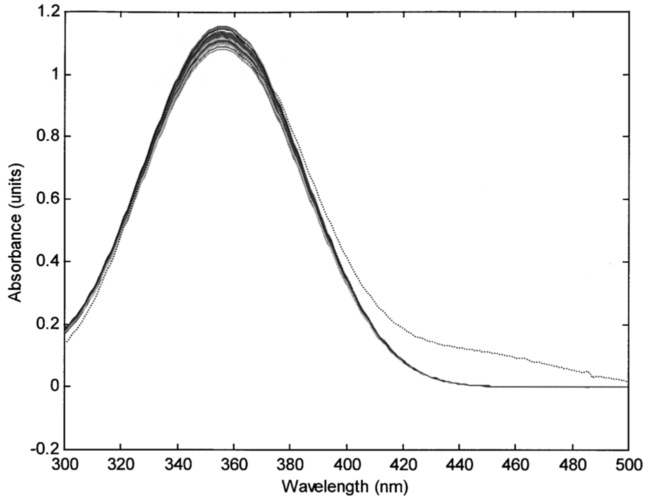


Figure 9. Spectra of the reactant prior to reaction initiation for the 32 NOC batches (—) and the batch with impurity present (····).

In the laboratory environment, SW-NIR (Bijlsma et al., 1998), UV-Vis (Bijlsma et al., 1999, 2001), FT-IR (Sans et al., 1997), and Raman (Svensson et al., 1999) spectroscopy have been used to study reaction kinetics in both first- and second-order systems. This approach could be extended to the more complex batch reactions common in the bioprocess and polymer industries. One possibility here is in the approach of tendency modeling (Rastogi et al., 1990, 1992), which aims to identify reaction schemes, stoichiometry, and kinetics based on a set of time-resolved concentration measurements. The ability to generate frequent concentration estimates without manual sampling of the process would provide a new opportunity in this field.

Monitoring of fundamental process parameters

The transfer of methods for estimation of fundamental process parameters such as reaction rate constants to an on-line scenario would provide a powerful insight into process changes at the chemical level. One goal here would be improved forecasting of where the process is going based on the current estimation of process conditions, thus enabling more accurate determination of endpoints and end-product quality. Prediction of future process conditions would also provide an improved solution to the problem of filling in future measurement during on-line monitoring (see the subsection titled "Calculation of control statistics and limits"). Another goal would be improved input information for closed-loop control, particularly for semibatch processes.

Specification limits and parametric release

If it is possible to build a reliable model between spectroscopic batch process data and end-product quality, it might be possible to transfer the specification limits on the end-product to the on-line spectroscopic measurements. This would then result in on-line control charts that give early warnings of eventual off-specification product. This would facilitate the quality control of the end-product, because this can then (partly) be done using the specification on the spectroscopic process measurements. This connects to the idea of parametric release, a quality-assurance system based on the idea that the product is of the intended quality if the information collected during the manufacturing process is compliant. Parametric release can be very important in, for example, biopharmaceutical processes where issues such as sterility mean that manual sampling is not an option. As a noninvasive source of high-quality chemical information, spectroscopy has an important role here.

Feedstock characterization

It is generally recognized within the process industries that the nature and quality of the raw feedstocks used to charge the process have an important effect on final product quality. However, strict analysis of feedstock quality is often not carried out due to the time and cost involved. Spectroscopy provides a method of instantaneous, multivariate characterization of the feedstock, allowing the process operator to (1) check whether the current feedstock is significantly different from previous feedstocks, (2) to measure quality characteristics that could then be used to set suitable process conditions for that feedstock, and (3) to give an early prediction of final product quality. An example of (1) was already seen in the case study.

Predicting end-product quality

Calibration-based spectral monitoring is able to provide a current estimate of process conditions within a batch reactor. However, sometimes an estimate of the final product quality would also be useful. Calibration models connecting on-line process measurements and final product quality have been used in the engineering literature to give an on-line prediction of final product quality for new batches (Boqué and Smilde, 1999; Nomikos and MacGregor, 1995b), but with limited success, perhaps due to the high level of unsystematic variation often found in industrial process data. As a source of high-quality chemical information, on-line spectroscopy could be expected to provide a large advantage in the role of predicting end-product quality.

Combining spectroscopic and engineering measurements

For industrial batch processes, spectroscopy can be viewed as a complement to the engineering trend data commonly measured at present, the combination of the two data types enabling a full description of the overall system in terms of both chemical and physical states. In cases where a calibration model is available, it is possible to make predictions at each time point and simply treat these predicted values as extra process variables. However, it is often advantageous to use the spectrum itself for monitoring in order to include all possible information about the process, including unexpected disturbances. Two approaches for this are as follows:

1. To consider the spectral data as an extra set of (highly correlated) process measurements and add them to existing data measurements, as shown in Figure 10a. It would be necessary to scale the two sets of data, according to their relative significance in describing the process.

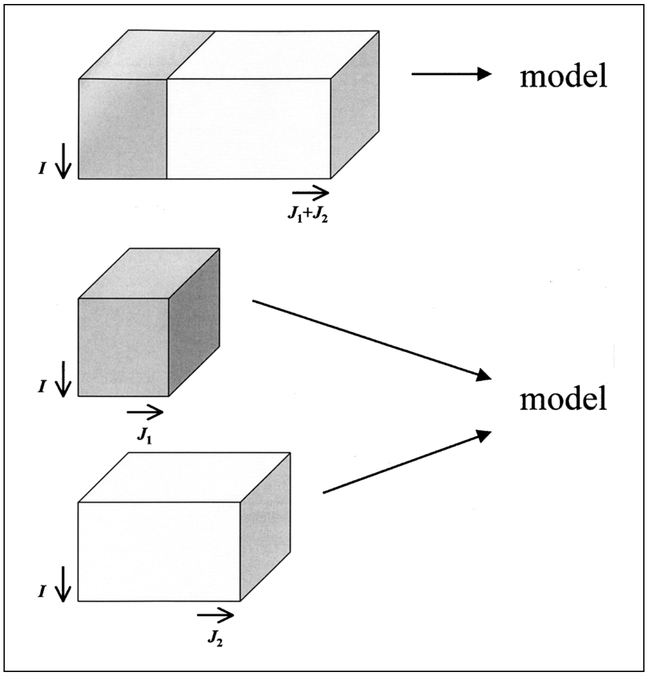


Figure 10. Combining different process data types.

(a) Engineering and spectroscopic data are appended to form one set of process data; (b) engineering and spectroscopic data are combined using a multiblock approach.

2. To treat the engineering and spectral data as separate pieces of information, but to combine them using a multiblock model (Frank et al., 1984; Kourti et al., 1995; Smilde et al., 2000; Wangen and Kowalski, 1988; Westerhuis et al., 1998; Wold et al., 1996), as shown in Figure 10b. The advantage of this approach is that each data type is modeled separately (and optimally), but the models are then combined to give, for example, one set of monitoring charts. This would also allow, for example, a nonlinear model to be used for the engineering data and a trilinear model to be used for the spectroscopic data.

Multiple instrumentation

Finally, another future prospect is the use of multiple instruments to give different views on the process. This could be envisioned in terms of combinations of spectroscopic techniques, such as NIR and fluorescence or, more generally, in terms of multivariate techniques already well established in other fields such as image analysis or acoustic spectroscopy. Techniques traditionally considered as off-line, such as HPLC and nuclear magnetic resonance (NMR), are now being adapted for on-line analysis. Coupled with the continual increase in computer data-handling capability and the widespread acceptance of multivariate statistical methodology, this points to an increasing role for on-line multivariate analyzers in the future.

Acknowledgments

These investigations were supported by the Council for Chemical Sciences of the Netherlands Organisation for Scientific Research (NWO-CW) with financial aid from the Netherlands Technology Foundation (STW). We also thank Sabina Bijlsma and Paula Dekker for performing the UV-Vis experiments.

Literature Cited

Amrhein, M., B. Srinivasan, D. Bonvin, and M. M. Schumacher, "Calibration of Spectral Reaction Data," *Chemometrics Intell. Lab. Syst.*, **46**, 249 (1999).

Baunsgaard, D., L Munck, and L. Nørgaard, "Evaluation of the Quality of Solid Sugar Samples by the Fluorescence Spectroscopy and Chemometrics," *Appl. Spectrosc.*, **54**, 438 (2000).

Bijlsma, S., H. F. M. Boelens, and A. K. Smilde, "Rapid Determination of Rate Constants in Second Order Kinetics," *Appl. Spectrosc.*, **55**, 77 (2001).

Bijlsma, S., D. J. Louwerse, and A. K. Smilde, "Estimating Rate Constants and Pure UV-vis Spectra of a Two-Step Reaction Using Trilinear Models," *J. Chemometrics*, **13**, 311 (1999).

Bijlsma, S., D. J. Louwerse, W. Windig, and A. K. Smilde, "Rapid Estimation of Rate Constants of Batch Processes Using On-Line SW-NIR," *AIChE J.*, **44**, 2713 (1998).

Bisby, R. H., and E. W. Thomas, "Kinetic Analysis by the Method of Nonlinear Least Squares. A Reaction Involving Consecutive Steps," *J. Chem. Educ.*, **63**, 990 (1986).

Boqué, R., and A. K. Smilde, "Monitoring and Diagnosing Batch Processes with Multiway Covariates Regression Models," *AIChE J.*, **45**, 1504 (1999).

Bro, R., "PARAFAC. Tutorial and Applications," *Chemometrics Intell. Lab. Syst.*, **38**, 149 (1997).

Carroll, J. D., and J. Chang, "Analysis of Individual Differences in Multidimensional Scaling via an N-Way Generalization of 'Eckart-Young' Decomposition," *Psychometrika*, **35**, 283 (1970).

Chalmers, J. M., *Spectroscopy in Process Analysis*, CRC Press, Sheffield (2000).

Chau, F. T., and K. W. Mok, "Multiwavelength Analysis for a First-Order Consecutive Reaction," *Comput. Chem.*, **16**, 239 (1992).

Cheng, K. L., and V. Y. Young, "Ultraviolet and Visible Absorption Spectroscopy," *Instrumental Analysis*, G. D. Christian and J. E. O'Reilly, eds., Prentice Hall, Englewood Cliffs, NJ (1986).

Cho, J.-H., P. J. Gemperline, P. K. Aldridge, and S. S. Sekulic, "Effective Mass Sampled by NIR Fiber-Optic Reflectance Probes in Blending Processes," *Anal. Chim. Acta*, **348**, 303 (1997).

Dahl, K. S., "Chemometrics' Experience in DuPont," *North Amer. Chap. Intl. Chemometrics Soc. Newslett.*, **20**, 7 (2000).

Dahl, K. S., M. J. Piovoso, and K. A. Kosanovich, "Translating Third-Order Data Analysis Methods to Chemical Batch Processes," *Chemometrics Intell. Lab. Syst.*, **46**, 161 (1999).

Dallin, P., "NIR Analysis in Polymer Reactions," *Process Control Qual.*, **9**, 167 (1997).

de Maesschalck, R., F. Cuesta Sánchez, D. L. Massart, P. Doherty, and P. Hailey, "On-Line Monitoring of Powder Blending With Near-Infrared Spectroscopy," *Appl. Spectrosc.*, **52**, 725 (1998).

Despagne, F., D. L. Massart, and P. Chabot, "Development of a Robust Calibration Model for Nonlinear In-Line Process Data," *Anal. Chem.*, **72**, 1657 (2000).

DiFoggio, R., "Guidelines for Applying Chemometrics to Spectra: Feasibility and Error Propagation," *Appl. Spectrosc.*, **54**, 94A (2000).

Dumoulin, M. M., R. Gendron, and K. C. Cole, "Techniques of Real Time Monitoring of Polymer Processing," *Trends Polym. Sci.*, **4**, 109 (1996).

Eastment, H. T., and W. J. Krzanowski, "Cross-Validatory Choice of the Number of Components from a Principal Component Analysis," *Technometrics*, **24**, 73 (1982).

Frake, P., D. Greenhalgh, S. M. Grierson, J. M. Hempenstall, and D. R. Rudd, "Process Control and End-Point Determination of a Fluid Bed Granulation by Application of Near Infra-Red Spectroscopy," *Int. J. Pharm.*, **151**, 75 (1997).

Frank, I. E., J. Feikema, N. Constantine, and B. R. Kowalski, "Prediction of Product Quality from Spectral Data Using the Partial Least-Squares Method," *J. Chem. Inf. and Comput. Sci.*, **24**, 20 (1984).

Gallagher, N. B., B. M. Wise, S. Watts Butler, D. D. White, Jr., and G. G. Barna, "Development and Benchmarking of Multivariate

Statistical Process Control Tools for a Semiconductor Etch Process: Improving Robustness Through Model Updating," *IFAC ADCHEM*, Banff, Canada, p. 78 (1997).

Gemperline, P. J., J.-H. Cho, B. Baker, B. Batchelor, and D. S. Walker, "Determination of Multicomponent Dissolution Profiles of Pharmaceutical Products by in Situ Fiber-Optic UV Measurements," *Anal. Chim. Acta*, **345**, 155 (1997).

Gurden, S. P., J. A. Westerhuis, S. Bijlsma, and A. K. Smilde, "Modelling of Spectroscopic Batch Process Data Using Grey Models to Incorporate External Information," *J. of Chemometrics*, **15**, 101 (2001).

Hailey, P. A., P. Doherty, P. Tapsell, T. Oliver, and P. K. Aldridge, "Automated Systems for the On-Line Monitoring of Powder Blending Processes Using Near-Infrared Spectroscopy: Part I. System Development and Control," *J. Pharm. Biomed. Anal.*, **14**, 551 (1996).

Hall, J. W., B. McNeil, M. J. Rollins, I. Draper, B. G. Thompson, and G. Macaloney, "Near-Infrared Spectroscopic Determination of Acetate, Ammonium, Biomass and Glycerol in an Industrial *Escherichia coli* Fermentation," *Appl. Spectrosc.*, **50**, 102 (1996).

Harshman, R. A., "Determination and Proof of Minimum Uniqueness Conditions for PARAFAC1," *UCLA Working Papers in Phonetics*, **22**, 111 (1970).

Hassell, D. C., and E. M. Bowman, "Process Analytical Chemistry for Spectroscopists," *Appl. Spectrosc.*, **52**, 18A (1998).

Jackson, J. E., *A User's Guide to Principal Components*, Wiley, New York (1991).

Jackson, J. E., and G. S. Muldholkar, "Control Procedures for Residuals Associated with Principal Component Analysis," *Technometrics*, **21**, 341 (1979).

Jolliffe, I. T., *Principal Component Analysis*, Springer Verlag, Berlin (1986).

Kassidas, A., J. F. MacGregor, and P. A. Taylor, "Synchronization of Batch Trajectories Using Dynamic Time Warping," *Process Syst. Eng.*, **44**, 864 (1998).

Kirsch, B. A., J. P. Chauvel, D. P. Denton, S. V. Lange, D. R. Lafevor, R. A. Bredeweg, W. K. Winnett, and J. L. Dinaro, "Improved Process Control in Polyolefin Polymerization by On-Line Fourier Transform Infrared Spectroscopy," *Process Control Qual.*, **8**, 75 (1996).

Kosanovich, K. A., K. S. Dahl, and M. J. Piovoso, "Improved Process Understanding Using Multiway Principal Component Analysis," *Industrial Eng. Chem. Res.*, **35**, 138 (1996).

Kourti, T., and J. F. MacGregor, "Multivariate SPC Methods for Process and Product Monitoring," *J. Qual. Technol.*, **28**, 409 (1996).

Kourti, T., P. Nomikos, and J. F. MacGregor, "Analysis, Monitoring and Fault Diagnosis of Batch Processes Using Multiblock and Multiway PLS," *J. Process Control*, **5**, 277 (1995).

Kresta, J. V., J. F. MacGregor, and T. E. Marlin, "Multivariate Statistical Monitoring of Process Operating Performance," *Can. J. of Chem. Eng.*, **69**, 35 (1991).

Litani-Barzilai, I., I. Sela, V. Bulatov, I. Zilberman, and I. Schechter, "On-Line Remote Prediction of Gasoline Properties by Combined Optical Methods," *Anal. Chim. Acta*, **339**, 193 (1997).

Louwerse, D. J., and A. K. Smilde, "Multivariate Statistical Process Control of Batch Processes Using Three-Way Models," *Chem. Eng. Sci.*, **50**, 1225 (2000).

Louwerse, D. J., A. K. Smilde, and H. A. L. Kiers, "Cross-Validation of Multiway Component Models," *J. Chemometrics*, **13**, 491 (1999).

Macaloney, G., I. Draper, J. Preston, K. B. Anderson, M. J. Rollins, B. G. Thompson, J. W. Hall, and B. McNeil, "At-Line Control and Fault Analysis in an Industrial High Cell Density *Escherichia coli* Fermentation, Using NIR," *Food and Bioprod. Process.*, **74**, 212 (1996).

McLennan, F., and B. Kowalski, eds., *Process Analytical Chemistry*, Blackie, London (1995).

Miller, C. E., "The Use of Chemometric Techniques in Process Analytical Method Development and Operation," *Chemometrics Intell. Lab. Syst.*, **30**, 11 (1995).

Miller, P., R. E. Swanson, and Ch. E. Heckler, "Contribution Plots: A Missing Link in Multivariate Quality Control," *Appl. Math. Comput. Sci.*, **8**, 775 (1998).

Neogi, D., and C. E. Schlags, "Multivariate Statistical Analysis of an Emulsion Batch Process," *Industrial Eng. Chem. Res.*, **37**, 3971 (1998).

Nomikos, P., "Detection and Diagnosis of Abnormal Batch Operations Based on Multiway Principal Component Analysis," *ISA Trans.*, **35**, 259 (1996).

Nomikos, P., and J. F. MacGregor, "Monitoring Batch Processes Using Multiway Principal Component Analysis," *AIChE J.*, **40**, 1361 (1994).

Nomikos, P., and J. F. MacGregor, "Multivariate SPC Charts for Monitoring Batch Processes," *Technometrics*, **37**, 41 (1995a).

Nomikos, P., and J. F. MacGregor, "Multi-Way Partial Least Squares in Monitoring Batch Processes," *Chemometrics Intell. Lab. Syst.*, **30**, 97 (1995b).

Podkulski, D. E., "How Do New Process Analyzers Measure Up?," *Chem. Eng. Prog.*, **93**, 33 (1997).

Quinn, A. C., P. Gemperline, B. Baker, M. Zhu, and D. S. Walker, "Fiber-Optic UV/Visible Composition Monitoring for Process Control of Batch Reactions," *Chemometrics Intell. Lab. Syst.*, **45**, 199 (1999).

Ramaker, H. J., E. N. M. van Sprang, S. P. Gurden, J. A. Westerhuis, and A. K. Smilde, "Improved Monitoring of Batch Processes by Incorporating External Information," *J. of Process Control*, **12**, 569 (2002).

Rastogi, A., J. Fotopoulos, C. Georgakis, and H. G. Stenger, Jr., "The Identification of Kinetic Expressions and the Evolutionary Optimization of Speciality Chemical Batch Reactors Using Tendency Models," *Chem. Eng. Sci.*, **47**, 2487 (1992).

Rastogi, A., A. Vega, C. Georgakis, and H. G. Stenger, Jr., "Optimization of Catalyzed Epoxidation of Unsaturated Fatty Acids by Using Tendency Models," *Chem. Eng. Sci.*, **45**, 2067 (1990).

Ränner, S., J. F. MacGregor, and S. Wold, "Adaptive Batch Monitoring Using Hierarchical PCA," *Chemometrics Intell. Lab. Syst.*, **41**, 73 (1998).

Rutan, S. C., O. E. de Noord, and R. R. Andréa, "Characterisation of the Sources of Variation Affecting Near-Infrared Spectroscopy Using Chemometrics Methods," *Anal. Chem.*, **70**, 3198 (1998).

Sans, D., R. Nomen, and J. Sempere, "Interactive Self-Modelling of Chemical Reaction Systems Using Multivariate Data Analysis," *Comp. Eng.*, **21**, S631 (1997).

Sekulic, S. S., H. W. Ward II, D. R. Brannegan, E. D. Stanley, C. L. Evans, S. T. Sciavolino, P. A. Hailey, and P. K. Aldridge, "On-Line Monitoring of Powder Blend Homogeneity by Near-Infrared Spectroscopy," *Anal. Chem.*, **68**, 509 (1996).

Shaw, A. D., N. Kaderbhai, A. Jones, A. M. Woodward, R. Goodacre, J. J. Rowland, and D. B. Kell, "Noninvasive, On-Line Monitoring of the Biotransformation by Yeast of Glucose to Ethanol Using Dispersive Raman Spectroscopy and Chemometrics," *Appl. Spectrosc.*, **53**, 1419 (1999).

Sherman, R. E., and L. J. Rhodes, eds., *Analytical Instrumentation*, Instrument Society of America, Research Triangle Park, NC (1996).

Shewhart W. A., *Economic Control of Quality of Manufactured Product*, Van Nostrand, Princeton, NJ (1931).

Skibsted, E., C. Lindemann, C. Roca, and L. Olsson, "On-Line Bioprocess Monitoring with a Multi-Wavelength Fluorescence Sensor Using Multivariate Calibration," *J. Biotechnol.*, **47** (2001).

Smilde, A. K., "3-Way Analyses—Problems and Prospects," *Chemometrics Intell. Lab. Syst.*, **15**, 143 (1992).

Smilde, A. K., J. A. Westerhuis, and R. Boqué, "Multiway Multiblock Component and Covariates Regression Models," *J. Chemometrics*, **14**, 301 (2000).

Svensson, O., M. Josefson, and F. W. Langkilde, "Reaction Monitoring Using Raman Spectroscopy and Chemometrics," *Chemometrics Intell. Lab. Syst.*, **49**, 49 (1999).

Svensson, O., M. Josefson, and F. W. Langkilde, "The Synthesis of Metoprolol Monitored Using Raman Spectroscopy and Chemometrics," *Eur. J. of Pharm. Sci.*, **11**, 141 (2000).

Swierenga, H., A. P. de Weijer, R. J. van Wijk, and L. M. C. Buydens, "Strategy for Constructing Robust Multivariate Calibration Models," *Chemometrics Intell. Lab. Syst.*, **49**, 1 (1999).

Swierenga, H., 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," *Anal. Chim. Acta*, **411**, 121 (2000).

Tracy, N. D., J. C. Young, and R. L. Mason, "Multivariate Control Charts for Individual Observations," *J. Qual. Technol.*, **24**, 88 (1992).

Tucker, L. R., "Some Mathematical Notes on Three-Mode Factor Analysis," *Psychometrika*, **31**, 279 (1966).

Wang, Y., and B. R. Kowalski, "Calibration Transfer and Measurement Stability of Near-Infrared Spectrometers," *Appl. Spectrosc.*, **46**, 764 (1993).

Wangen, L. E., and B. R. Kowalski, "A Multiblock Partial Least Squares Algorithm for Investigating Complex Chemical Systems," *J. Chemometrics*, **3**, 3 (1988).

Ward, H. W., II, S. S. Sekulic, M. J. Wheeler, G. Taber, F. J. Urbanski, F. E. Sistare, T. Norris, and P. K. Aldridge, "On-Line Determination of Reaction Completion in a Closed-Loop Hydrogenator Using NIR Spectroscopy," *Appl. Spectrosc.*, **52**, 17 (2000).

Westerhuis, J. A., S. P. Gurden, and A. K. Smilde, "Generalized Contribution Plots in Multivariate Statistical Process Monitoring," *Chemometrics Intelligent Lab. Syst.*, **51**, 95 (2000a).

Westerhuis, J. A., S. P. Gurden, and A. K. Smilde, "Spectroscopic Monitoring of Batch Monitoring for On-Line Fault Detection and Diagnosis," *Anal. Chem.*, **72**, 5322 (2000b).

Westerhuis, J. A., T. Kourti, and J. F. MacGregor, "Analysis of Multiblock and Hierarchical PCA and PLS Models," *J. Chemometrics*, **12**, 301 (1998).

Westerhuis, J. A., T. Kourti, and J. F. MacGregor, "Comparing Alternative Approaches for Multivariate Statistical Analysis of Batch Process Data," *J. Chemometrics*, **13**, 397 (1999).

Wise, B. M., N. L. Ricker, D. E. Veltcamp, and B. R. Kowalski, "A Theoretical Basis for the Use of Principal Component Models for Monitoring Multivariate Processes," *Process Control Qual.*, **1**, 41 (1990).

Wold, S., "Cross-Validatory Estimation of the Number of Components in Factor and Principal Components Models," *Technometrics*, **20**, 397 (1977).

Wold, S., K. H. Esbensen, and P. Geladi, "Principal Component Analysis," *Chemometrics Intell. Lab. Syst.*, **2**, 37 (1987).

Wold, S., H. Antti, F. Lindgren, and J. Öhman, "Orthogonal Signal Correction of Near-Infrared Spectra," *Chemometrics Intell. Lab. Syst.*, **44**, 175 (1998a).

Wold, S., N. Kettaneh, H. Friden, and A. Holmberg, "Modelling and Diagnostics of Batch Processes and Analogous Kinetic Experiments," *Chemometrics Intell. Lab. Syst.*, **44**, 331 (1998b).

Wold, S., N. Kettaneh, and K. Tjessem, "Hierarchicial Multiblock PLS and PC Models for Easier Model Interpretation and as an Alternative to Variable Selection," *J. Chemometrics*, **10**, 463 (1996).

Workman, J., Jr., D. J. Veltkamp, S. Deherty, B. B. Anderson, K. E. Creasy, M. Koch, J. F. Tatera, A. L. Robinson, L. Bond, L. W. Burgess, G. N. Bokerman, A. H. Ullman, G. P. Darsey, F. Mozayeni, J. A. Bamberger, and M. S. Greenwood, "Process Analytical Chemistry," *Anal. Chem.*, **71**, 121R (1999).

Workman, J. J., Jr., "Review of Process and Non-Invasive Near-Infrared and Infrared Spectroscopy," *Appl. Spectrosc. Rev.*, **34**, 1 (1999).

Wülfert, F., W. Th. Kok, O. E. de Noord, and A. K. Smilde, "Correction of Temperature-Induced Spectral Variation by Continuous Piecewise Direct Standardization," *Anal. Chem.*, **72**, 1639 (2000).

Manuscript received May 29, 2001, and revision received Mar. 8, 2002.