Presented at the World Batch Forum European Conference Mechelen, Belgium 11-13 October 2004	WORLD BATCH FORUM	900 Fox Valley Drive, Suite 204 Longwood, FL 32779-2552 +1.407.774.5764 Fax: +1.407.774.6751 E-mail: info@wbf.org www.wbf.org

Statistical Process Monitoring of Industrial Batch Processes

O. Marjanovic and B. Lennox
School of Engineering
University of Manchester,
United Kingdom
Phone:
Email: barry.lennox@man.ac.uk

D. Sandoz and D. Lovett
Perceptive Engineering Limited
Cheshire,
United Kingdom
Phone:
Email:

KEY WORDS

ABSTRACT

The manufacture of high-value products involves many different batch processes, for example industrial fermenters. Such processes require high levels of consistency in their operation to ensure minimal losses of raw materials, utilities and product. Recent application studies have indicated that multivariate statistical technology can provide some support when trying to maintain consistent operation in complex batch processes. This paper summarizes the findings from three case studies involving the application of multivariate statistics to batch processes. Two of the studies are taken from the fermentation industry with the third study involving a comprehensive penicillin production simulation model. The main focus of this paper is to demonstrate how realistic upsets in process operation can be detected and graphically presented using statistical process monitoring technologies. Further to this the paper provides a comparison of different approaches to monitoring batch process operations.

1. Introduction

Copyright © 2004 World Batch Forum. All rights reserved.

Biochemical fermentation systems are highly sensitive to abnormal changes in operating conditions with product quality and yield typically being very much dependent upon the consistency of process operation. Since such processes are used in the manufacturing of high-quality products and as a result, it is particularly important to detect incipient degradation of batch performance so that the effects of any abnormal condition does not have a detrimental effect upon product quality. Furthermore, the pharmaceutical industry together with the food and beverage industry are obliged to comply with increasingly stringent regulatory requirements enforced by agencies such as the FDA. For many compounds, these agencies demand proof that consistent operation is adhered to and without this proof the product cannot be sold.

Recent application studies have indicated that multivariate statistical technology can provide some support when trying to maintain consistent operation in complex batch processes. Multivariate statistical technologies capture complex interactions between process variables during satisfactory (high yield or low cost) process operation and extract these inter- variable relationships in the form of statistical models. These statistical models can then be used in real- time in order to benchmark satisfactory performance of a process and, thereupon, continuously improve reliability and profitability of a manufacturing plant. The ability of these models to predict the behaviour of important process variables also makes them suitable for integration within advanced process control applications and for the construction of sensor validation schemes and inference engines for difficult-to- measure quality variables. Finally, statistical models can often act as an aid to process understanding.

Optimisation of process operation through the use of advanced process monitoring technologies is usually seen as a natural choice for reducing production costs, improving product quality as well as meeting safety requirements and environmental regulations. However, the task of achieving optimal performance of industrial batch processes represents a very difficult challenge to production and systems engineers. Highly non-linear dynamics, large batch- to- batch variations, in addition to difficulties in measuring many quality variables means that maintaining consistent operation of such systems can be problematic. Nevertheless, successful applications of multivariate statistical batch process monitoring have been reported in literature (Kourti et al. 1995; Lennox et al. 2000).

This paper presents results from three case studies in which both traditional univariate and advanced multivariate statistical analysis have been applied to batch processes taken from the biochemical industry. Two of these studies have been conducted using industrial data while the third was performed using a simulation of a penicillin production facility. The main focus of this paper is to demonstrate how realistic upsets in the operation of a batch process can be detected and graphically presented to operators and engineers using both univariate and multivariate statistical process monitoring technologies. Also, a comparison between these different monitoring schemes is conducted in terms of their ability to detect the existence of faults and abnormal conditions and to diagnose the root cause of such problems.

2. Statistical Process Monitoring Technologies

2.1 Univariate Analysis

Univariate batch process analysis considers all the process variables to be independent of each other, analysing them one by one. The mean and standard deviation of the trajectory that each variable follows for a set of satisfactory batches are determined and these statistics are then used to establish quality control limits, describing an envelope of satisfactory operation for each recorded process variable. Consistent violation of these limits during a batch progression would then indicate that the conditions of the current batch are inconsistent with what is expected for satisfactory performance, suggesting that the batch may be of poor quality. A drawback with this approach is that it ignores

Copyright © 2004 World Batch Forum. All rights reserved.

any relationships that may exist between process variables. Therefore its applicability is limited to more severe types of faults which may obly be detected when the relationships between variables is considered. Also, many variables may be recorded, necessitating the need for multiple charts to be monitored, which can be impractical.

2.2 Multivariate Statistical Process Monitoring

Multivariate statistical analysis captures relationships that exist between different process variables and condenses this information into a small number of important metrics. This analysis represents a more comprehensive attempt, when compared to univariate analysis, to capture the nominal operation of the process in the form of a statistical model.

Multivariate statistics relies heavily upon the statistical routines referred to as Principal Component Analysis (PCA) and Partial Least Squares (PLS). Principal Component Analysis is generally used to develop a statistical model representing satisfactory process operation. As such a PCA model identifies the inter- variable relationships that exist during satisfactory process operation. PCA is then able to extract the main features of process operation which can be extracted and stored in a small number of composite variables, commonly referred to as scores. These composite variables can then be easily monitored in real- time in order to benchmark process performance and highlight potential problems, leading to continuous improvement of the process operation.

PLS is a regression tool that is able to identify cause to effect relationships in process systems that contain many highly correlated variables. Such relationships can be difficult to identify using more traditional regression tools, such as multiple linear regression. PLS, like PCA, can be used to identify statistical models that capture satisfactory operation of a process. These models can then be used in process monitoring applications. Furthermore, through their cause- effect structure and predictive powers, PLS- based models can also be used for advanced process control and inference engine developments.

One of the most frequently used metrics in multivariate statistical process monitoring is the composite prediction error for all of the monitored process variables, termed Squared Prediction Error (SPE). This metric provides a single measure of process deviation from the statistical model. Hence, the SPE metric is expected to have high values during abnormal operation and low values when the process operates in a satisfactory manner. Another frequently used metric is the T^2 statistic. This metric is composed of all of the scores, which have been extracted during PCA or PLS analysis and represents the main features of the process. Hence, the T^2 metric provides a measure of how far away the current operating conditions of the plant are from the conditions present in the data that was used for statistical model development.

As a new batch progresses predictions of each variable are obtained from the PCA or PLS model and are compared with the actual measurement values. Hence, in addition to the correct detection of the under performing batches, mainly through the use of SPE and T^2 quality control charts, the multivariate statistical models are able to describe precisely how each variable deviates from its expected behaviour, i.e. its model prediction. This predictive feature of PCA and PLS based models is crucial during identification of the root cause of the fault, commonly known as fault isolation or diagnosis. This information could be exploited so that process operators are able to make corrective action in the early stages of the batch progression, i.e. before the abnormal conditions have a significant impact upon the batch performance.

Copyright © 2004 World Batch Forum. All rights reserved.

2.3 Transformations of Batch Process Data

The data collected from a single batch tends to come in a standard matrix form with each column representing a particular recorded variable and each row corresponding to a particular sampling instant during local batch time. In order to consolidate data sets from a number of batches into a single standard data matrix two alternative transformation approaches can be applied, shown in figure 1 and figure 2. The first of these two methods is shown graphically in figure 1 and is referred to in this paper as the unfolding approach. The second method is shown graphically in figure 2 and will be referred to in this paper as the concatenated approach.

In figures 1 and 2 the data from three different batches is shown on the left hand side. For each batch there are three process variables (V1, V2 and V3) recorded at different instances of local batch time. In the case of the unfolding method, demonstrated in figure 1 and originally proposed by Nomikos and MacGregor (1995), each observation in consolidated in to a data set corresponding to all of the process variables' measurements collected during the progression of a single batch.

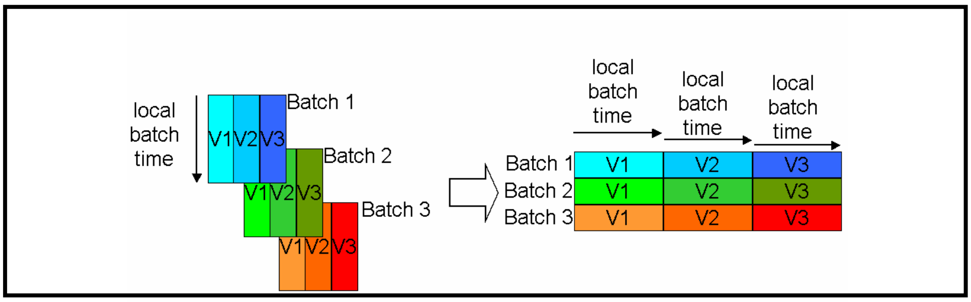

Figure 1: Unfolding approach

On the other hand, concatenated approach, demonstrated in figure 2 and originally proposed by Wold et al. (1998), consolidates the data by simply joining the end point of one batch, e.g. batch 2, with the starting point of another batch, e.g. batch 3.

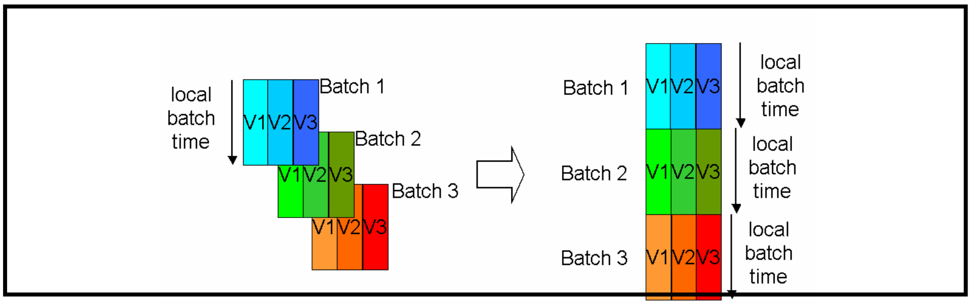

Figure 2: Concatenated approach

The unfolding method is by far the more popular one in academic community, described by the leading experts in the field as the most meaningful technique for batch analysis and monitoring (Nomikos and MacGregor 1995; Westerhuis et al. 1999). There are however a number of practical problems that limit the applicability of this method. In particular, the approach is not well suited for processes with high variability in batch lengths, which are occur frequently in the process industries. On the other hand, this is not an issue with concatenated approach, at least not during the model identification. Also, during the progression of a batch, statistical models that are based on the unfolding approach require the estimation of the future evolution of the batch. Such estimation is

Copyright © 2004 World Batch Forum. All rights reserved.

generally made using crude models, described by Nomikos and MacGregor (1995). This problem does not arise when using models that are based on the concatenated approach. However, in the case of the unfolding approach, the mean trajectories of all of the monitored process variables are extracted during the pre- processing of the data. Such processing is not the case with concatenated approach. This implicit feature of the unfolding approach removes the main source of non- linearity in the data, allowing accurate and yet linear statistical models to be developed.

3. Case Studies

Three case studies, detailed below, were conducted in order to assess the capabilities of different statistical process monitoring approaches in terms of fault detection as well as isolation of the root cause of abnormality. Due to the space limitations and for the reasons of confidentiality, only the summaries of these studies are reported here. The results displayed in this section were obtained using a commercially available software product.

3.1 Fermentation Process No. 1

The first case study was conducted on the data set from an industrial fermentation process. A single failed batch was used to test the capabilities of univariate as well as the multivariate approaches. The reason for the failure of this batch was diagnosed by operations staff as a drift in an important sensor. Analysis of this problem using univariate analysis failed to identify this fault. Hence, the fault with this batch could not be detected by analysing the the behaviour of the failed sensor alone. The univariate chart for the failed sensor measurement, during this particular batch is displayed in figure 3. In this figure the black lines represent the upper and lower confidence limits, together with the mean trajectory for this variable. The blue trend in this figure displays the progression of the failed sensor measurements during the failed batch. Figure 3 indicates that there are no violations of the control limit during the progression of this batch.

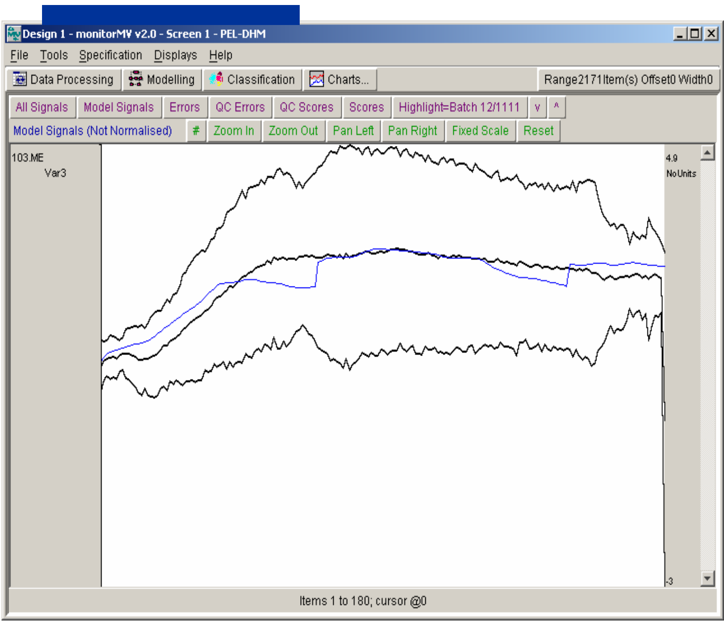

Figure 3: Univariate chart

Copyright © 2004 World Batch Forum. All rights reserved.

Figure 4: Multivariate chart

The multivariate statistical models, based on both the unfolding and concatenated approach, detected the existence of the fault. The SPE chart for PCA model based on the unfolding approach is shown in figure 4. The top trend in this figure refers to the actual SPE which is coloured according to whether its trend violates 99% confidence limit (red coloured), 95% confidence limit (brown coloured) or neither of these limits (blue coloured). As seen in Figure 4, it is clear that SPE trend remains red for large portion of the failed batch progression indicating detection of abnormal behaviour, i.e. significant deviation from its predicted behaviour.

In this example, careful consideration of the process had to made for the concatenated approach to detect this fault. For example, unlike the unfolded PCA approach, the concatenated PCA model failed to identify any abnormality. Concatenated PLS models that were developed to estimate the failed sensor measurement however were able to identify the fault. In the case of the unfolding approach, the detection of the fault was much clearer when the model was developed using a reduced set of process variables, identified as important parameters by process engineers, rather than all of the available variables.

3.2 Fermentation Process No. 2

The second case study was conducted using data obtained from a yeast production facility. Two unsatisfactory batches were used for the testing of the different monitoring schemes.

The results from the univariate analysis show that this approach correctly identified abnormalities associated with both failed batches. Hence, in this case study the faults were considered to be more obvious than in the previous study.

Both the concatenated and unfolding approach delivered satisfactory results in this particular case study. However, due to the predictive behaviour of these approaches, these multivariate models also provided additional insight into the possible reasons for the abnormality, which was not possible using univariate techniques. Figure 5 shows the trends of each process variable (coloured blue, magenta and green) for one of the failed batches with their model predictions superimposed (brown trend). The predictions, displayed in figure 5, were computed using a concatenated PCA model. For this particular batch the fault was attributed to the production of ethanol instead of yeast towards the end of the batch. However, figure 5 indicates that a significant upset occurred in the process at the very beginning of the batch, this is seen by the large difference in predicted and actual measurements at the left hand side of the figure. In particular, the pH measurement is seen to go through a noticeable transient at the very beginning of this batch, which was not highlighted using univariate

Copyright © 2004 World Batch Forum. All rights reserved.

analysis. Note that the disturbance in terms of molasses and NH3, half- way through the batch, was equally detectable by both univariate analysis and multivariate methods.

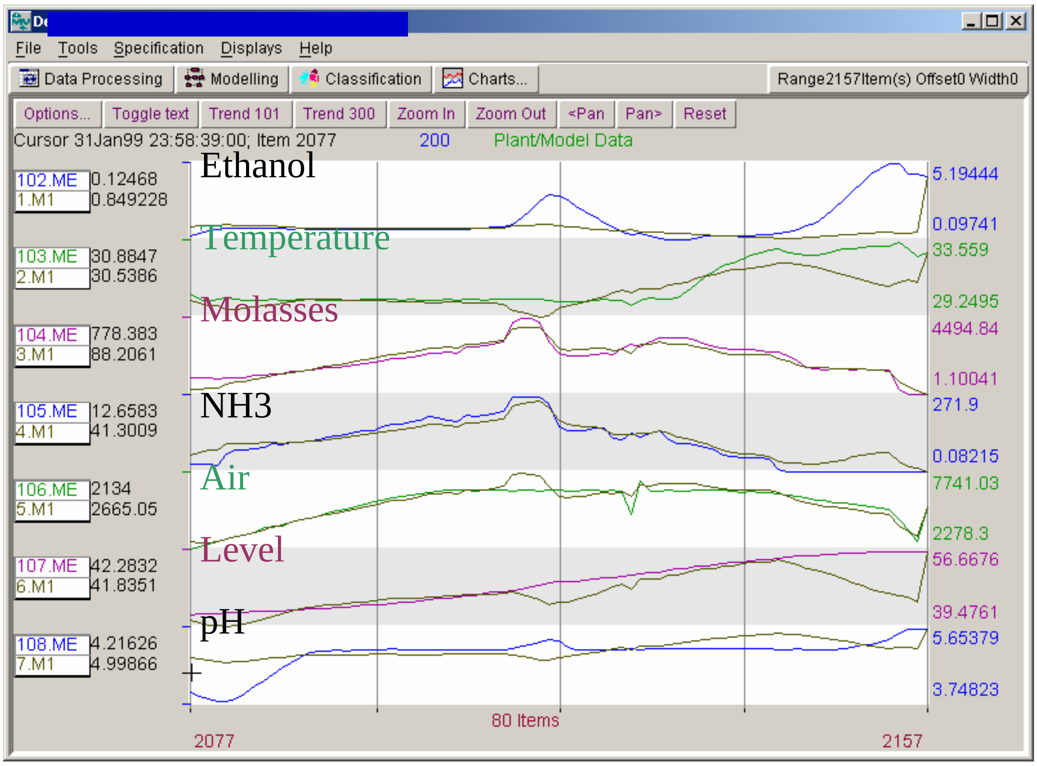

Figure 5: Multivariate predictions

In summary, detection and diagnosis of the faults through the use of univariate quality control charts was found to be sufficient tool for identification and isolation of the root causes of the faults in this particular case study. However, multivariate analysis methods provided additional insight into the exact type of process deviation in terms of each process variable by employing its prediction capabilities.

3.3 Simulation of Penicillin Production

The final case study that is reported in this paper is concerned with the condition monitoring of the simulated pencilin production process documented in Birol et al. (2002). The simulation was performed in MATLAB and the faults were injected as offsets of varying magnitude to the dissolved oxygen (DO_2) measurements. By varying the magnitude of the faults the relative sensitivity of the different approach can be analysed

Results indicate that in this case study multivariate models based on concatenated approach performed equally well as the univariate analysis. For a positive or negative offset of 1.5 units it was observed that both univariate analysis and PCA/PLS models generated using concatenated approach provided detection of the fault and the diagnosis of the root cause of the abnormality by the same level of clarity and consistency. In contrast, the multivariate models developed using the unfolding approach were found to be more sensitive to this particular type of fault than univariate quality control charts or the multivariate models based on concatenated approach.

4. Conclusions

Copyright © 2004 World Batch Forum. All rights reserved.

This paper details three case studies that demonstrate the applicability of statistical process monitoring technologies for batch processes. The case studies were conducted using real process data from industrial fermenters as well as a comprehensive simulation of a penicillin production vessel. Both univariate and multivariate analysis were performed. In terms of multivariate analysis two alternative methods of consolidating data from multiple batches into a single matrix format were applied and compared with each other.

The results from this study indicate that the multivariate techniques were consistently able to detect and isolate the cause of process abnormalities. This was not the case for the univariate approaches which failed to detect quite subtle faults. Comparisons between the concatenated and unfolding multivariate approaches indicate that in terms of their ability to detect and isolate the cause of process abnormalities, there is very little difference between the approaches. Whilst the unfolding approach appears to be more sensitive to abnormal conditions, it has the disadvantage that it does not cope well when the length of batch cycles vary. The concatenated approach does not suffer from this problem.

REFERENCES

MonitorMV, more info available at www.perceptive-engineering.co.uk.

Birol, G., C. Undey and A. Cinar (2002). A modular simulation package for fed- batch fermentation: penicillin production. Computers and Chemical Engineering, Vol. 26, pp. 1553- 1565.

Kourti, T., P. Nomikos and J. F. MacGregor (1995). Analysis, monitoring and fault diagnosis of batch processes using multiblock and multiway PLS. Journal of Process Control, Vol. 5(4), pp. 277- 284.

Lennox, B., H. G. Hidden, G. A. Montague, G. Kornfeld and P. R. Goulding (2000). Application of multivariate statistical process control to batch operations. Computers and Chemical Engineering. Vol. 24, pp. 291- 296.

Nomikos, P. and J. F. MacGregor (1995). Multivariate SPC charts for monitoring batch processes. Technometrics, Vol. 37(1), pp. 41- 59.

Westerhuis, J. A., T. Kourti and J. F. MacGregor (1999). Comparing alternative approaches for multivariate statistical analysis of batch process data. Journal of Chemometrics, Vol. 13, pp. 397- 413.

Wold, S., N. Kettaneh, H. Friden, A. Holmberg (1998). Modelling and diagnostics of batch processes and analogous kinetic experiments. Chemometrics and Intelligent Laboratory, Vol. 44, pp. 331- 340.

Copyright © 2004 World Batch Forum. All rights reserved.