Process Monitoring of an Industrial Fed-Batch Fermentation

B. Lennox,[1] G.A. Montague,[3] H.G. Hiden,[2] G. Kornfeld,[4] P.R. Goulding[1]

[1]*Control Technology Centre, School of Engineering, University of Manchester, School of Engineering, Simon Building, Oxford Road, Manchester, M13 9PL, England; telephone: 0161 275 4324; fax: 0161 275 4346; e-mail: Barry.Lennox@man.ac.uk*
[2]*GSE Systems Ltd, Hexham, Northumberland*
[3]*Department of Chemical and Process Engineering, University of Newcastle, England*
[4]*Biochemie Ltd, Kundl, Austria*

Abstract: Market demand places great emphasis in industry on product quality. Consequently, process monitoring and control have become important aspects of systems engineering. In this article we detail the results of a 2-year study focusing on the development of a condition monitoring system for a fed-batch fermentation system operated by Biochemie Gmbh in Austria. We also demonstrate the suitability and limitations of current state of the art technologies in this field and suggest novel modifications and configurations to improve their suitability for application to a fed-batch fermentation system. © 2001 John Wiley & Sons, Inc. *Biotechnol Bioeng* **74**: 125–135, 2001.
Keywords: fault detection; fault diagnosis; fermentation; multivariate statistical process control; principal component analysis

INTRODUCTION

When operating fermentation systems it is vitally important to maintain operation within strict limits. There are two main reasons why this is the case. First, biologically based systems by nature are highly sensitive to abnormal changes in operating conditions. To ensure that the maximum possible yield of product is obtained from the system it is necessary to make sure that conditions within the fermentor remain closely fixed around a prespecified "ideal" trajectory. Second, for many compounds as part of the procedures to guarantee product chemical consistency, regulatory authorities (such as the FDA in the USA) demand proof that consistent operation is adhered to. Without this the product cannot be sold. In industrial fermentation systems, process operators are typically employed to achieve consistent operation through manual monitoring and control. The operator uses his experience and knowledge of the fermentation process, together with information provided by supervisory control systems, to detect potential problems and make modifications when necessary. The importance of effective

operator control cannot be underestimated as the performance of a fermentation is very much dependant upon the ability to keep the system operating smoothly. A fermentation that is free from major upsets is likely to be more productive than one subject to significant disturbances. Therefore, the earlier a potential problem to the system can be detected, the less severe its influence will be and the resulting corrective action will consequently be more restrained.

Although operator control may be adequate in certain situations it makes little direct use of the historical data that is routinely gathered and logged. Historical data will typically contain information on high- and low-productivity batches, as well as information on the consequences of performing particular actions in response to problem situations. It is possible therefore to develop simple rule-based structures that compare individual variables with historical records. Any deviations that indicate reduced productivity from the current batch can then be brought to the attention of the operators. Rules themselves can be formulated through the use of such methodologies as rule-induction procedures and case-based reasoning (Leake, 1996). With the rules formulated, it is relatively simple to implement them, particularly with recent developments in real-time knowledge-based systems, such as G2 from Gensym Ltd. These approaches are fine in certain situations, such as final productivity analysis. However, the problem when applying the methods to on-line fermentation monitoring is the sheer complexity of the system together with the need to account for temporal patterns.

An alternative technique for monitoring batch processes that also utilizes historical data is Statistical Process Control (SPC) (Wetherill and Brown, 1991). Traditional univariate SPC may be suitable for selected fermentation systems (Hahn and Cockrum 1987; Vander Wiel et al., 1992), however, such processes in general pose a variety of problems that make univariate SPC inappropriate. For example, many variables may be recorded, necessitating the need for multiple charts to be interpreted, which can be difficult. Other

problems include the fact that steady state is not achieved in batch operation and deviations may be caused by interactions between variables which may not appear on SPC charts.

To overcome the problems associated with univariate SPC, multivariate SPC (MSPC) techniques have been developed (Nomikos and MacGregor, 1995a; Wise and Gallagher, 1996) and subsequently applied to batch-fermentation systems. The aim of this work was to develop a process monitoring system for a fed-batch fermentation system operated by Biochemie in Austria. In developing the system the capabilities of a variety of MSPC techniques were compared and assessed. In addition, a number of modifications to the existing technologies are described and it is demonstrated how they can improve the monitoring capabilities of the system. The utilization of artificial neural networks (ANNs) to infer product quality measurements is also detailed. This was required due to the problems associated with making direct on-line measurements of quality. The work with ANNs represents an initial study into the development of a closed loop control system to regulate the fermentation system.

This article is divided into several sections. First, we describe the mathematical algorithms that were used in this study. This is followed with a description of a number of practical considerations that were investigated to make the algorithms more suitable for batch fermentation systems. Then, we detail the results of applying the algorithms to the industrial fed-batch fermentation system and finally, outline the important conclusions from the work.

For propriety reasons many details from the fermentation process, including labels on graphs have been omitted.

COMPONENT TECHNOLOGIES

This article describes the on-line application of multivariate statistical process control (MSPC) and artificial neural network (ANN) techniques to a fermentation process. Before the application is detailed a brief description of the algorithms employed in this study is provided. These descriptions assume some knowledge of the technologies, which, if required, can be obtained from the publications detailed in Table I which provide excellent introductions to the topics.

Multivariate Statistical Process Control

The main focus of this work was the application of MSPC technologies to the industrial fermentation system. The

Table I. References for further information.

Topic	Reference
Principal Component Analysis	Wise and Gallagher (1996)
Partial Least Squares	Geladi and Kowalski (1986)
Multi-way Techniques	Kourti and MacGregor (1995)
Artificial Neural Networks	Willis et al (1991)
Radial Basis Functions	Warnes et al (1998)

work concentrates on the application of principal component analysis (PCA) and partial least squares (PLS), also known as "projection to latent structures." The basic concepts of these algorithms are described in the following sections. The algorithms were originally formulated for application to continuous systems, however, it is explained in a following section how simple data preprocessing procedures can be employed to enable their application to batch processes.

Principal Component Analysis

Principal component analysis is used to analyze the covariance of a set of plant variables. The approach transforms a matrix containing m measurements from n process variables, $[\mathbf{Z}]$, into a matrix of mutually uncorrelated variables, $\mathbf{t}_k$ (where $k =; 1$ to n) of length m. These variables, called principal components (PCs), are transforms of the original data into a new basis defined by a set of orthogonal *loading* vectors, $\mathbf{p}_k$, of length n. The individual values of the principal components are called "scores." The transformation is defined by

$$[\mathbf{Z}] = \sum_{k=1}^{np<n} \mathbf{t}_k \mathbf{p}_k^T + \mathbf{E} \qquad (1)$$

The loadings are defined here as being orthonormal, and so they become the eigenvectors of the data covariance matrix, $\mathbf{Z}^T\mathbf{Z}$. The $\mathbf{t}_k$ and $\mathbf{p}_k$ pairs are ordered so that the first pair capture the largest amount of variation in the data and the last pair capture the least. In this way, it is generally found that a small number of PCs (np) can account for much of the power in the covariance matrix. The remaining power constitutes the error term E. When Eq. (1) is applied to a single vector of new process measurements, $\mathbf{z}^T$, the resulting term $\mathbf{E}$ is called the "prediction error." There are several methods for determining the suitable value for np. One method is to continue to add PCs until the variation explained in the retained PCs exceeds a particular value, however, a more suitable approach, and the technique used in this work, is to use cross validation (Wold, 1978).

By selecting a value of np lower than n the PCA algorithm is able to project highly correlated process data into a low-dimensional space defined by the principal components. Control charts that are simple to interpret can then be constructed using this data. The formulation of these charts is described in the section on process monitoring using PCA.

Partial Least Squares

Partial least squares is a tool suitable whenever plant variables can be partitioned into cause ($\mathbf{X}$) and effect ($\mathbf{Y}$) values. The method may be used for regression or similarly, to PCA, reduction of the effective dimensionality of data. The approach works by selecting factors of cause variables in a sequence that successively maximizes the explained covari-

ance between the cause and effect variables. Given a matrix of cause data, $\mathbf{X}$, and effect data, $\mathbf{Y}$, a factor of the cause data, $\mathbf{t}_k$, and effect data, $\mathbf{u}_k$, is evaluated, such that

$$\mathbf{X} = \sum_{k=1}^{np<nx} \mathbf{t}_k \mathbf{p}_k^T + \mathbf{E} \text{ and } \mathbf{Y} = \sum_{k=1}^{np<nx} \mathbf{u}_k \mathbf{q}_k^T + \mathbf{F} \qquad (2)$$

where $\mathbf{E}$ and $\mathbf{F}$ are residual matrices, np is the number of inner components that are used in the model and nx is the number of causal variables. These equations are referred to as the "outer relationships." The vectors $\mathbf{t}_k$ are mutually orthogonal. These vectors and $\mathbf{u}_k$ are selected so as to maximize the covariance between each pair, $(\mathbf{t}_k, \mathbf{u}_k)$. Linear regression is performed between the $\mathbf{t}_k$ and the $\mathbf{u}_k$ vectors to produce the inner relationship, such that:

$$\mathbf{u}_k = b_k \mathbf{t}_k + \varepsilon_k$$

where b_k is a regression coefficient, and ε_k refers to the prediction error. The PLS method provides the potential for a regularized model through selecting an appropriate number of latent variables, $\mathbf{u}_k$ in the model (np). The number of latent variables is typically made through the use of cross validation.

The section on process monitoring on PLS explains how this algorithm can be used to monitor the progress of a process.

Multiway Multivariate Statistical Process Control

Conventional PCA and PLS are linear procedures and therefore limited in their effectiveness when applied to nonlinear batch fermentation problems. Two options exist for improving the capabilities of the techniques when applied to batch systems. The first is to develop nonlinear counterparts to PCA and PLS and the second is to transform the batch data in such a way as to remove the nonlinear characteristics. Although nonlinear MSPC techniques exist and have been applied successfully to fermentation systems (Dong and McAvoy, 1996), the transformation of batch data has proved to be a more effective option and was therefore adopted in this study. The most common form of data transformation, termed multiway PCA and PLS, was initially proposed by Nomikos and MacGregor (1994). Since then, other researchers have adopted the approach and applied it to a variety of processes. For example, Gallagher et al. (1996) applied the technique to monitor nuclear waste storage vessels and Gregersen et al. (1997) investigated the detection of faults in a fed-batch fermentation process.

Multiway Principal Component Analysis

The concept of multiway PCA is a relatively straightforward extension to the approach taken for continuous systems but deviations from mean trajectories rather than steady state are considered. The following description together with Figure 1 explains the multiway approach to PCA.
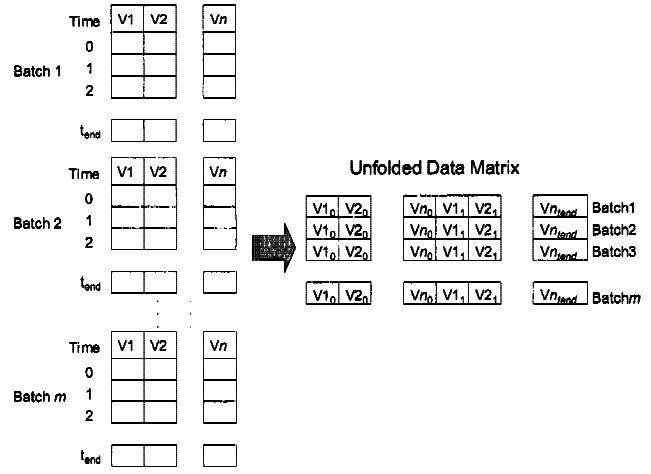

Figure 1. Multi-way MSPC.

Select m historical batches which all yielded high product concentrations. These batches are referred to as "nominal" batches and will be used for comparison purposes. The duration of each batch is likely to differ and therefore the data from each batch is considered only until the shortest run length. Techniques exist which allow all the measured process data to be used regardless of the shortest run length (Lakshminarayanan et al., 1996). In this application however, the run lengths did not differ significantly and thus using the reduced sized data sets was not considered to be a problem.

The next step is to identify the, n, process variables that are to be monitored. For each variable the mean trajectory over all the nominal batches is calculated and subtracted from each process measurement. This effectively removes the major nonlinearity from the data and leaves a zero mean trajectory for each variable.

Unfold the individual data matrices from each batch into a single unfolded data matrix, as depicted in Figure 1. In this Figure $V1$ refers to variable 1, $V1_t$ refers to the value of variable 1 at time t and t_{end} refers to the time at which the batch finished. Finally, all columns of the unfolded data matrix are standardized to unit variance. Principal component analysis can now be applied to this unfolded data matrix using two techniques:

1. Projection PCA

   A PCA model is developed using the entire unfolded data matrix, i.e., there is a loading variable calculated for every measured process variable at each sampling instant. The subsequent use of this model to monitor a batch on-line poses the problem that it is necessary to know the values of all process measurements through to the end of the batch. This means that with the exception of the end point of the batch, it is necessary to estimate the future values of all the measured variables. The prediction of future process values is normally referred to as "filling up" the matrix and is typically achieved by assuming that either all future scaled process values remain at the mean value of zero (filling method 1); or the future

process values remain at the current offset from the mean value (filling method 2).

An alternative technique that has been utilized in this work is to use the capability of PCA and PLS to handle missing data (Nelson et al., 1996). Research work in this area has tended to concentrate on using PCA and PLS to infer process measurements at the current sampling instant when a particular measurement is unavailable. However, the same techniques can be used to fill up the data matrix. The technique was employed in this work and compared with the other approaches described above (filling method 3).

2. Moving Window PCA

A PCA model is developed on a moving window of data. For example at the 50th sampling instant, a PCA model may be developed using data collected from the 41st to the 50th sampling instants. In this case, the moving window length would be said to be equal to 10. Such a model effectively compares operating conditions over the last few sample points with those experienced at the same time in the nominal batches. The advantage of this technique is that it does not require the values of future measurements to be predicted. However, it does require that the PCA model be re-evaluated at each sampling point. Both PCA algorithms are applied and compared in this study.

Multiway Partial Least Squares

The concept of multiway PLS is very similar to multiway PCA. The unfolded data matrix represents the cause matrix and is created in exactly the same way. The construction of the effect matrix is problem dependent and is detailed in the section below on process monitoring using partial least squares.

Artificial Neural Networks

Artificial neural networks (ANNs) are mathematical functions that provide a nonlinear description of the relationship between cause and effect variables. Although many differing types of ANN exist (Lippman, 1987), they do possess some common features. They are generally composed of numerous process elements, termed nodes, which are arranged together to form a network. A common type of ANN model used in many applications is the feedforward network. This type of network comprises an input layer where input information is presented to the network, one or more *hidden* layers where neuron processing takes place and an output layer from which the network outputs are obtained. It is termed a feedforward network because the outputs from one layer are fed-forward as inputs to the subsequent layer. In such a network the processing element is one that weights the input signals and then sums them together with a bias term. The neuron output is then obtained by passing the summed, weighted inputs through a nonlinear activation function, such as the hyperbolic tangent. An alternative type of neural network and the one chosen for this study is the radial basis function (RBF) neural network.

The RBF networks utilize a clustering process on the input data before presentation to the network and uses non-linear activation functions that are locally tuned to cover a region of the input space. The network structure consists of an input layer, a single hidden layer containing the same number of nodes as cluster centers, and an output layer. In this application the hidden layer nodes are made up of symmetrical Gaussian density functions. The activation, and hence, the output of the hidden units is dependent upon the distance between the given input vector and the unit center. The closer to the center the input lies, the higher the activation of the unit, i.e., the larger the value sent on to the output layer in the network. Only the connections from the hidden to the output layer are weighted, leading to a fast training rate. This fast training rate means the neural network models can be quickly developed, and the fixed layer structure of the RBF network (i.e., only one hidden layer) places a convenient restriction on topology selection when compared to alternative network structures. An additional property of RBF networks is that it is relatively easy to introduce on-line calculation of confidence limits for the model estimations, providing a measure of local reliability.

The training procedure for the RBFNs can be decomposed naturally into three distinct stages: (1) locating the centers of the hidden layer radial units; (2) determining the width of each radial unit; and (3) calculating the network weights for interconnections between the radial basis layer and the output layer. The scheme proposed by Moody and Darken (1989) to perform the training is used in this work. Further details of the precise methodology used for RBF network construction and confidence-bound determination can be found in Warnes et al. (1998).

Process Monitoring Using Principal Component Analysis

Process monitoring is achieved with PCA by developing the PCA model on example data from the process. The data chosen for this stage should have been collected during successful, high-yield fermentation runs and is referred to as "nominal data." The model developed with this data is later recalled using on-line data and the consistency of this data is assessed. Various approaches for assessing this consistency have been suggested in the literature. In this and previous work (Goulding et al., 2000), it has been found that for PCA the analysis should be applied separately to data variation in the space of the principal components and in the prediction errors.

Variation in the space of the principal components is provided through the T^2 statistic that is defined as:

$$T^2 = \sum_{k=1}^{np} \mathbf{t}_k \sigma_k^{-2} \mathbf{t}_k^T \tag{3}$$

where σ_k^2 is the variance of the kth score.

Variation in the prediction error is expressed as the squared prediction error (SPE) and is defined as $\mathbf{E}^2$.

Control limits can be applied to the T^2 and SPE charts based on the assumption that the statistics follow a normal and chi-squared distribution, respectively. These limits are termed the T^2 and Q limits, respectively, and a description of how they are calculated is provided in Appendix A.

While a new batch is running, the T^2 and *SPE* values can be plotted. Any violations of the confidence limits should indicate that the current batch is deviating from nominal conditions. However, the manner in which it is deviating from normality differs depending upon which chart is violating the confidence limits. Goulding et al. (2000) demonstrated that changes in the relationships between variables, such as that experienced if a sensor failed, tended to be detected on the *SPE* chart, while changes in operating condition, for example a grade change, were typically identified on the T^2 chart. There will be exceptions to this, for example, a high-impact fault which significantly affects a number of variables is likely to be detected on both the T^2 and *SPE* charts.

An important consideration when constructing the *SPE* and T^2 charts is setting the confidence limits. Most reported applications tend to place 95% or 99% confidence limits on the charts. However, experience shows that the correct setting of the confidence limits is critical to the success of the monitoring system. Process operators will quickly lose confidence with a system that gives many false alarms, and therefore the limits should be set so that such alarms are minimized. This aspect is further detailed in the section on application of PCA.

While highlighting abnormal conditions is very useful, assigning the cause of the abnormalities is arguably even more important. Technology in this aspect of MSPC is limited and is an area of significant research (Gertler et al., 1999). In this work simple contribution charts were used to indicate cause.

Contributions to the T^2 statistic are obtained by taking the gradient of T^2 with respect to each variable. The contribution to the *SPE* statistic from a given variable is simply the squared prediction error on that variable.

Process Monitoring Using Partial Least Squares

Partial least squares is a technique for determining a linear relationship between cause and effect variables. Its advantage over ordinary least squares is that it is capable of handling large ill-conditioned matrices, as are typically encountered in batch systems. Previous work has demonstrated that it can be used for process monitoring in two very different ways. The first approach uses PLS as a prediction tool to estimate the end concentration of biomass in the fermentor. The second technique is very similar to PCA and involves monitoring the inner latent variables produced by the algorithm. In this study, it was found that the second approach offered no tangible benefits over PCA and was therefore not integrated into the final monitoring system (Lennox et al.,

1999). Using a PLS model to predict the final biomass concentration was, however, found to be beneficial. The PLS model acts as a classification system and can be used to indicate if the current batch is more consistent with high- or low-yield batches. Because it is used as a classifier it is important that the model is developed using historical data from both high- and low-yield batches. The construction of the cause and effect matrices is relatively straightforward. The cause matrix is identical to that used in multiway PCA and the effect matrix is a column vector containing the final product concentration from each of the historical batches.

Confidence limits can be placed around the PLS prediction (Nomikos and MacGregor, 1995b) and if the PLS model predicts that the required amount of biomass will not be produced from the batch then a warning message can be relayed to the operator.

Process Monitoring Using Artificial Neural Networks

Previous research has demonstrated that ANNs can be used as a non-linear counterpart to PCA and PLS algorithms, for example, auto-associative ANNs (Kramer, 1992). In this study, linear PCA and PLS have been found to provide adequate results and therefore, ANNs have not been investigated in this capacity. However, the longterm plans for this work are to provide automatic feedback control of biomass production in the fermentor. It is anticipated that employing ANNs within a model-based control system will provide a suitable controller for this application. As a preliminary study, ANNs have been employed as soft sensors to predict the concentration of biomass in the fermentor during the batch.

ON-LINE CONSIDERATIONS

Data Preprocessing

It is important that any monitoring system that is developed be capable of handling common data problems such as noise and outliers. Tham (1994) provides a thorough review of data-handling procedures and the recommended techniques suggested in his article have been successfully employed in this work. These procedures involve filtering out noise using low-pass filters and detecting and removing outliers through the identification of trends in the data.

Missing Data

A common problem in process systems is that some measurements may not always be available. For example, a faulty thermocouple may be withdrawn from service for the duration of a batch. Such a situation would mean that it would not be possible to apply any previously developed PCA or PLS model that used this thermocouple measurement to the current batch. In fact, any blank entries in the

unfolded data matrix, which would result from a single missing measurement, would render the PCA or PLS model invalid during an on-line batch. In this application it was considered likely that one or more process measurements would be unavailable during a typical batch run. It was therefore considered to be essential that this problem be addressed.

Problems associated with missing data values have been researched at length by Nelson et al. (1996). In their studies they identified three methods for estimating missing data values, each of which makes use of the predeveloped PCA or PLS models. Each of these approaches was applied in this work with mixed results. It was found that two of the methods, projection to the model plane and replacement by the conditional mean created matrix singularity problems and were unsuitable for this application. However, the single-component regression technique was found to provide acceptable results. Complete details of these algorithms can be found in Nelson et al. (1996).

Pre-Culture Data

The progress of a batch is dependent upon not only the conditions in the fermentor, but also the conditions when the biomass was in its pre-culture, or seed, stage. It is possible to employ measurements taken in the pre-culture stage in the monitoring system by simply appending the information to the unfolded data matrix that is used in both the PCA and PLS routines. Such information was not available in this work and thus, has not been applied. However, Ignova et al. (1999) have demonstrated how important information regarding the progress of the batch can be extracted from the pre-culture data.

APPLICATION OF CONDITION MONITORING TECHNOLOGIES TO THE FED-BATCH FERMENTATION

The work detailed in this article represents the first phase of an ongoing project to implement a condition-monitoring system onto an industrial fed-batch fermentation system. The aim of this work is to assess the suitability and benefits of applying advanced monitoring technologies to the batch process. Based upon the outcome of this phase of the work, the next stage will be to apply the monitoring tools to a research facility and test their performance over a period of time. Based upon the results of this phase, the monitoring system will be applied to the industrial fermentation system.

Application of Principal Component Analysis

A PCA model was developed from historical data from 10 high-yield industrial batches. The ability of this model to monitor a number of subsequent batches was then determined. Some of these subsequent batches operated smoothly while others were affected by disturbances.

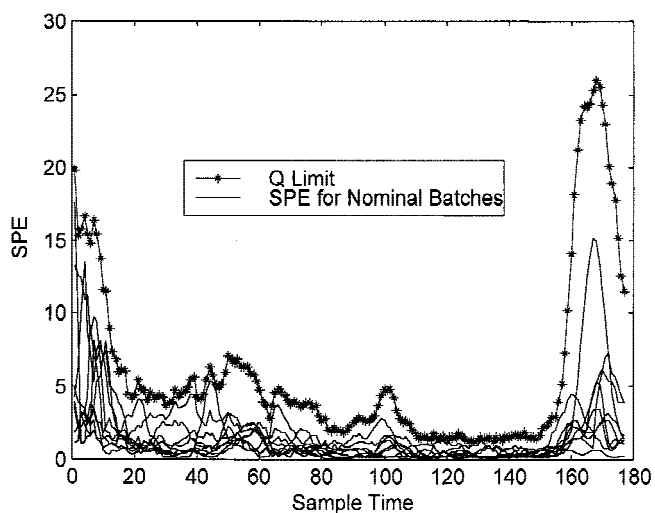Figure 2 displays the SPE for each of the 10 reference

Figure 2. SPE chart for nominal batches.

batches. Based on this information, a 97% confidence limit has been evaluated and also plotted. In this example the unfolded data matrix has been filled using filling method 2. A 97% confidence limit has been specified in this application because at this level none of the nominal batches violate the limit, thus reducing the potential for false alarms.

Figure 3 shows the SPE, along with 97% confidence limits, for a particular test batch. During this fermentation an intermittent drift on a sensor measurement was experienced. This drift was most apparent between sample numbers 40–70 and 100–150. The SPE chart displayed in Figure 3 demonstrates that the PCA model has detected this fault entering the system.

Further investigation of the PCA model showed that the T^2 limit was never exceeded during this batch and therefore, the chart was unable to identify this fault. This result agrees with the analysis in the section on process monitoring using PCA. The T^2 chart is able to detect significant changes in operation and high-impact faults. A fault such as this, a
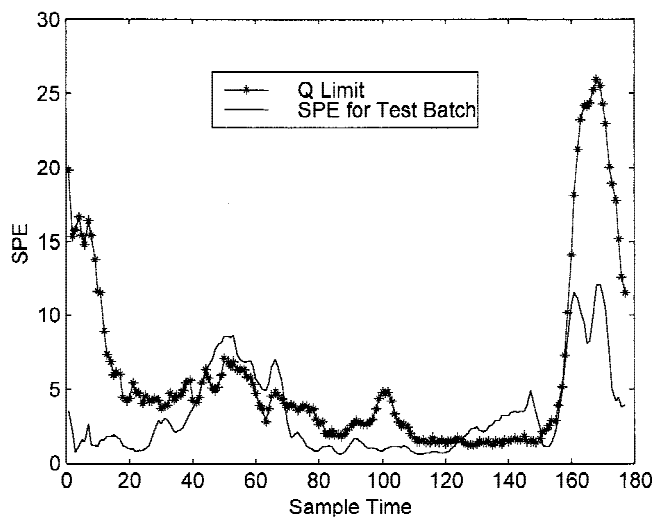

Figure 3. SPE chart for test batch.

small drift on a sensor is unlikely to be detected in the T^2 chart.

Section on multiway PCA described two alternative techniques for filling up the unfolded data matrix. The suitability of these techniques was compared with the results obtained using filling method 2. Figure 4 shows the SPE chart for the test batch investigated above, this time using filling method 1. By using this technique the PCA model is no longer able to identify the sensor fault. Similar reduction in monitoring capabilities of this approach have been reported by Lakshminarayanan et al. (1996) and Nomikos and MacGregor (1995a).

The reason for the reduction in performance using filling method 1 can be explained by considering Figure 5. This figure shows a particular process variable measurement recorded during a complete batch cycle (bold line). The batch chosen in this example was a high-yield batch free from process disturbances. Using the data collected up to sample point 90, the value of the process variable between sample times 91 and 180 has been inferred using each of the three methods to fill up the unfolded data matrix. It is evident from this figure that filling method 2 most accurately matches the actual process measurement. Because the performance of the process monitor will be dependent upon the accuracy with which it fills the unfolded matrix, filling method 2 would appear to be the more suitable in this application.

Similar results were repeated in many other tests using different variables and batches, confirming the suitability of filling method 2.

The low-power principal components that are combined to create the SPE chart capture relationships between the process variables and as such, can be considered to be models of the system. The more accurate these models are the more sensitive the SPE chart will be to abnormal process conditions. The value of the Q limit, which in this work is specified as being greater than the maximum SPE over all of the nominal batches, provides a measure of the accuracy of
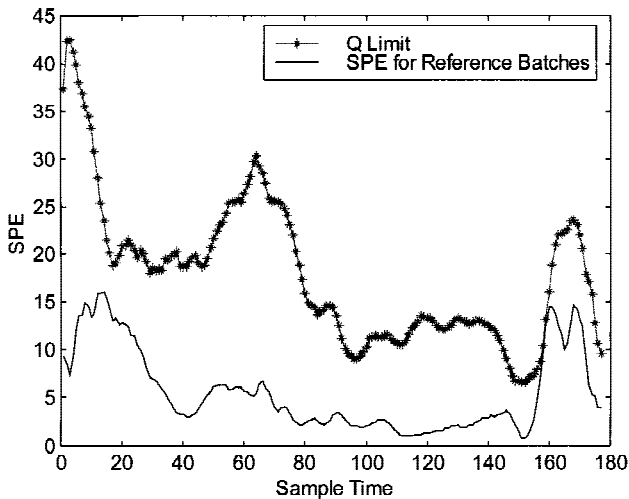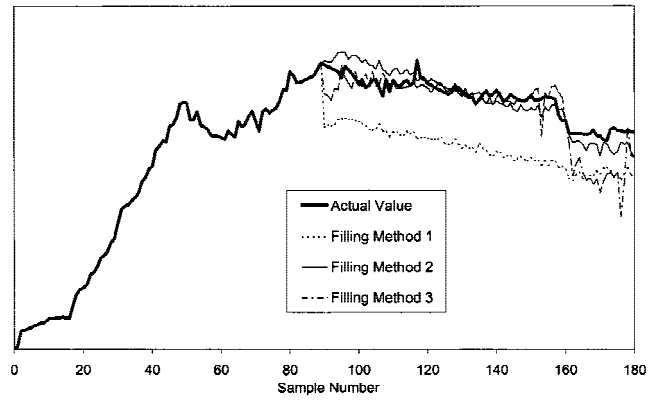

Figure 5. Inferred process measurement using various filling methods.

the low power PCA models. The lower the Q limit, the more accurate the models are.

Figure 2 indicates that the sensitivity of the SPE is low at the beginning and end of the batch and high during the middle. The low sensitivity at the start is because so little data is available from the batch and therefore the unfolded data matrix is being filled with inaccurate data. The sensitivity at the end of the batch is believed to be because the relationships between the variables differ during the progression of the batch, for example, cell death phenomena will affect the latter stages of the batch. Relationships between variables that are valid at the early stages of the batch may not hold towards the end of the batch. To account for this, it is possible to construct PCA submodels. Using this technique, PCA models are constructed using data collected from only part of the batch. For example, one PCA model may be applied up to sample number 60, a second between sample 60 and 150, and a third between 150 and 180.

Figure 6 compares the Q limit obtained using such a submodel approach compared with that using a single PCA model over the full duration of the batch. The figure shows that the SPE and the corresponding Q limit using submodels is significantly lower than that using the single PCA model. This indicates that the SPE chart should be more sensitive to process abnormalities, such as faults, if submodels are em-
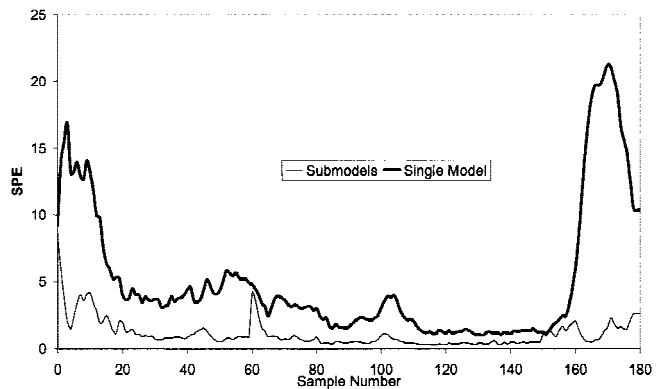

Figure 4. SPE chart for test batch using filling method 1.

Figure 6. Effect of using sub-models.

ployed and is therefore a more suitable method for process-monitoring purposes. One minor drawback with using sub-models is that the Q limit can increase sharply following the change from one model to the next. This effect can be seen most notably in Figure 6 at sample number 60.

The reason for this is the same as the high SPE experienced at the start of the single model technique and is because so little data is available when filling up the unfolded data matrix. To reduce this problem it is possible to construct submodels that operate during particular periods but are developed using data collected both before and during the period. For example, a submodel that operates between sample numbers 80 and 120 may actually be developed using data collected between sample numbers 60 and 120. Therefore, when the model is introduced at sample number 80 the data it uses to fill up the unfolded data matrix will be more reliable.

The ability of PCA to handle missing data values is demonstrated in Figure 7. The solid line in this chart shows the SPE over the course of a batch.

Data collected from this batch was passed through the PCA algorithm once more. However, this time the measurement from a particular variable was set to be missing between samples 41 and 60 and 101 and 130. Using the single-component regression method to replace the missing values the SPE chart obtained for this batch is given by the dotted line in Figure 7. It is evident from this chart that there is very little difference between the two SPE values through the duration of the batch and therefore in this particular example, the routines for handling missing data values have worked very well. Similar results to that displayed in Figure 7 were recorded from a number of subsequent missing data tests involving other process variables and batches.

Further tests using the moving window PCA technique showed that it produced similar results to those obtained using the projection method with submodels. It is therefore difficult to establish conclusions as to which is the more suitable technique for this application. An advantage in using the moving window approach is that there are fewer variables that require specifying, such as the number of submodels and the periods over which these submodels should be used. However, it was found that the moving window approach required accurate specification of the time delays, while the projection method was more robust to this information.

Application of Partial Least Squares

As discussed earlier, it was found that monitoring the latent variables of the PLS model gave no further insight into the operation of the fermentation system than is available through PCA. Therefore, in this study the inner relationships in the PLS models were ignored and only the prediction accuracy of the models was monitored.

Partial least squares' models were developed to estimate the final concentration of biomass in the fermentor. The aim of this exercise is to use this estimate to categorize the on-line fermentations into high- and low-yield production.

The PLS model was developed on historical data and applied to a number of alternative batches. Figure 8 shows an example of a PLS-monitoring chart. This chart shows the final biomass concentration (solid line), estimated at each sample time during the batch, 95% confidence limits (dashed line) and the actual final concentration of biomass (thin solid line). Details of the confidence limit calculations can be found in Nomikos and MacGregor (1995b).

It can be seen from Figure 8 that for this batch, which proceeded upset-free, the PLS estimate of the final product concentration was reasonably accurate throughout. It is also noticeable that the biomass estimate during the first half of the run is relatively noisy and for the second half of the run it is reasonably consistent. This is because at the start of the batch most of the data in the unfolded data matrix has been estimated and is therefore less reliable than in the later stages of the batch.

The accuracy of the PLS model displayed in Figure 8 was consistent with that produced for several other batches. This level of accuracy makes the PLS model a useful tool for monitoring the progress of the fermentation system for this particular application.

Application of Neural Networks

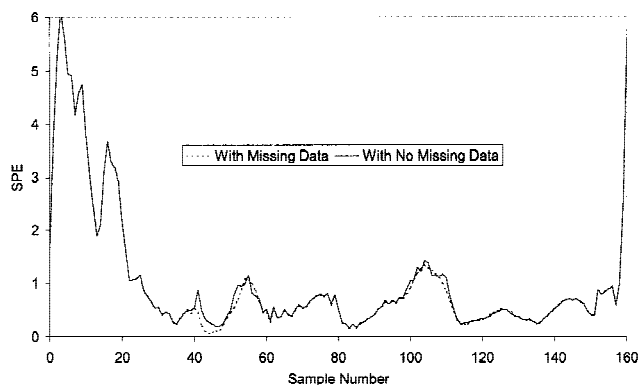The ability to formulate a relationship between on-line measurements and off-line assays using a RBF network pro-
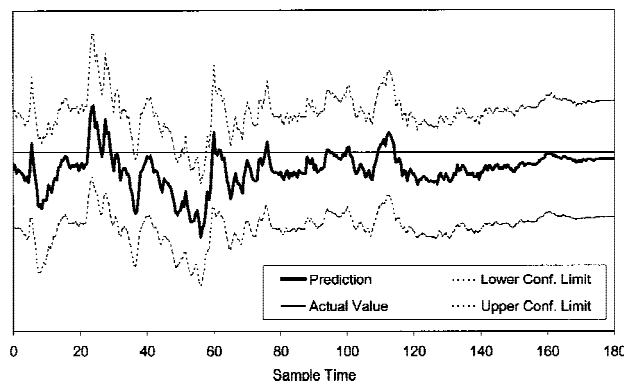

Figure 7. Effect of missing data.

Figure 8. PLS monitoring chart.

vides on-line estimates of performance. This information is complementary to that provided by the MSPC methods. To build the RBF model it is necessary to first select batches for RBF network parameterization. The batch data requirements are similar to those for PLS-based modeling and require a wide coverage of operational experience to maximize the model coverage.

Figure 9 demonstrates the predictive capability of the RBF network. Inputs to the network are on-line measurements from the process and the output is the biomass level measured in the laboratory from off-line samples. The network was trained using a selected set of batches exhibiting varying behavior spreading the region of operation. The inputs to the network were chosen using a combination of process knowledge and off-line verification of network behavior. The network topology was specified using cross-validation to determine the topology that maximized the predictive capability. Figure 9 shows the data that was used to test the network behavior and it is important to note that it did not feature in the training procedure. The crosses represent off-line samples and the continuous line is the network prediction. The dashed lines either side of this are the 95% confidence bounds in prediction accuracy. It can be seen that a good agreement between on-line prediction and performance verified by off-line analysis is achieved. The estimates are currently available to the process operator for information purposes but in the longer term the option to close the loop exists.

CONCLUSIONS AND FURTHER WORK

This article has provided details of a 2-year study aimed at developing a condition monitoring system for a fed-batch fermentation process. The primary conclusion from the work is that existing technologies in the field of process monitoring, such as PCA and PLS, provide a suitable tool for the detection of process abnormalities.
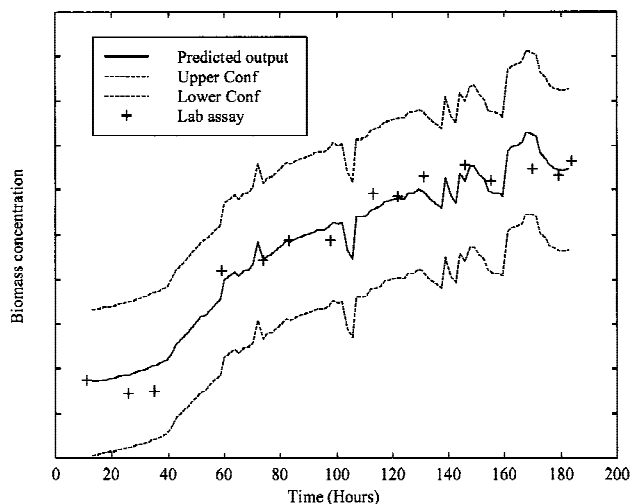
Principal component analysis has been employed successfully to detect and isolate process faults while PLS has been applied to estimate final product composition in the fermentor. The PLS estimate was shown to be sufficiently accurate to enable it to categorize on-line batches into high and low yield.

The integration of the PCA and PLS methods with information from other sources such as inferences from RBF networks serves to provide operators with a rich source of operational assistance.

Described herein are many of the practical considerations that have been encountered in this application. Such considerations have ranged from data preprocessing issues to computation problems.

The monitoring procedures developed in this work are currently being applied to a research facility operated by Biochemie. Subject to the success of trials conducted in this facility the intention is to then transfer the monitoring system to the industrial production equipment. Further analytical work that will be conducted during this time will focus on assessing whether or not there are advantages in using data collected during the pre-culture stage of the batch and if advantages exist to determine the optimal way to exploiting this information.

NOMENCLATURE

b_k	Regression co-efficient from PLS
$\mathbf{E}$	error matrix from PCA
$\mathbf{F}$	error matrix from PCA and PLS
$F_{np,r-l,\alpha}$	value of the F-distribution at the α% confidence limit for np principal components and r samples
g	function of the traces of the residual covariance matrix
h	function of the traces of the residual covariance matrix
h_O	function of the traces of the residual covariance matrix
k	index for principal components and latent variables
m	number of reference batches selected
n	number of process variables
np	number of principal components or latent variables retained
nx	number of cause variables
$\mathbf{p}_k$	loading vector from PCA and PLS
$\mathbf{q}_k$	loading vector for effect data from PLS
Q limit	confidence limit for T^2 values
Q_α	value for Q limit at the α% level of confidence
r	number of samples in the matrix $\mathbf{Z}$
SPE	square prediction error
$\mathbf{t}_k$	principal component or latent variable
T^2	measure of the variation in the space of the principal components
$\mathbf{u}_k$	score vector for effect data in PLS
V	residual covariance matrix
X	matrix of cause data
$\mathbf{Y}$	matrix of effect data
z	vector of new process data
$\mathbf{Z}$	matrix of process data

Figure 9. On-line biomass concentration estimation.

APPENDIX A

Statistical confidence limits for T^2 are calculated using the F-distribution as follows:

$$T_\alpha^2 = \frac{np(r-1)}{r-np} F_{np,r-1,\alpha} \qquad (A1)$$

where r is the number of samples in $\mathbf{Z}$, α is the confidence limit, expressed as a fraction and np is the number of principal components retained in the model.

Confidence limits for the SPE chart are based upon the chi-squared distribution (Jackson and Mudholkar, 1979).

$$Q_\alpha = \theta_1 \left[1 - \frac{\theta_2 h_0 (1 - h_0)}{\theta_1^2} + z_\alpha \frac{\sqrt{2\theta_2 h_0^2}}{\theta_1} \right]^{\frac{1}{h_0}} \qquad (A2)$$

where:
Q_α is the confidence limit for the SPE chart at the $\alpha\%$ level.

$$\theta_1 = \text{trace}(\mathbf{V}), \; \theta_2 = \text{trace}(\mathbf{V}^2), \; \theta_3 = \text{trace}(\mathbf{V}^3), \; h_0 = 1 - \frac{2\theta_1\theta_3}{3\theta_2^2}$$

$$(A2)$$

$\mathbf{V} = \dfrac{\mathbf{E}\mathbf{E}^T}{r-1}$, α is the confidence limit (0–100%) and z is the *Normal* variable.

The calculation of traces of powers of the residual matrix can become impractical with large data sets. This is particularly relevant in applying these techniques to batch systems as the control limits and PCA models are often calculated at each sampling instant. Where the calculations may be excessive it is possible to employ the confidence limits proposed by Box (1954) and given by:

$$Q_\alpha = g\chi_{h,\alpha}^2 \qquad (A4)$$

where: χ_h^2 is the chi-squared variable, $g = \dfrac{\theta_2}{\theta_1}$ and $h = \dfrac{\theta_1^2}{\theta_2}$

g and h can be approximated by matching moments (Nomikos and MacGregor, 1995a) as shown below.

$$g = \frac{\sigma_E^2}{\mu_E}, \; h = \frac{2\mu_E^2}{\sigma_E^2}, \; \sigma_E \text{ and } \mu_E \text{ refer to the standard deviation}$$
and mean of $\mathbf{E}$.

Figure A1 shows a comparison of the confidence limits produced using the trace of the residual matrix compared with those obtained by matching moments. The graph indicates that there is very little difference between the two
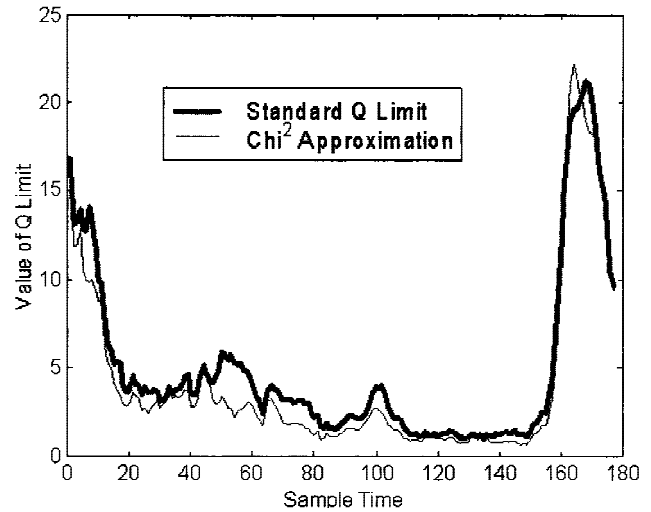

Figure A1. Comparison of confidence limits.

techniques and that if large data sets are encountered (they weren't in this application) then the matching moments algorithm can be used as an alternative. Care should be taken when using this algorithm as errors can be introduced if outliers are present in the data (Nomikos and MacGregor, 1995a).

References

Box GEP. 1954. Some theorems on quadratic forms applied in the study of analysis of variance problems: Effect of inequality and variance in one-way classification. Ann Math Stat 25:290–302.

Dong D, McAvoy TJ. 1996. Batch tracking via non-linear principal component analysis. AIChE J 42(8):2199–2208.

Gallagher NB, Wise BM, Stewart CW. 1996. Application of multiway principal components analysis to nuclear waste storage tank monitoring. Comp Chem Eng 20S:739–744.

Geladi P, Kowalski BR. 1986. Partial least squares regression: A tutorial. Anal Chim Acta 185:1–17.

Gertler J, Li WH, Huang YB, McAvoy T. 1999. Isolation enhanced principal component analysis. AIChE J 45(2):323–334.

Goulding PR, Lennox B, Sandoz DJ, Smith K, Marjanovic O. 2000. Fault detection in continuous processes using multivariate statistical methods. Intl J Syst Sci 31(11).

Gregersen L, Jorgensen SB, Andersen MY. 1997. Industrial fed-batch fermentation monitoring. International Symposium on Advanced Control of Chemical Processes. Banff p. 49–54.

Hahn GJ, Cockrum MB. 1987. Adapting control charts to meet practical needs: A chemical processing application. J Appl Stat 14:35–52.

Ignova M, Montague GA, Ward AC, Glassey J. 1999. Fermentation seed quality analysis with self-organizing neural networks. Biotechnol Bioeng 64:82–91.

Jackson JE, Mudholkar GS. 1979. Control procedures for residuals associated with principal analysis. Technometrics 21:341–349.

Kourti T, MacGregor JF. 1995. Process analysis, monitoring and diagnosis, using multivariate projection methods. Chemometr Intelli Lab Syst 28:3–21.

Kramer MA. 1992. Autoassociative neural networks. Comp Chem Eng 16(4):313–328.

Lakshminarayanan S, Gudi RD, Shah SL, Nandakumar K. 1996. Monitoring batch processes using multivariate statistical tools: Extensions and practical issues. Proceedings of IFAC World Congress, June, San Francisco. p. 241–246.

Leake DB. 1996. Case based reasoning: Experiences, lessons and future directions. Cambridge, MA: MIT Press.

Lennox B, Montague GA, Hiden HG, Kornfeld G. 1999. Case study investigating multivariate statistical techniques for fermentation supervision. Comp Chem Eng 23S:827–830.

Lippman RP. 1987. An introduction to computing with neural nets. IEEE ASSP Mag, Volume 90(5), April, 4–22.

Moody J,Darken CJ. 1989. Fast learning in networks of locally-tuned processing units. Neural Comp 1:281–294.

Nelson PRC, Taylor PA, MacGregor JF. 1996. Missing data methods in PCA and PLS: Score calculations with incomplete observations. Chemometr Intell Lab Sys 35:45–65.

Nomikos P, MacGregor JF. 1994. Monitoring batch processes using multiway principal component analysis. AIChE J 40(8):1361–1375.

Nomikos P, MacGregor JF. 1995a. Multivariate SPC charts for monitoring batch processes. Technometrics 37(1):41–59.

Nomikos P, MacGregor JF. 1995b. Multi-way partial least squares in monitoring batch processes. Chemometr Intell Lab Sys 30:97–108.

Tham MT. 1994. Succeed at on-line validation and reconstruction of data. Meas Contr May:46–56.

Vander Wiel SA, Tucker WT, Faltin FW, Doganaksoy N. 1992. Algorithmic statistical process control: concepts and an application. Technometrics 34:286–297.

Warnes MR, Glassey J, Montague GA, Kara B. 1998. Application of radial basis function and feedforward artificial neural networks to the *Escherichia coli* fermentation process. Neurocomp 20(1–3):67–82.

Wetherill GB, Brown DW. 1991. Statistical process control: Theory and practice. London: Chapman and Hall.

Willis MJ, Di Massimo C, Montague GA, Tham MT, Morris AJ. 1991. Artificial neural networks in process engineering. IEEE Proceedings-D 138(3):256–266.

Wise BM, Gallagher NB. 1996. The process chemometrics approach to process monitoring and fault detection. J Proc Cont 6(6):329–348.

Wold S. 1978. Cross-validatory estimation of the number of components in factor and principal components models. Technometrics 34:286–297