10th International IFAC Symposium on
Computer Applications in Biotechnology
Preprints Vol.1, June 4-6, 2007, Cancún, Mexico

MULTISCALE FAULT DETECTION AND DIAGNOSIS IN FED-BATCH FERMENTATION

Ahmed Alawi[1] and Julian Morris[2]

[1]Genzyme Ltd., Hollands Road. Haverhill, Suffolk. CB9 8PU UK
[2]Centre for Process Analytics and Control Technology
School of Chemical Engineering and Advanced Materials
University of Newcastle, Newcastle Upon Tyne, NE1 7RU, UK

Abstract: The importance of the FDA PAT guidelines in pharmaceutical process design space can be influenced by the introduction of robust process malfunction and senor fault detection and diagnosis tools. The paper compares a multi-scale multi-block modelling approach with conventional multiway PCA approaches for batch process monitoring. A benchmark penicillin fermentation simulation is used to evaluate the two methodologies. Contributions plots with confidence bounds enhance the fault diagnosis potential of the approaches studied. The methodology is in the process of being evaluated in fermentation and batch cooling crystallisation.

Keywords: Monitoring; Wavelets; Multiblock methods; Fault detection and diagnosis.

1. INTRODUCTION

This work uses a combination of multiblock statistical modelling approaches together with multiscale wavelet decomposition. A number of multiblock algorithms have been proposed and investigated. Several algorithms have been proposed to deal with multiple data block including Hierarchical PCA (HPCA), Consensus PCA (CPCA) introduced by Wold et al., (1987), Hierarchical PLS (HPLS) and Multiblock PLS (MBPLS), e.g. Westerhuis et al., 1998; Qin et al., 2001; Smilde et al., 2003; Choi and Lee, 2004. Kourti et al., (1998) presented a theoretical review of MPCA and MPLS and compared the modelling capabilities of HPLS and MPLS on data from a multizone low-density polyethylene tubular reactor. Other applications have been to chemical process modelling and performance monitoring (e.g. Choi et al., 2004; Wong et al., 2005). It is suggested that as the FDA PAT guidelines are implemented the need for multiblock approaches will become even more important given the drive for deeper process understanding its implications for understanding the design space and the need to integrate properly large amounts of data collected using different sensors and analytical instruments including the integration of spectroscopic and process data. For example, Bra's et al., (2004) evaluated the potential benefits of combining NIR and MIR spectra for the prediction of protein, moisture, fat and fibre content of soybean flour using PLS models with those obtained from MPLS.

Chemical and biological processes typically exhibit dynamic behaviour with the measurements exhibiting autocorrelation (dynamics). Current multiblock approaches do not take into account such autocorrelated data structures, being based on the assumption that the process variables are stationary and normally distributed. In addition, process measurements are known to exhibit multi-scale behaviour as a consequence of representing the cumulative effect of a number of underlying process phenomena including process dynamics, measurement noise and disturbances. Thus a methodology is required to address (i) the multiscale nature of batch process data, and (ii) the inability of the existing algorithms to handle auto-correlation. It has been shown that frequency domain can be utilized to improve PCA based process monitoring and diagnosis (e.g. Bakshi, 1998; Misra et al., 2002). Multiresolution analysis based on wavelet decomposition can take into account the autocorrelated nature of the process data and that the process data can be decomposed into different scales. Coefficients from each decomposed scale are then used as the basis for statistical process monitoring taking into account process dynamics in an indirect way. In continuous processing examples have been presented by Bakshi, 1998; Misra *et al.*, 2000; Teppola & Minkkinen, 2000; Rosen and Lennox, 2001; Yoon and MacGregor, 2004; and Alawi *et al.*, 2005. Bakshi's work laid the groundwork for exploiting the benefits of multi-scale modelling and eliminating the risk of losing of useful information or changing the multivariate structure by pre-processing data through multi-scale PCA (MSPCA). MSPCA combines the ability of PCA to decorrelate variables by extracting a linear relationship with that of wavelets to extract deterministic features and approximately decorrelate auto-correlated measurement at each scale. The Multiscale nature of the MSPCA formulation makes it suitable to work process data that are typically non-stationary and represent a cumulative effect of many underlying process phenomena, each operating at different scales. In this paper, a multiscale-PCA approach is proposed (figure 1) for process monitoring and fault detection of batch processes. At every time point, the batch process variables are decomposed into scales to the wavelet domain and then reconstructed back to the time domain. The scales/details and the approximations are collected into separate matrices (block). Multiblock PCA is then applied to the wavelets details and approximation. Fault detection based on the total T_s^2 and Q_s statistics can then be used along with contribution plots incorporating confidence bounds to enhance fault diagnosis.

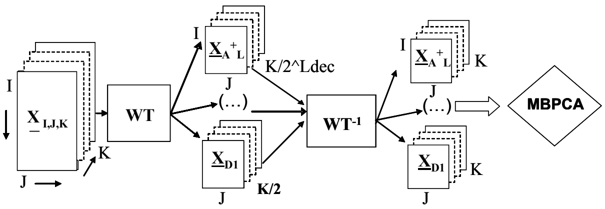

Fig. 1: Multiscale batch monitoring scheme

2. PROCESS MONITORING BASED ON MBPCA

Model building commences by decomposing the batch block data sets into different scales through the application of wavelets. The batch data matrix is arranged in tri-linear form $\underline{\mathbf{X}} \in \Re^{I \times K \times J}$ where K is the batch duration, J is the number of variables and I is the number of batches. First, a nominal operations model is constructed form historical data.

2.1 The nominal operating condition (NOC) model

1) I historical batches are selected which represent nominal operations and unfolded to give $\underline{\mathbf{X}}\,(I \times J \times K)$ to $\mathbf{X}_j(I \times K)$, where J is the number of variables. The matrix $\mathbf{X}_j$ contains each variable for all batches at all time point with each ith row, $i = 1, \cdots, I$ of the individual variable matrix $\mathbf{X}_j$ being decomposed by applying the discrete wavelet transform (DWT), $\mathbf{W}(K \times K)$ with decomposition level L above. It is noted that the same wavelet transform with the same level, L is applied to each of the $j = 1, \cdots J$ variables. The wavelet reconstruction detail functions from $\mathbf{X}_{D_1^i}(I \times K)$ to $\mathbf{X}_{D_L^J}(I \times K)$ and approximation $\mathbf{X}_{A_L^J}(I \times K)$ are collected for each batch. The objective is to extract the correlation within the batch variables across the batch duration (c.f. Misra *et al.*, 2002).

2) Refold the wavelets detail transformation matrices $\mathbf{X}_{D_1^i}(I \times K)$ to $\mathbf{X}_{D_L^J}(I \times K)$ and approximation $\mathbf{X}_{A_L^J}(I \times K)$ to be of the form $\mathbf{X}_{D_L}(I \times KJ)$ and $\mathbf{X}_{A_L^*}(I \times KJ)$.

3) The unfolded wavelet transformation matrices (blocks), $\mathbf{X}_{D_1}^{I \times KJ} \cdots \mathbf{X}_{D_L}^{I \times KJ}$ and $\mathbf{X}_{A_L^*}^{I \times KJ}$ are normalized using the mean and the standard deviation of each reconstructed wavelets variables at each time in the batch cycle over all batches. Apply the multiblock PCA algorithms to the block matrices, $\mathbf{X}_B = \lfloor \mathbf{X}_{D_1} \cdots \mathbf{X}_{D_L}\ \mathbf{X}_{A_L} \rfloor$.

2.2. On-line monitoring

For on-line implementation, there are a number of approaches that can be used. The approach adopted requires that the measured data is decomposed within a window of dyadic length and only the last data point of the reconstructed signal is retained for on-line monitoring. When the measured data becomes available, the window is shifted to include the most recent measurement whilst maintaining the same window length. Nounou and Bakshi (1999) used a similar scheme for on-line data rectification. An alternative approach is to use an autoregressive model or PCA model to estimate the unknown future variables. That is, a separate MPCA model is built on the original data using this model to estimate the rest of the batch using a missing data approach. Once the variables are estimated for the reminder of the batch, the wavelet transform is applied.

Yoon and MacGregor (2004) discussed the effect of scaling before and after wavelet decomposition for continuous processes. Their work concluded that scaling after the application of wavelets will place greater weight on the faults occurring at high frequency whereas scaling before the application of wavelets places greater emphasis on faults occurring at low frequency. In this work scaling was performed after the application of wavelet decomposition.

1) For new batch data, a window of samples is recorded up to time point k such that the size of the window is $J * 2^{L^w}$, where L^w is the number of decompositions.
2) The resultant blocks of wavelet transformations are refolded batch-wise to obtain $\mathbf{x}_{D_L^{new}}(1 \times kJ)$ and $\mathbf{x}_{A_L^{*new}}(1 \times kJ)$, and then the missing values are in-filled up to $\mathbf{X}_{D_L^{new}}(1 \times KJ)$ and $\mathbf{X}_{A_L^{*new}}(1 \times KJ)$.

The block of wavelet transformations are then scaled using the same mean and variance obtained from nominal operating conditions.

3) The block of wavelets are projected on the CPCA model and the super, block scores and residuals are calculated: $\mathbf{t}_S^{new}(1 \times A) = \mathbf{t}_B^{new}\mathbf{P}_T$, $\mathbf{t}_b(1 \times A) = \mathbf{X}_b^{new}\mathbf{P}_b$, $e_b^{new}(JK \times 1) = \mathbf{X}_b^{new} - \mathbf{P}_b\mathbf{t}_b$
4) Determine whether the super-block or block Hotelling's T^2 or Q-statistic exceed the confidence bounds.

It is noted that there are two issues with regard to the application of multiresolution analysis using wavelets for batch processes: (i) identification of the optimal decomposition level, L^w and (ii) the on-line implementation. In the decomposition of the batch process variables, identifying the optimal wavelet decomposition level L^w is of particular importance. In the multivariate case, each variable may have a different optimal decomposition level. However, in most practical applications, only a single decomposition level will be applied to all variables for computational simplification (*c.f.* Bakshi, 1998; Wang and Romagnoli, 2005). Four monitoring statistics and their control limits were used (Table 1), where I is the number of batches, A bias the number of principal components, s is the sample variance of Q_s, μ is the sample mean of Q_s, S_b and μ_b are the sample variance and mean for Q_b respectively. The optimal decomposition level is selected such that the underlying features of each variable are adequately preserved in the approximation function with minimum noise. Maulud *et al.* (2005) used PCA to identify 'the optimal' number of levels by examining the noise content of the PCA model residuals of the signal approximation at level L^w. In this work the level of decomposition is identified using the procedure proposed by Maulud *et al.,* (2005). More specifically, a PCA model is applied to the approximation reconstruction function. Recalling that the residual of the PCA relationship consists mainly of noise, as the wavelet decomposition is recursively applied (starting with $l^w = L^w$), the magnitude of the residual is reduced as more noise is captured by the detail functions. However, the A retained principal components remain more or less constant as the underlying features of the signal are adequately preserved (Maulud *et al.* 2005). As the decomposition level l^w increases, some of the underlying features of the signal in the approximation function start to be lost to the detail function and the number of retained principal components, A, start to change significantly. This change can be detected by observing the explained variance of the first principal component of the approximation reconstruction function for different decomposition levels.

Table 1 Summary of the monitoring statistics and their control limits

	Statistic		Control limits
Total	T_S^2	$\mathbf{t}_S^T\Lambda_S^{-1}\mathbf{t}_s$	$A(I^2-1)/N(I-A) .. F_{A,m-A,\alpha}$
	Q_s	$\|e\|^2$	$(s/2\mu)\chi^2_{2\mu^2/s,\alpha}$
Block	T_b^2	$\mathbf{t}_b^T\Lambda_b^{-1}\mathbf{t}_b$	$A(I^2-1)/N(I-A) .. F_{A,m-A,\alpha}$
	Q_b	$\|e_b\|^2$	$(s_b/2\mu_b)\chi^2_{2\mu_b^2/s,\alpha}$

3. FED-BATCH PENICILLIN APPLICATION

In the study presented, fermentation data are generated using a detailed mathematical model and a simulator PenSim (Birol *et al*, 2002)[1, 2]. In the final paper results will be presented from an industrial scale fed batch fermentation. The model has five input variables, nine process variables and five quality variables. The variables can be categorized as follows - load variables including aeration rate, agitator power, and substrate feed rate and substrate feed temperature; manipulated variables includong acid/base flow rates and heating/cooling water flow rates; internal state variables such as culture volume, generated heat, carbon dioxide, dissolved oxygen, biomass, penicillin and substrate feed concentrations; and controlled variables such as pH and bioreactor temperature. A nominal operating data set from 50 batches is generated with the process inputs small perturbations added to the inputs to mimic the variations in normal operating conditions encountered in the real process. In addition, measurement noise was added to the 14 monitored variables. All batches are assumed to have the same

[1] Birol, G., Undey, C., Cinar, A., (2002). "A Modular Simulation Package for Fed-Batch Fermentation: Penicillin Production". Comp. Chem. Eng., 26, 1553-1565.

[2] http://www.chee.iit.edu~control/software.html

length. The batch duration is 400 hrs comprises of two stages, pre-culture stage for about 45 hrs and a fed-batch stage of 355 hrs. The sampling period is two hourly. Three principal components were selected for both the multiscale and the conventional MPCA modelling using cross-validation.

In the MSPCA methodology presented here, data collected during good operations (e.g. high yield batches) are decomposed into wavelet coefficients for the MSPCA model. The wavelet function and the decomposition level needs to be determined properly and needs to be consistent throughout the model building and application process. The choice of the wavelet is subject to many influential factors depending on process measurements and characteristics. Although the separation between deterministic and stochastic depends on the wavelet chosen; here the focus is on wavelet applications rather than the detailed characteristics. Selecting the decomposition level is based on frequency bandwidth and increasing the decomposition level gives better separation of high-frequency (noise) signals. In this study, the decomposition level was selected to be 2 using the method proposed by Maulud *et al*. (2005) and the data decomposed using the MATLAB Wavelet Toolbox. Two sets of experiments were conducted. In the first set, sensor faults (sensor drift and sensor degradation) were introduced and investigated; the second set examined process faults (process drift and process bias).

In the fermentation studied, pH and bioreactor temperature were used for control purposes with other sensors used for process monitoring. Typically, four types of sensor faults are observed: complete failure, bias, drifting and degradation. Two sets of experiments were conducted. In the first set, sensor faults (sensor drift and sensor degradation) were introduced, whilst the second set consisted of process faults (process drift and process bias). The objective is the earliest possible detection and identification of abnormal conditions. A CPCA model based on the reconstruction approximation and detail blocks was then built as well as a conventional multi-way PCA model. Three principal components were selected for both the proposed multiscale method and the conventional MPCA using cross-validation.

3.1 Sensor drift

Space allows only a brief description of a dissolved oxygen (%DO) probe drift and a drift in the fermenter substrate feed rate. The impact of the probe drift is confined to the low frequency bandwidth. The drifting sensor was simulated to start at time 65 an eventual fault magnitude of ϖ. Figure 2 shows five different %DO2 drift fault magnitudes all starting at time point 65. The magnitude of this fault is confined to the low frequency bandwidth (approximation). This change in signal variance is difficult to detect because it is often masked by noise and other events.

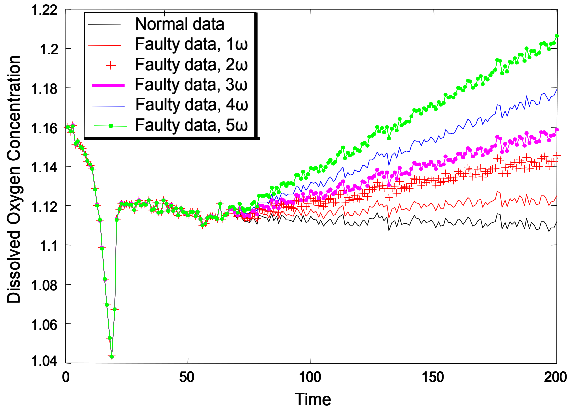

Fig. 2: Time series plots of %DO2 for different magnitudes of drift from time 65

The study compared the proposed MSPCA and standard MPCA for 5 different drift fault magnitudes and which are summarised in Figure 3. It can be observed that such faults can be detected much faster using the proposed multiscale approach. MPCA is generally better for capturing faults that contain contributions with the same localization everywhere in the time-frequency domain.

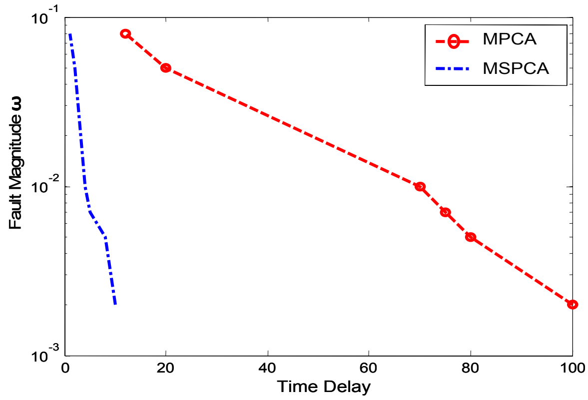

Fig. 3: Comparison of fault magnitude ω against fault detection time delay (sensor drift)

Figure 4 shows the results for a particular drift fault of magnitude 5ω. The fault was detected at time point 65 in both the MSPCA T_s^2 and the Q_s statistics in contrast to standard multiway PCA where it is only detected in the Q-statistic at time point 120. To identify the fault, a hierarchical contribution plot is first investigated to identify the block which significantly contributes to the fault. In this case it can be seen from Figure 5 that the approximation block significantly contributes to the fault. Drilling further down into the monitoring statistics for the individual blocks, figure 6, indicates that both blocks detect the fault with the approximation block significantly contributing to the fault. In particular, the approximation block can be further interrogated to identify the root cause of the fault. The approach used here is through contribution plots, where the contribution of each variable to Hotelling's T^2 is calculated in contrast to the convention contributions approach of examining the contribution of the individual scores.

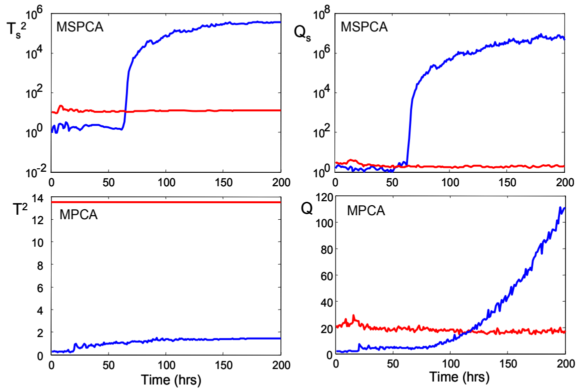

Fig. 4: Monitoring charts for MSPCA (upper plots) for Super T_s^2 and super Q_s-statistic; MPCA (lower plots) T^2 and Q – statistics

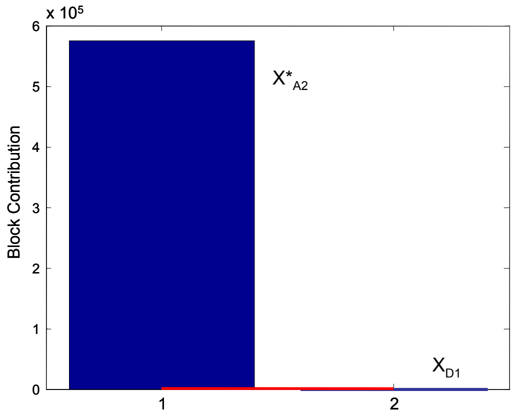

Fig. 5: Identification of the faulty block(s) contributing to the super T_s^2 statistic

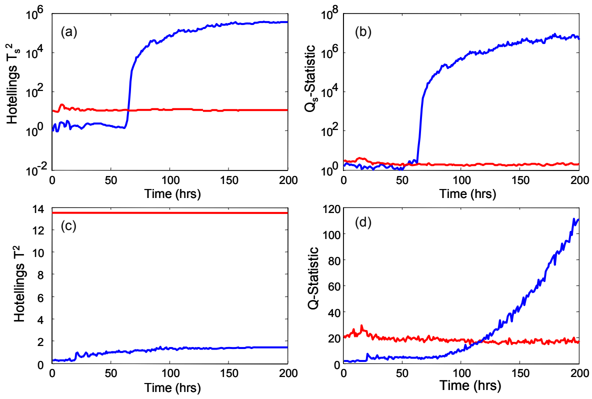

Fig.6: Monitoring statistics for individual blocks

Contribution plots based on the Q-statistic can also be derived. Figure 7 shows the contribution plots for the Q-statistic and Hotelling's T^2 for the first approximation respectively. It can be clearly observed that variable 7, dissolved oxygen concentration, breach's the contribution plot confidence bounds and hence makes a major contribution to the out-of-control signal.

3.2 Process drift

Process drift is a complex fault whereby its effect can propagate to other measurements and which can be difficult to detect and diagnose as they affect the mass, energy and reaction kinetics of the process resulting in the fault eventually appearing in a number of variables. The process fault introduced in this case is a process drift in the substrate feed rate which leads to a change in DO2 which may affect the culture volume. Figure 8 shows the time series plot of the substrate feed rate with fault magnitude 5$\boldsymbol{\omega}$.

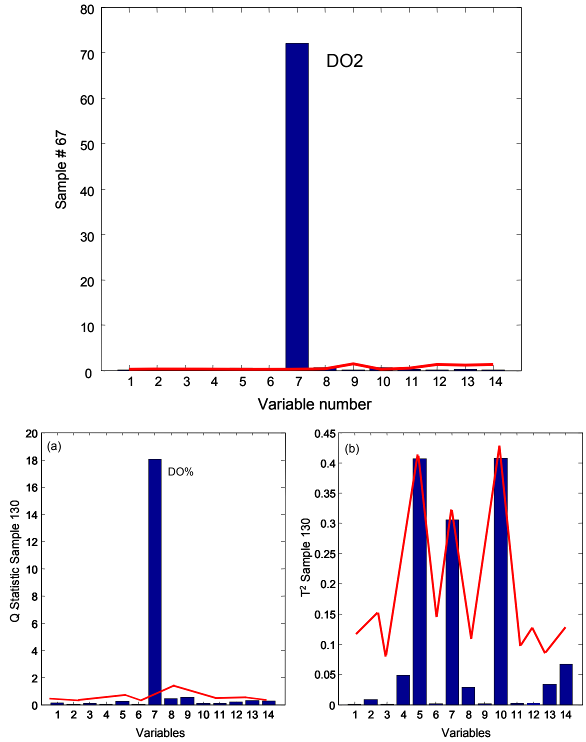

Fig. 7: Upper Plots – MSPCA variable contributions to the Q-statistic (approximation block) at t = 67; Lower Plots – MPCA variable contribution to the Hotelling's T^2 at sample time point 130

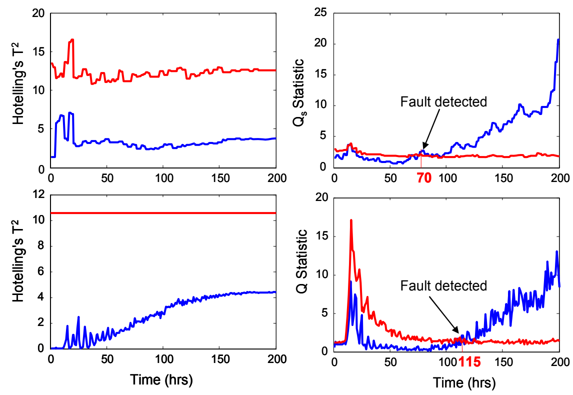

Fig. 8: Monitoring charts for a substrate feed rate drift: MSPCA Super T_s^2 and Q_s statistics (upper plots); MPCA T^2 and Q statistics (lower plots)

Figure 8 shows the monitoring statistics for both MSPCA and MPCA. It can be seen that MSPCA outperforms conventional MPCA. The fault was detected at sampling point 70 using multiscale PCA compared to sampling point 116 using conventional multiway PCA. From the approximation block (Figure not shown) it was observed that variable 3 (substrate feed rate), variable 7 (biomass concentration) and variable 9 (culture volume) violated the confidence limits due to the control and process interactions and loops. From the contribution plots (not shown) the only variable that breached the confidence limits at time point 70 was that associated with variable 3, substrate feed rate – the primary contributor to the fault which has not, at this time, propagated to other process measurements.

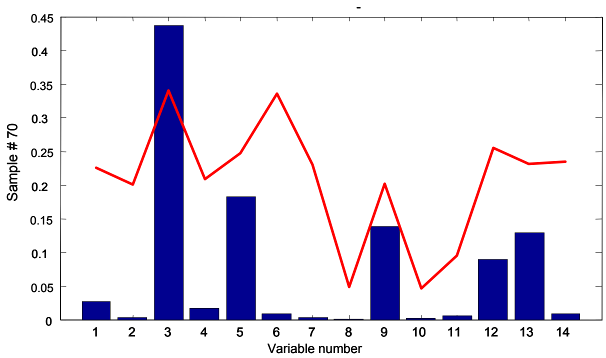

Figure 6.30 Variable contribution to the *Q*-statistic for drift type fault at sample time point 70

3.3 Wavelets for Fault Detection

There are many questions that can be raised with respect to multi-scale monitoring from which wavelet transform to use to false alarm rates. The basis function determines the information extracted from the data and allows study of different signal structures. The Haar wavelet provides orthogonal wavelets with features including compact support, symmetry and orthogonality at the expense of frequency localization and tends to emphasize discontinuities in the raw data. Other basis functions such as the Daubechies family are less discontinuous and probably better suited for representing smother variations. An investigation of the average detection delay for four wavelet functions computed over 100 runs gave figures of - Haar (1), Daubechies 4 (10), Coiflet (1) and Symlet (1) compared to no detection using standard MCPA. In this particular study the detection delay was not greatly affected by the choice of the wavelets functions.

4. CONCLUSIONS

The paper has presented some of the results of a larger study into assessing the characteristics of number different faults and to examine where multiscale PCA provides enhanced performance over conventional MPCA has been demonstrated to be superior in capturing faults that contain contributions with the same localization everywhere in the time–frequency domain. Furthermore, since the drift behavior in the sensor outputs are separated from other events (e.g. noise) by wavelet decomposition, the sensitivity of detecting sensor drift can be increased and results in a more sensitive fault detection approach. The approach is being evaluated in fermentation and batch cooling crystallisation.

5. ACKNOWLEDGMENTS

The authors acknowledge Dr S.W. Choi, for his contributions, and the UK ORS Scheme and CBBII (EPSRC GR/43853/01) for funding A Alawi.

REFERENCES

Alawi, A., Martin, E. B., & Morris, A. J., (2005) Statistical Process Monitoring Using State Space Modeling and Wavelet analysis. *Proc. ESCAPE-15*, 1459-1465

Bakshi, B. R., (1998) "Multiscale PCA with Application to Multivariate Statistical Process Monitoring", *AIChE Journal*. 44, 1596.

Bra´s, L.P., Bernardino, S.A, Lopes, J.A., Menezes, J.C., (2004) "Multilock PLS as an approach to compare and combine NIR and MIR spectra in calibrations of soybean flour", *Chemom, Intell. Lab. Syst*. 77, 1, 91-99.

Choi, W. C., Lee, I-B. (2004) "Multiblock PLS-Based Localized Process Diagnosis", *Journal of Process Control*, 15 (3), 295-306.

Felicioa, C. Bra´s, C., Lopes, L.P., Cabrita, J. A., Menezes, J.C., (2005) "Comparison of PLS algorithms in gasoline and gas oil parameter monitoring with MIR and NIR", *Chemom.& Intell. Lab. Syst*, 78,1-2, 74-80.

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

Misra, M., Yue, H., Qin, S.J., and Ling, C., (2000) "Multivariate Process Monitoring and Fault Diagnosis By Multiscale PCA", *Computers and Chemical Engineering*, 26 (9), 1985.

Nomikos, P. & MacGregor, J.F. (1994) "Monitoring batch processes using multi-way principal components analysis". *AIChE Journal*, 40(8), 1361–1375.

Nounou, M., & Bakshi, B. R. (1998). "On-Line Multiscale Filtering of Random and Gross Errors Without Process Models". *AIChE Journal*, 45 (5), 1041–1058.

Qin, S.J., Valle, S., Piovoso, M. J. (2001) "On Unifying Multiblock Analysis With Application to Decentralized Process Monitoring", *J. of Chemometrics*, 15, 715-742.

Rosen, C, Lennox, J.A, (2001) "Monitoring Wastewater Treatment Operation. PartII: Multiscale monitoring", *Water Research*, 35,3402-3410.

Smilde, A.K., Westerhuis, J. A., Jong, S., (2003) "A Framework for Sequential Multiblock Component Methods", *J. of Chemometrics*, 17, 323-337.

Teppola, P., & Minkkinen, P., (2000) "Wavelet-PLS regression models for both data exploratory data analysis and process monitoring", *J. of Chemometrics*, 14, 383.

Wold, S.; Geladi, P.; Esbensen, K.; Ohman, J. (1987) Multi-way principal components and PLS Analysis. *J. of Chemometrics* 1**,** 41–56.

Yoon, S., J.F. MacGregor (2001) ,"Fault diagnosis with multivariate statistical models part I:Using Steady State Fault Signature", *J. Process Control*, 11, 387-400.

Westerhuis, J. A., Kourti, T., MacGregor, J. F., (1998) "Analysis of Multiblock and Hierarchical PCA and PLS Models". *Journal of Chemometrics*, 12, 301-321.

Wong, C.W. L., Escott, R. E., Morris, A. J., Martin, E. B., (2005) "The Integration of Process and Spectroscopic Data for Enhanced Knowledge Extraction in Batch Processes". *Proc. ESCAPE-15*, Barcelona, 1141-1147.