Automated Production Support for the Bioprocess Industry

Barry Lennox,*,† Kathryn Kipling,‡ Jarka Glassey,‡ Gary Montague,‡ Mark Willis,‡ and Hugo Hiden§

School of Engineering, University of Manchester, Manchester, U.K., Department of Chemical and Process Engineering, University of Newcastle, NE1 7RU, Newcastle upon Tyne, U.K., and GSE Systems Ltd., Hexham, Northumberland, U.K.

This paper describes the application of Artificial Intelligence and Multivariate Statistical Techniques to two industrial fermentation systems. In the first example, an Expert System is shown to provide tighter control of an important process parameter. This is shown to lead to improved consistency of operation. In the second application, Principal Component Analysis is applied to a final stage fermentation production facility. The results presented indicate that the algorithm can provide concise indicators of process faults that can be presented to the operators to assist them in taking suitable corrective actions.

Introduction

Production facilities in the process industries are under continuous demands to improve their efficiency of operation and the quality and consistency of their product. This represents a real and difficult challenge to systems engineers. The complexity of the problem comes as a result of the many generic features of process plants, a selection of which are listed below:

- Production facilities are becoming increasingly complex and consequently more difficult to fully understand.
- The quality of information obtained from the process (by way of on-line sensors and off-line laboratory analysis) is often limited.
- There are many problems that may adversely affect the operation of the process.
- There are often multiple responses that can be made to rectify any given problem, some of which will provide better results than others.
- There can be large delays before the effects of a change in the process are observed, thus often making it difficult to ensure that future production meets quality requirements.

While the above features can be considered to be general to all process plants, they are further compounded in the bioprocess industry by issues relating to biological aspects of production, which tend to lead to greater variability. This variation is often unacceptable, particularly in fermentation systems where it is important that operation is maintained within strict limits. The reasons for this are first, for many compounds regulatory authorities, such as the FDA in the USA, demand proof that consistent operation is adhered to, so that product chemical consistency is guaranteed, and second, biological systems are highly sensitive to abnormal changes in operation and therefore product yield and quality is highly dependent upon the consistency of operation.

Over time procedures tend to evolve on industrial fermentation systems that attempt to maintain consistent operational conditions. These procedures tend to rely heavily on the expertise of process operators and engineers rather than automatic control systems, which are hindered by difficulties associated with the measurement of the actual condition within the bioprocess. Significant problems that result through the application of such operational procedures include the fact that plant operators and engineers will each have varying levels of expertise and that situations will often arise when such expertise is unavailable.

The purpose of this paper is to demonstrate how assistance can be provided to process operators and engineers responsible for maintaining product consistency in industrial fermentation systems. This assistance is provided through the use of Artificial Intelligence (AI) techniques and by application of statistical analysis tools.

Pharmaceutical companies are beginning to recognize the value of AI techniques in general and Expert Systems (ES) in particular. An ES is a computer-based program that encodes rules obtained from process experts, usually in the form of "if–then" statements. These systems can operate on-line, where they receive measurements from the plant in real-time and pass on information to the operators. The information they transmit to the operators is dependent upon the responses from the rule sets that make up the ES. Such systems have been applied successfully to several bioprocess operations (*1*, *2*). These publications served to indicate future potential, but many implementation problems still remain and maximum use of available information from process data is not generally achieved.

A significant issue in the development of an ES is the elicitation of knowledge from the experts and the validation, storing, and encoding of this knowledge in a form that is maintainable. A number of approaches for developing ES have been proposed and include repertory grid analysis, card sort, and structured interviews (*3*). These approaches typically involve one to one interviews, during which the knowledge engineer attempts to elicit the reasoning and decision-making process of an expert

* To whom correspondence should be addressed. Email: barry.lennox@man.ac.uk.
† University of Manchester.
‡ University of Newcastle.
§ GSE Systems Ltd.

10.1021/bp0101839 CCC: $22.00
Published on Web 02/23/2002

within their domain. Unfortunately, each of the mentioned techniques suffers from difficulties relating either to the time required for the interviews to be undertaken or the construction of rules following the interview. An alternative approach to developing ES that has recently been developed is known as the Knowledge Acquisition Technique (KAT). This approach has been shown to reduce knowledge elicitation time and result in a complete, correct, and consistent knowledge base (*4*).

While the ES can provide useful information to operations staff, further support can be made available through the exploitation of the latest developments in statistical data analysis. This type of analysis makes use of the historical data that is routinely gathered and logged in the majority of bioprocess plants. These data contain information that will have been collected from both high- and low-yield batches, as well as information on the consequences of performing particular actions in response to various situations. Simple rule-based structures can be developed from these data that can 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. The rules can be formulated using a variety of methodologies, such as Rule Induction procedures and Case Based Reasoning (*5*). These approaches are acceptable in certain situations, for example in final productivity analysis.

The problem with applying rule-based supervision methods alone to on-line fermentation monitoring is the sheer complexity of the system together with the need to account for temporal patterns. To overcome this difficulty historical data can be utilized in Statistical Process Control (SPC) (*6*). Conventional univariate SPC may be suitable for selected fermentation systems (*7, 8*); however, bioprocesses in general pose a variety of problems that make this type of analysis 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 furthermore deviations may be caused by interactions between variables, which may not appear on SPC charts that monitor individual variables only.

To overcome the problems associated with univariate SPC, multivariate SPC (MSPC) techniques have been developed (*9*) and subsequently applied to batch fermentation systems. Preliminary results reported in the literature indicate the suitability of this approach to industrial fermentation systems.

This paper provides details of two studies that have been undertaken to assess the merits of using ES and statistical routines to monitor the progress and quality of industrial fermentation systems. These two investigations identify some limitations with existing state of the art technologies in the field and demonstrate that with some modifications MSPC and ES have the capability of providing significant benefits to the biochemical industry.

Problem Description

The two industrial fermentation systems that are investigated in this study are both very similar in that they involve the production of a secondary metabolite. Biomass is initially grown in vessels of increasing size. The product from the penultimate stage is referred to as the seed and provides the initial inoculum to the final production fermentation vessel. To ensure that the biomass remains healthy and grows at a suitable rate, conditions such as temperature and pH in the fermentation vessel are closely controlled. In the final stage the product is synthesized and subsequently recovered in downstream processing.

The aim for both of the industrial investigations detailed in this paper is to provide the process operators and engineers with greater insight into what is happening within the fermentation. The first application was undertaken in collaboration with Synpac Pharmaceuticals in the U.K. The research efforts concentrated upon the final production stage with the objective of maximizing process yield. The control policy at this stage is to maximize the carbon feeds within operational constraints and to reduce the feed rates only when problems arise within the vessel. This maximum productivity can be achieved when environmental conditions approach a biological constraint in the process. The closer the operation can be moved toward the constraint the higher the productivity. If the constraint is exceeded irreparable damage can occur to production. The implication of this is that tight control would allow operation closer to the constraint without violating it and thus lead to higher productivity.

The second investigation was supported by Biochemie Gmbh in Austria. The aim of this investigation was to determine whether MSPC techniques would be able to provide early warning of abnormalities occurring in the final production stage of their fermentation process. The amount of biomass present in their fermentation vessels is measured infrequently via laboratory analysis. An early indication that problems are occurring that may affect the growth of the biomass in the fermentation vessel is highly desirable as this information can be used to either rapidly respond to the problem or to stop the fermentation if necessary.

In the first application a rule-based control system was considered appropriate since heuristic expertise existed within the company for the maximization of productivity but could not be consistently applied by the process operators. It was felt that the ES application would result in a more consistent action. In the second application a wealth of process data was available and the requirement was to present it in a more meaningful manner. In this case the task was more suited to algorithmic methods.

Knowledge Elicitation for ES

The quality of the ES is highly dependent upon the correctness of the knowledge elicited from the experts, and the selection of suitable experts to be interviewed is therefore critical. In the first application, six of the plant operators who operated the plant on a daily basis and controlled the feeds through the existing DCS were considered as the process experts. In the KAT elicitation sessions they were asked “What would make you reduce the feed rates below the maximum level?” Their responses were coded in a tree-like structure.

The interview process was completed within the planned schedule, with the longest interview taking two man-days. The majority of the interviews took around one man-day. The exhaustive nature of the interviews limited single elicitation sessions to no longer than half a day. In theory there is no requirement for the knowledge engineer to be experienced in the subject domain. Best practice is to start with the most experienced expert who has most extensive coverage of the domain and is cooperative in offering all possible scenarios. This acts to focus the subsequent expert interviews more efficiently. In the case of this application, the knowledge engineers had an appreciation of the basic features of the

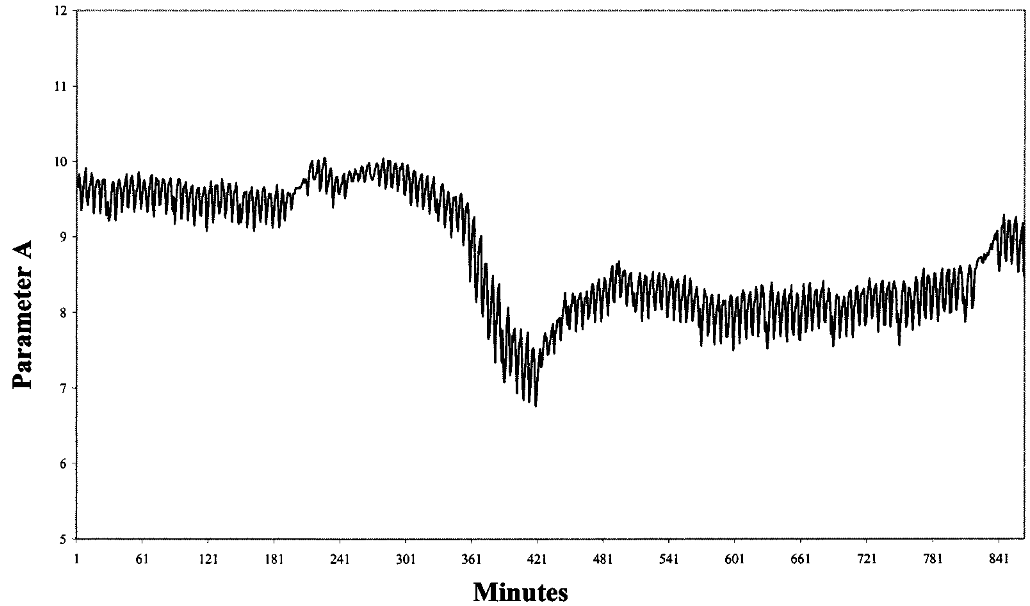

Figure 1. Example of raw process data.

domain. This resulted in a faster, more complete elicitation being achieved. However, there are several hidden dangers if the knowledge engineer is too experienced in the domain: biasing the knowledge base (KB) from preconceived opinions, modifying (rather than simply recording) the knowledge, and subconsciously intimidating the expert during the interview. Every effort was made to avoid these problems in this particular application.

As a panel of experts with varying experience was interviewed, it was necessary to combine the individual expert knowledge. Inevitably, some degree of overlap exists, and in some cases, apparent conflicts arise as a result of individual differences in experience. In these situations, arbitration is necessary and the KB "owner" is called upon to make the recommendation.

Using the KAT technique it was discovered that there were two main criteria, parameters A and B, which were critical. The knowledge was specified in terms of a flowchart that included the order in which actions should be taken and the magnitude of the carbon nutrient changes that were required. Flowcharts were developed for the control of both parameters A and B. These flowcharts with the explanatory notes provided the basis for the development of the knowledge base.

Validation and Implementation of the Knowledge Base. There are no prescribed ways in which to validate a knowledge base, and the tests that were carried out on this framework were designed using the flowchart from which the rule base was developed. Each path was considered, and the possible outcome of every action was analyzed. The first problem that was identified was that there was a high degree of noise in the system. This is illustrated in Figure 1. To solve the problem a first-order filter was used on the noisy variable, and this filtered value was used in any calculations. Although there was a small phase lag in the filtered data, this was not seen as a major problem because of the long time constant of the process.

Following the extensive testing of the code, the control system was implemented within the Gensym G2 system, and its advice and the actions of operators were monitored extremely closely. The control of parameter A will be used to demonstrate the capabilities of the real-time knowledge-based system. In assessing the benefits of improved control it is important to be able to compare the performance prior to control improvement to that achieved after. Historical data was obtained, and Figure 2 illustrates the changes in parameter A over a 13-hour period in the middle of the batch prior to the implementation of G2. Data from five batches are shown. Note that to maintain industrial confidentiality the *y* axis has been scaled and units removed. In interpreting these results it is important to realize that the first control objective is to maintain within batch consistency. The variations in parameter A observed in Figure 2 are significant and resulted in operators reducing carbon feed levels. Following the implementation of the G2 rule-based control system it was observed that the control parameter A was more consistent, as indicated in Figure 3.

It can be seen that the fluctuation in the value of parameter A is of the order of 1 when the system was under G2 control, whereas for the non-G2 controlled system the change in the value of parameter A was of the order of 2. This reduction in process variance following the introduction of improved control strategies is in keeping with general experiences of other workers (*10*). This is a useful metric for those wishing to carry out cost-benefit analysis.

Multivariate Statistical Process Control

As mentioned earlier, it can often be difficult to interpret the information that is collected, in the form of process measurements, from an industrial fermentation system. Primary reasons for this are the large number of measurements that may be available and the complex interactions that occur between these variables. A series of mathematical tools, collectively referred to as Multivariable Statistical Process Control (MSPC) have been developed in recent years to help interpret such data. The overall effect of MSPC algorithms is that correlated data with high dimensionality can be compressed onto a lower dimension without the loss of important information. The lower dimensional data set can then be interpreted with greater ease than the original data.

There are a number of MSPC tools that are available to help interpret process data, the most commonly applied methods being Principal Component Analysis

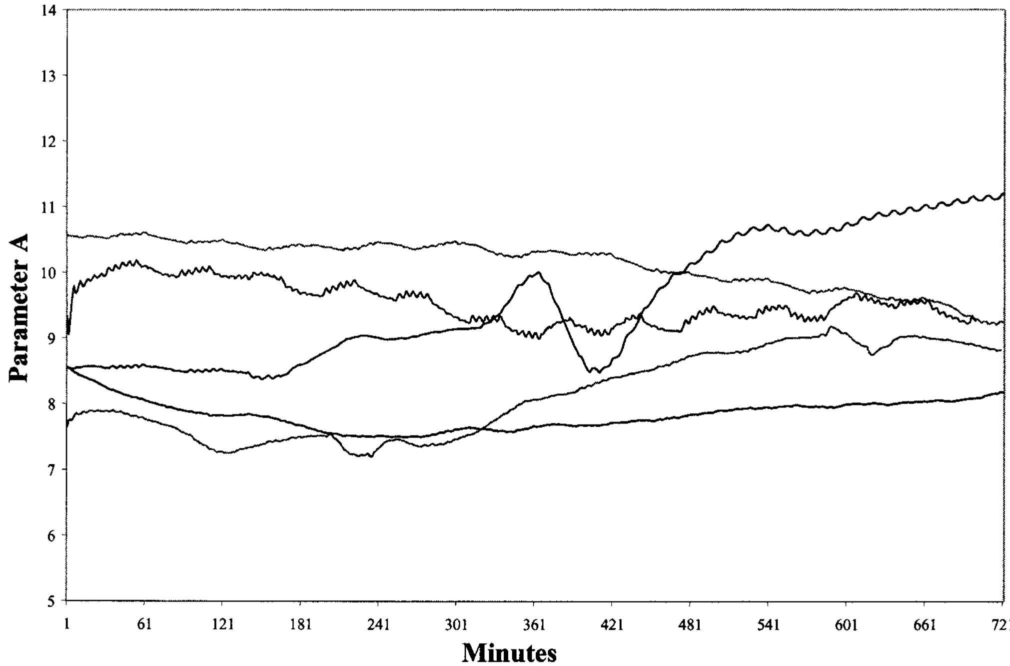

Figure 2. Variation in Parameter A prior to G2 based control.

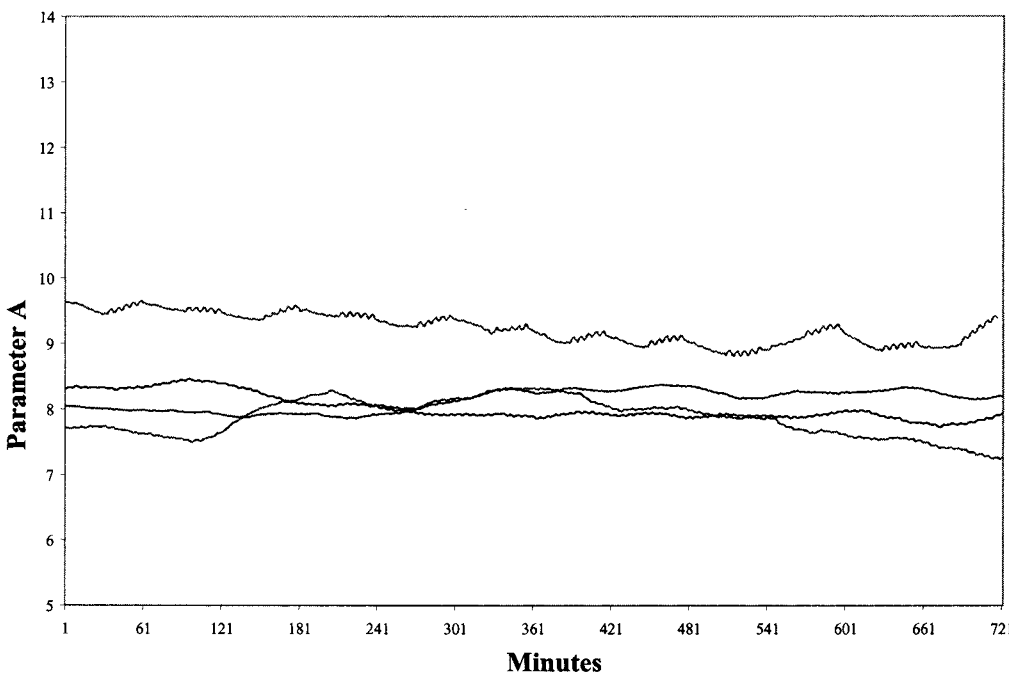

Figure 3. Variation in Parameter A with G2 based control.

(PCA) and Partial Least Squares (PLS). On the basis of the results obtained in previous studies by the authors and other researchers (*9, 11, 12*), PCA was believed to be a suitable algorithm to apply to the problem discussed here. A brief description of PCA is provided below, and further details of the algorithm are provided in ref 13.

Principal Component Analysis. PCA transforms a matrix containing m measurements of n process variables, [**Z**], into a matrix of mutually uncorrelated variables, termed scores, $\mathbf{t}_k$ (where $k = 1$ to n) of length m as defined by

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

where $\mathbf{t}_k\mathbf{p}_k^T$ are termed principal components (PCs) and $\mathbf{p}_k$ represents the set of orthogonal vectors, of length n, referred to as *loadings*. These loadings are defined here as being orthonormal and are equal to the eigenvectors of the covariance matrix $\mathbf{Z}^T\mathbf{Z}$. The $\mathbf{t}_k$ and $\mathbf{p}_k$ pairs are ordered so that the first pair captures the largest amount of variation in the data and the last pair captures the least. In this way it is generally found that a small number of PCs (np) can account for much of the variation in the data. Any variation in the data that is not explained by the np PCs is contained within the error term, $\boldsymbol{E}$. 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 a suitable value for np. One method is to

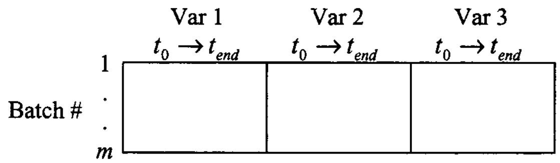

Figure 4. Unfolded matrix.

continue to add PCs until the variation explained in the retained PCs exceeds a particular value; however, a more suitable approach, the technique used in this work, is to use cross validation (*14*).

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. This low dimensional data can then be viewed in several different ways to detect patterns in the data and to identify abnormal operating conditions that may be the result of characteristics such as poor operational procedures or an instrumentation fault.

Multiway PCA. PCA is a linear procedure and is therefore limited in its effectiveness when applied to nonlinear batch fermentation problems. To overcome this limitation a number of researchers have developed nonlinear counterparts to PCA. Kramer (*15*), for example, proposed an autoassociative neural network that was demonstrated to be capable of reducing the dimensionality of a simulated batch reactor. Dong and McAvoy (*16*) integrated principal curves with a neural network and showed how the resulting algorithm could be used to detect the drift of a process as it moved from one operating region to another.

Despite the successful application of nonlinear MSPC techniques, for batch processes it has been found that a more suitable approach is to remove the major nonlinearities from the data prior to applying the linear MSPC algorithms. The most common form of data transformation, termed multi-way PCA (MPCA), was initially proposed by Nomikos and MacGregor (*17*). Since then it has been adopted by several other researchers and applied to a variety of processes. For example, Gallagher et al. (*18*) applied the technique to monitor nuclear waste storage vessels and Gregersen and Jorgensen (*19*) investigated the detection of faults in a fed-batch fermentation process.

The concept of MPCA is a relatively straightforward extension to the approach taken for continuous systems, but deviations from mean trajectories rather than steady state are considered. In the approach *m* historical batches, referred to as *nominal* batches, are selected and used for comparison purposes. The duration of each of these nominal batches is likely to differ, and for reasons that will become apparent later the data from each batch is considered up until the shortest run length.

The next step is to identify the process variables, *n*, 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.

The individual data matrices from each batch are then unfolded into a single data matrix, as illustrated in Figure 4 (assuming 3 variables and *m* batches are used), t_0 and t_{end} refer to the batch start and end times and Var refers to Variable. Finally all columns of the unfolded data matrix are standardized to unit variance. PCA can now be applied to this unfolded data matrix in the same way as for continuous systems.

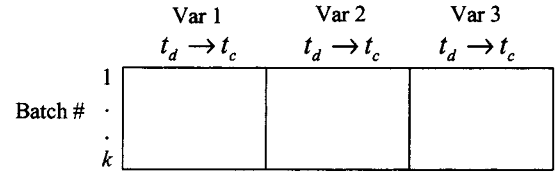

Figure 5. Moving Window PCA.

Despite the reported success of this approach, it has a significant weakness in that it can only be used on-line up until the shortest run length encountered in the nominal batches. This is of particular concern in the bioprocess industry where variations in batch length are common. Lakshminarayanan et al. (*12*) proposed that the data from the shorter run length batches could be extrapolated to simulate the future progression of the batch had it been allowed to continue. This does however make assumptions regarding the future profile of the batch that may be invalid. In this paper a technique referred to here as Moving Window PCA is introduced and applied. The advantage of this approach is that it is able to cope with variations in batch length seamlessly. The following section details the Moving Window PCA approach, and the application of the approach to an industrial fermentation system is then compared with the results obtained using MPCA.

Moving Window PCA. The concept of Moving Window PCA is very similar to that of MPCA. Initially, *m* historical batches are selected for comparison purposes. The mean trajectory for each variable over all the nominal batches is then calculated and removed from each process measurement, thus creating a matrix of scaled data.

At any given time during an on-line fermentation, a matrix is constructed, from the scaled data matrix, that contains data collected over a moving window of time, as displayed in Figure 5. In this figure t_{c} represents the current sample time and t_{d} the length of the moving window, i.e., information for each process variable is collected between sample times t_{d} and t_{c}. Notice that rather than containing *m* batches, the matrix contains data from *k* batches, where *k* refers to the number of batches for which data is available during the particular time window. For example if t_c refers to a time that is beyond the run length of a particular batch then this batch is simply removed from the collection of nominal batches. Similarly if for some reason there is data missing from a particular batch between sample numbers 40 and 60 then this batch can be used in the analysis at all periods except between these times.

Once the matrix has been constructed, all columns of the data matrix are standardized to unit variance. This matrix is referred to as the *unfolded moving window matrix*. A new unfolded moving window matrix is generated at each sampling instant, and a PCA model is calculated on it and then applied to data from the current batch.

This approach has several advantages over MPCA. The first is that no assumptions need to be made regarding the future progress of the batch (this issue is discussed further in the next section) and the second is that, unlike MPCA, moving window PCA copes seamlessly with variable run lengths.

There are, however, disadvantages with using the Moving Window PCA approach. There is a greater computational burden; as a new PCA model must be determined at each sampling instant, it is necessary to determine the length of the moving window and time delays must be considered.

The ability of the moving window PCA and MPCA approaches to monitor an industrial fermentation system are compared in the following section.

Application of PCA

To determine the benefits of monitoring a fermentation during the final production stage, an investigation was carried out for Biochemie Gmbh on their research plant in Austria. A MPCA model was developed using historical data collected from 10 high yield batches. The ability of this model to monitor a number of subsequent batches was then determined. Some of these subsequent batches operated smoothly, whereas others were affected by disturbances.

The MPCA model was developed using the entire unfolded data matrix. This poses a problem when the model is applied to a batch on-line as the full unfolded data matrix for this batch will not be known until the end of the run. This means that with the exception of the end point of the batch, it is necessary to estimate the future values of all measured variables. The prediction of future process values is normally referred to as *filling up* the matrix and can be achieved using three possible approaches:

(1) It is assumed that all future scaled process values remain at the nominal mean trajectory.

(2) It is assumed that all future process values remain at the current offset from the nominal mean trajectory.

(3) The PCA model itself is used to estimate the future values of the process variables using the approach detailed by Nelson et al. (*20*).

In this study it was found that the second method was the most suitable.

The purpose of the MPCA analysis is to identify process faults such as sensor failures and drifts as well as to monitor the quality of the bioproduct. Analysis of the information provided by MPCA can be achieved in a variety of ways. For example, the variation of an individual score can be viewed or a two-dimensional plot of one score against another. In this work it was found that monitoring the square prediction error (SPE) and T^2 value during the batch was a simple and informative approach to monitoring the batch. Goulding et al. (*21*) demonstrated that such plots are capable of identifying abnormal operating conditions caused by hardware faults and poor operating practices. Hardware faults tend to manifest themselves through an increase in SPE, whereas changes in operating procedures are more likely to be indicated on the T^2 chart. By imposing statistical confidence limits on these charts it is possible to create warning levels that, when exceeded, indicate an abnormality. The values of the limits are based upon the assumption that the SPE and T^2 values follow a Chi squared and Normal distribution, respectively. Further details of their calculation can be found in ref 22.

Figure 6 shows the SPE, along with 97% confidence limits, for a particular test batch. The confidence limits have been set at 97% because at this value the confidence limit is not exceeded by the nominal batches. A value lower than this may result in excessive false alarms, and a value higher may be too insensitive to identify abnormalities on the process. 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 6 demonstrates that the MPCA model has detected this fault entering the system. The T^2 chart failed to identify this fault, which supports the hypothesis earlier regarding the SPE and T^2 charts.

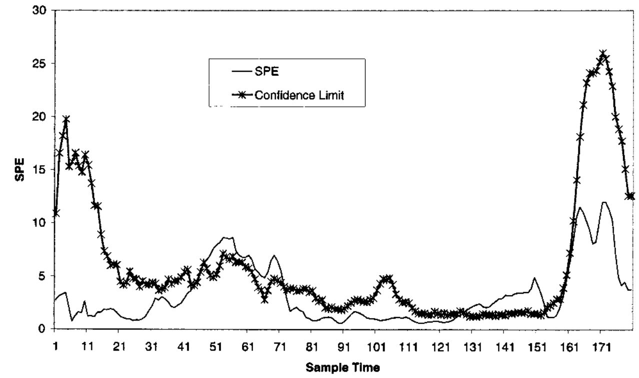

Figure 6. SPE chart for test batch.

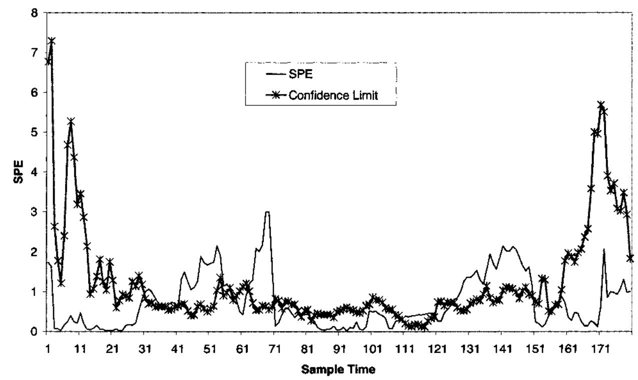

Figure 7. SPE chart for Moving Window PCA.

Figure 7 shows the SPE chart that was generated using the moving window PCA model. The moving window PCA model was developed using the same historical data as for the MPCA model, and the length of the moving window was chosen to be 5. It was found that varying the length of the window had limited impact upon the results.

The results displayed in Figures 6 and 7 are fairly representative of the results obtained in several tests. These results indicate that for this industrial fermentation system both techniques are capable of identifying the process fault. It is also evident from these figures that for the earlier fault, the moving window approach detects the fault more readily than MPCA. This is indicated by the fact that the SPE exceeds the confidence limits to a greater degree for the moving window approach than with MPCA. This may be expected as during the early stages of the run, MPCA is making significant assumptions regarding the future progress of the batch when filling up the matrix. Errors in these assumptions may mask any minor faults that may occur early in the batch.

Figure 8 demonstrates the ability of the moving window approach to cater for batches with varying run lengths. Initially the nominal data comprises measurements taken from 10 batches. At sample numbers 100, 120, 140, and 160 one, two, three, and then four batches are assumed to have completed and have been removed from the nominal data set. The data passed through the moving window PCA model this time is a batch that contained no abnormal behavior. The chart shows that as expected the SPE keeps below the confidence limit during the entire batch. For this example traditional approaches, such as MPCA, would mean that the on-line analysis would only continue until the shortest run length was reached, which in this case would be sample number 100.

The monitoring techniques detailed here have now

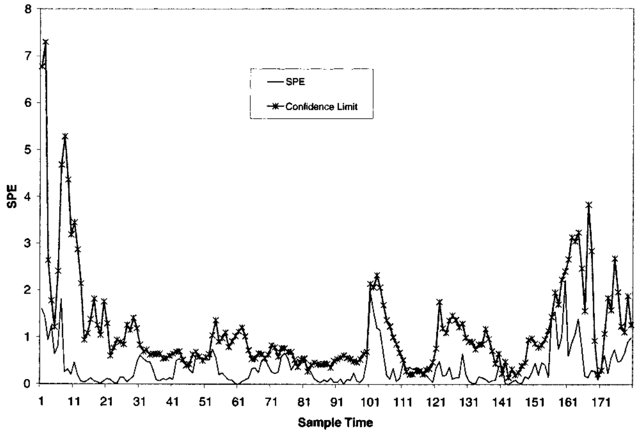

Figure 8. SPE chart for variable run lengths.

been consolidated in an on-line monitoring package that is currently undergoing further trials. The intention is that following successful on-line tests the approach will be applied to an industrial fermentation system operated by Biochemie Gmbh.

Conclusions

This article has provided details of an investigation into the use of AI and statistical techniques to aid the monitoring of two industrial fermentation processes. In both cases the benefits of these methods have been established. In the Synpac case study a tighter control of parameter A has been established through the expert system. In the first instance this will lead to improved consistency of operation. In the longer term it will be possible to lower the value of the parameter A set point, which will ultimately lead to enhanced productivity. Currently the system relies on the operators to perform the actions proposed by the system. To reap the full benefits an automated feedback system would be necessary with the operator acting as a supervisor of the loop. This will, however, require further validation of the system. The effort required in validation can only be justified once the company fully assesses the benefits of the system in its present form.

The second case study demonstrated that a sensitive fault indicator can be implemented using statistical data analysis procedures. Results presented indicate that with appropriate assumptions concise indicators of process faults can be presented to the operators to assist them in taking suitable corrective actions. The methods rely upon comparing the current batch behavior with historical profiles. It is therefore important to select representative historical batch data and appropriate process variables with due care.

The expert systems approach has been shown to solve the particular problem considered. Similarly, the algorithmic methods proved to be valuable in fault detection. For overall supervision purposes it is highly likely that a combination of expert systems and data-based analysis approaches will be required. Their integration within a knowledge base framework can potentially be relatively straightforward given appropriate software tools.

Acknowledgment

The authors would like to acknowledge the contributions of Gordon Miller and Brian Smith of Synpac Pharmaceuticals and Georg Kornfeld of Biochemie. Thanks go to both companies for giving permission to publish the results of the studies.

References and Notes

(1) Fowler, G. et al. Development of real-time expert systems approach for the on-line analysis of fermentation respiration data. *Proceedings of the 2nd IFAC Symposium on Modelling and Control of Biotechnical Processes*; Pergamon Press: Keystone, CO, 1992; pp 173−178.

(2) Cooney, C. L. et al. An expert system for intelligent supervisory control of fermentation processes. *Proceedings of the 8th International Biotechnology Symposium*; Société Francaise de Microbiologie: Paris, France, 1988; pp 563−575.

(3) Glassey, J.; Montague, G. A.; Mohan, P. Issues in industrial bioprocess advisory system development. *Trends Biotechnol,* **2000**, *18*(4), 136−141.

(4) Ross, D. et al. Gensym User Symposium, 1998.

(5) Leake, D. B. *Case Based Reasoning: Experiences, Lessons and Future Directions*; MIT Press: Cambridge, MA, 1996.

(6) Wetherill, G. B.; Brown, D. W. *Statistical Process Control: Theory and Practice*; Chapman & Hall: New York, 1991.

(7) Vander Wiel, S. A.; Tucker, W. T.; Faltin, F. W.; Doganaksoy, N. Algorithmic statistical process control: concepts and an application. *Technometrics* **1992**, *34*, 286−297.

(8) Hahn, G. J.; Cockrum, M. B. Adapting control charts to meet practical needs: a chemical processing application. *J. Appl. Statistics* **1987**, *14*, 35−52.

(9) Nomikos, P.; MacGregor, J. F. Multivariate SPC charts for monitoring batch processes. *Technometrics* **1995**, *37*(1), 41−59.

(10) Anderson, J. S.; Brisk, M. L. Estimating the economic benefits of advanced process control. *I.Chem.E. Symposium, Advances in Process Control III*; York, September, 1992; Institute of Chemical Engineers: New York.

(11) Lennox, B.; Montague, G. A.; Hiden, H. G.; Kornfeld, G. Case study investigating multivariate statistical techniques for fermentation supervision. *Comput. Chem. Eng.* **1999**, *23*S, 827−830.

(12) Lakshminarayanan, S.; Gudi, R. D.; Shah, S. L.; Nandakumar, K. Monitoring batch proceses using multivariate statistical tools: extensions and practical issues. *Proceeding of IFAC World Congress*; San Francisco, 1996; Elsevier Science Ltd.: Oxford; pp 241−246.

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

(14) Wold, S. Cross-validatory estimation of the number of components in factor and principal components models. *Technometrics* **1978**, *34*, 286−297.

(15) Kramer, M. A. Autoassociative neural networks. *Comput. Chem. Eng.* **1992**, *16*(4), 313−328.

(16) Dong, D.; McAvoy, T. J. Batch tracking via nonlinear principal component analysis. *AIChE J.* **1996**, *42*(8), 2199−2208.

(17) Nomikos, P.; MacGregor, J. F. Monitoring batch processes using multi-way principal component analysis. *AIChE J.* **1994**, *40*(8), 1361−1375.

(18) Gallagher, N. B.; Wise, B. M.; Stewart, C. W. Application of multiway principal components analysis to nuclear waste storage tank monitoring. *Comput. Chem. Eng.* **1996**, *20*S, 739−744.

(19) Gregersen, L.; Jorgensen, S. B. Supervision of fed-batch fermentations. *Chem. Eng. J.* **1999**, *75*, 69−76.

(20) Nelson, P. R. C.; Taylor, P. A.; MacGregor, J. F. Missing data methods in PCA and PLS: Score calculations with incomplete observations. *Chemom. Intell. Lab. Syst.* **1996**, *35*, 45−65.

(21) Goulding, P. R.; Lennox, B.; Sandoz, D. J.; Smith, K.; Marjanovic, O. Fault detection in continuous processes using multivariate statistical methods. *Int. J. Syst. Sci.* **2000**, *31*(11), 1459−1471.

(22) Lennox, B.; Montague, G. A.; Hiden, H. G.; Kornfeld, G.; Goulding, P. R. Process monitoring of an industrial fed-batch fermentation. *Biotechnol. Bioeng.* **2001**, *74*(2), 125−135.

Accepted for publication December 11, 2001.

BP0101839