MONITORING PERFORMANCE IN FLEXIBLE PROCESS MANUFACTURING

E.B. Martin and A.J. Morris

Centre for Process Analytics and Control Technology
School of Chemical Engineering and Advanced Materials, University of Newcastle,
Newcastle upon Tyne, NE1 7RU, England.

Abstract: Partial Least Squares (PLS) is a popular method for the development of a framework for the detection and location of process deviations. A limitation of the approach is that it has generally been used to monitor one recipe, one product, for example, consequently applications may have been ignored because of the need for a large number of process models to monitor multi-product production. This paper introduces two extensions - multi-group and multi-group-multi-block PLS. These techniques enable a number of similar products, manufactured across different unit processes, to be monitored using a single model. The methodologies are demonstrated by application to a multi-recipe industrial manufacturing process.

Keywords: Modelling; Statistical Analysis; Manufacturing Processes

1. INTRODUCTION

Of particular strategic importance, especially in highly competitive world markets, is the drive to reduce the amount of off-specification production, re-work and waste, and minimize operating costs whilst continuously striving to reduce the levels of variability inherent within the manufacturing process. These drivers have led to the development of performance monitoring systems that not only provide early warning of the onset of process upsets, equipment malfunctions or other special events, but which also contribute to the identification and the location of the assignable causes of these events.

One of the more recent approaches for assessing and improving the performance and operation of manufacturing processes, and the quality and consistency of production, has been Multivariate Statistical Process Control (MSPC). The underlying methodologies are equally applicable to continuous and batch processes and have been investigated by a number of researchers including Kosanovich and Piovoso, (1995), Kourti *et al*, (1995), Martin *et al*, (2002); Simoglou *et al,* (2000); Rius *et al* 1997, Weighell *et al*, (2001) and Wise and Gallagher (1996).

Today production processes are required to meet the needs of changing markets and product diversification resulting in a drive towards flexible manufacturing. The process performance methodologies reported in the literature have typically focused on single product manufacture. Thus to accommodate market driven changes and multi-product manufacturing, there is a real need for process models which allow a range of product types, grades or recipes to be monitored using a single process representation. Traditionally multiple product manufacture has been handled through the development of separate control charts for each type of product, recipe or grade. In many process monitoring situations this may be impractical because of the large number of control charts required to monitor all the products being manufactured and the limited amount of data available from which to develop a process representation. Thus for the statistical monitoring of multiple products, there is a need to develop a methodology whereby the between group variation is removed, so that interest can focus on the within process (product) variability.

An extension to the statistical projection technique of Principal Component Analysis has previously been reported, Lane *et al.* (2001). Within this paper the methodology is extended to Partial Least Squares (PLS). In particular both multiple group and multi-group multi-block algorithms are proposed based on combining the variance-covariance matrices of each of the individual groups. The latent variable loadings are then calculated from the pooled variance-covariance matrix of the individual groups. For the calculation of the pooled correlation (variance-covariance) matrix, it is assumed that a common eigenvector subspace spanned by the first *a* eigenvectors of the individual correlation (variance-covariance) matrices exists.

The approach is illustrated through the development of multi-group PLS models for the monitoring of a commercial semi-discrete batch manufacturing operation, which involves the production of a variety of products (recipes), some of which are only manufactured in small quantities. In this particular application the multi-group methodology is able to handle different recipes that contain a number of different raw materials, as well as several ingredients that are specific to a particular recipe, and different numbers of monitored process variables. By grouping the recipes into a number of sub-groups (families), separate monitoring schemes for each family can be developed. The advantage of this approach is that instead of developing a separate model for each individual product type, which could result in a large number of models, a multi-group model is built for each family. An extension to this application was considered where several recipes undergo a pre-mixing as well as a main mixing stage. For this study a multi-block multi-group methodology was developed.

2. CURRENT METHODOLOGIES

2.1 *Introduction*

The main advantage of statistical projection techniques such as PCA and PLS is their ability to handle large numbers of highly correlated process variables. At the same time they have the ability to project the information contained within large data sets onto lower dimensional subspaces defined by a few latent variables, which are a weighted combination of the original variables. However, when the number of variables contained within a data set becomes extremely large the monitoring charts and in particular the diagnostic charts can become difficult to interpret. When this occurs there is often a temptation to exclude a number of the measured variables in an attempt to make the charts more manageable. Excluding variables can lead to a loss of information, model mismatch or incorrect diagnostic information. Consequently the concept of the multi-block techniques has grown in popularity.

2.2. Hierarchical and Multi-block Modeling

The objective of hierarchical principal component analysis (HPCA) and multi-block projection to latent structures (MBPLS) is to divide the process into a number of meaningful blocks and then apply PCA or PLS to the individual blocks. The individual block scores are combined into a consensus matrix and PCA or PLS is once again applied. The consensus scores are then used to monitor the overall process. Consequently when a process disturbance occurs, it is possible to isolate the block in which the disturbance has occurred. The individual block scores are then used to provide information on the specific variables that are indicative of the source of the disturbance. A schematic of the MBPLS monitoring scheme is shown in Fig. 1. Here the consensus monitoring chart detects the process disturbance and the corresponding contribution plot identifies the block in which the disturbance occurred and finally the contribution plot for that particular block identifies the variables indicative of the disturbance.

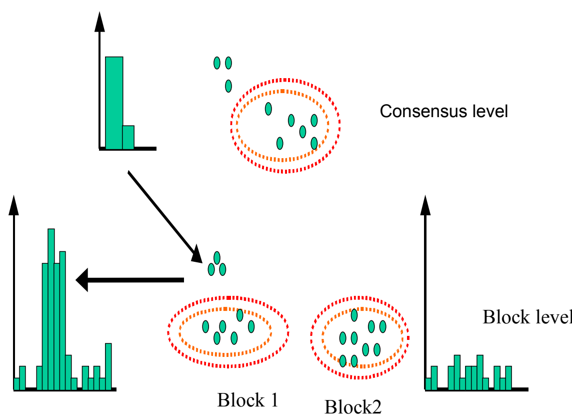

Fig. 1. Multi-block process performance monitoring.

The concept of hierarchical multi-block PCA and PLS was initially introduced around 1986. More recently papers on hierarchical algorithms have been published by Wold *et al* (1996) and Westerhuis *et al* (1998). The original algorithm was termed Consensus PCA (CPCA) and was presented as a method for comparing several blocks of descriptor variables measured on the same object. The data was divided into a number of blocks and the block loadings computed. In turn these were used to calculate the block scores, which were combined into a "super" block. The super block loadings were then computed and normalised to unit length. Finally a super score is computed. After convergence all the blocks are deflated using the super scores and a second super score, orthogonal to the first, is computed using the block residuals.

The algorithm presented by Wold *et al* (1996), Hierarchical Principal Component Analysis (HPCA), is similar to CPCA. The only difference being that instead of normalising the super loadings Wold *et al* (1996) normalised the super scores. In that paper the HPCA algorithm was demonstrated through its application to a catalytic cracker where both the process and quality variables were divided into a number of separate blocks and the resulting blocks combined into two consensus matrices, one for the process and one for the quality data. A second PLS was then carried out between the two consensus matrices.

Both the CPCA and HPCA algorithms were reviewed by Westerhuis *et al* (1998). They reported that both algorithms had convergence problems and needed modification. For the CPCA algorithm the block loadings as well as the super loadings had to be normalised. In a similar manner, the individual block scores of the HPCA algorithm also required normalising to unity. However, this still does not guarantee convergence with the algorithm converging to different solutions depending on the initial super score starting vector.

3. MULTI-GROUP PLS

The multiple group modeling approach was initially developed for principal component analysis, Lane *et al*. (2001). It can however be extended to PLS. PLS is a projection based method which maximizes the correlation between the process variables **X** and the quality variables **Y**. The objective being to summarize the variation in the data set using surrogate or latent variables that are linear combinations of the process variables. Dimensionality reduction is achieved if the number of latent variables is less than the number of process variables in the data set. A detailed description of PLS is given by Geladi and Kowalski (1986) and Garthwaite (1994).

For the multi-group approach, the algorithm presented by Lindgen et al. (1993) forms the basis of the computational approach adopted here. This method uses the sample variance-covariance matrix, $\mathbf{X}^T\mathbf{Y}\mathbf{Y}^T\mathbf{X}$, to estimate the latent variable loadings vector **W**, where **X** is the standardised process data matrix and **Y** is the standardised quality data matrix. The loading vector $\mathbf{W}_1$ is the first eigenvalue of the matrix $\mathbf{X}^T\mathbf{Y}\mathbf{Y}^T\mathbf{X}$. Once $\mathbf{W}_1$ has been estimated, the latent variable scores **t** are calculated as:

$$\mathbf{t}_1 = \mathbf{X}\mathbf{W}_1 \tag{1}$$

With the regression loading vectors, **P** and **Q**, given by:-

$$\mathbf{P} = \frac{\mathbf{t}_1\mathbf{X}}{\mathbf{t}_1\mathbf{t}_1} \tag{2}$$

$$\mathbf{Q} = \frac{\mathbf{t}_1\mathbf{Y}}{\mathbf{t}_1\mathbf{t}_1}$$

The **X** and **Y** data block matrices are then deflated to give:-

$$\mathbf{X}_{\text{new}} = \mathbf{X} - \mathbf{t}_1\mathbf{P}^T \tag{3}$$

$$\mathbf{Y}_{\text{new}} = \mathbf{Y} - \mathbf{t}_1\mathbf{Q}^T$$

The second latent variable is then calculated as:

$$\mathbf{t}_2 = \mathbf{X}_{\text{new}}\mathbf{W}_2 \tag{4}$$

where $\mathbf{W}_2$ is the first eigenvalue of :-

$$\mathbf{X}_{\text{new}}^T\mathbf{Y}_{\text{new}}\mathbf{Y}_{\text{new}}^T\mathbf{X}_{\text{new}} \tag{5}$$

This iterative procedure is continued until all the required latent variables have been computed. To extend this PLS algorithm to multiple group situations requires pooling the sample variance-covariance matrices of each of the groups and estimating the common loadings vectors $\mathbf{W}_i$ from the pooled variance-covariance matrix. For example for the two group case, this would require pooling the variance-covariance matrices:

$$\mathbf{X}_1^T\mathbf{Y}_1\mathbf{Y}_1^T\mathbf{X}_1 \text{ and } \mathbf{X}_2^T\mathbf{Y}_2\mathbf{Y}_2^T\mathbf{X}_2 \tag{6}$$

where $\mathbf{X}_i$ are the standardised matrices of the process data for each data group and $\mathbf{Y}_i$ are the standardised matrices of quality data for each data group. In both cases $i = 1,2$. The pooled variance-covariance matrix is given by:

$$\frac{(n_1 - 1)\mathbf{X}_1^T\mathbf{Y}_1\mathbf{Y}_1^T\mathbf{X}_1 + (n_2 - 1)\mathbf{X}_2^T\mathbf{Y}_2\mathbf{Y}_2^T\mathbf{X}_2}{n_1 + n_2 - 2} \tag{7}$$

where n_i are the number of observations in each group, $i = 1,2$. The loadings vector **W** is the first eigenvector of the pooled variance-covariance matrix. The iteration procedure can then be followed in a similar manner to that described previously.

The pooled correlation (variance-covariance) approach is based on the existence of a common eigenvector subspace spanned by the first *a* eigenvectors of the individual correlation (variance-covariance) matrices. A formal statistical model was defined by Flury (1987), who computed the common principal components using Maximum Likelihood Estimation (MLE). Previously Krzanowski (1984) had demonstrated that the common principal components derived using the pooled correlation (variance-covariance) matrix were almost identical to those computed from MLE. In practice the pooled correlation (variance-covariance) approach proposed by Krzanowski (1984) is simpler to apply than the MLE approach which requires the implementation of an iterative algorithm.

Furthermore, the pooled correlation (variance-covariance) approach compares the subspaces defined by the eigenvectors associated with the largest eigenvalues whereas no conditions are placed on the MLE approach proposed by Flury (1987). This is a major consideration when determining the method to be used for calculating the latent variables for process monitoring. In process monitoring, it is convention to construct the process models using the eigenvectors corresponding to the largest eigenvalues. As a consequence determining the common principal components from the pooled correlation (variance-covariance) matrix is more appropriate for the industrial processes being considered (Lane *et al*, 2001).

4. MULTIPLE GROUP MULTI-BLOCK PROJECTION TO LATENT STRUCTURE

The algorithm for multiple group PLS can be extended to situations were the process and/or quality variables can be divided into separate blocks. Multi-block PLS is particularly useful in situations where the number of variables is very large making the monitoring schemes unwieldy and difficult to interpret (MacGregor *et al*. 1994). As an example, the algorithm for a multiple-group two-block case is presented.

1. Construct the individual variance-covariance matrices for each group and each block:

$$\mathbf{R}_{kl} = \frac{(\mathbf{X}_{skl}^T \mathbf{Y}_{sk})(\mathbf{X}_{skl}^T \mathbf{Y}_{sk})^T}{(n_k - 1)^2} \tag{8}$$

   where $\boldsymbol{X}_{skl}$ is the scaled process data for group k and block (l), $\boldsymbol{Y}_{sk}$ is the scaled quality data for group k and n_k is the number of observations in group k.

2. The pooled variance-covariance matrix is then calculated for each block (l):

$$\mathbf{R}_l = \frac{\sum_{k=1}^{2} (n_k - 1)\mathbf{R}_{kl}}{N - g} \tag{9}$$

   where N is the total number of observations in all k groups and g is the number of groups.

3. The loading vector for the first latent variable for each block ($\mathbf{w}_{l1}$), i.e. the leading eigenvector of the respective pooled variance-covariance matrices ($\mathbf{R}_l$), is then calculated.

4. The latent variable scores ($\mathbf{t}_{kl}$) for each group and each block are given by:

$$\mathbf{t}_{kl} = \mathbf{X}_{skl}\mathbf{w}_{l1} \quad k = 1,\ldots, g \text{ and } l = 1, 2. \tag{10}$$

5. The scores for each block are then combined into a consensus matrix:

$$\mathbf{T}_k = [\mathbf{t}_{k1} : \mathbf{t}_{k2}] \tag{11}$$

   where $\mathbf{T}_k$ is the consensus matrix for group k.

6. The consensus pooled variance-covariance matrix is then constructed with the scores matrices ($\mathbf{T}_k$) replacing the process data ($\mathbf{X}_{skl}$) in the following way:

$$\mathbf{R}_c = \frac{\sum_{k=1}^{2} (n_k - 1)\mathbf{R}_k}{N - g} \tag{12}$$

   where ($\mathbf{R}_k$) are the individual consensus variance-covariance matrices given by: -

$$\mathbf{R}_k = \frac{(\mathbf{T}_k^T \mathbf{Y}_{sk})(\mathbf{T}_k^T \mathbf{Y}_{sk})^T}{(n_k - 1)^2} \tag{13}$$

7. The latent variable loading vector ($\mathbf{v}_1$) for the consensus model is the leading eigenvector of the matrix ($\mathbf{R}_c$).

8. The consensus latent variable scores are then calculated for each group:

$$\mathbf{t}_{ck} = \mathbf{T}_k \mathbf{v}_1 \tag{14}$$

9. The loading vectors for each group and each block are then calculated: -

$$\mathbf{p}_{kl} = \frac{\mathbf{t}_{kl}^T \mathbf{X}_{skl}}{\mathbf{t}_{kl}^T \mathbf{t}_{kl}} \quad \text{and} \quad \mathbf{q}_{kl} = \frac{\mathbf{t}_{ck}^T \mathbf{Y}_{sk}}{\mathbf{t}_{ck}^T \mathbf{t}_{ck}} \tag{15}$$

10. The process data matrices ($\mathbf{X}_{skl}$) and the quality data matrices ($\mathbf{Y}_{sk}$) are then finally deflated:

$$\mathbf{X}_{skl\,new} = \mathbf{X}_{skl} - \mathbf{t}_{kl}\mathbf{p}_{kl}^{T} \qquad (16)$$
$$\mathbf{Y}_{sk\,new} = \mathbf{Y}_{sk} - \mathbf{t}_{ck}\mathbf{q}_{k}^{T}$$

Steps 1 to 10 are repeated with each new set of residual matrices replacing the previous matrices until all the required latent variables have been extracted.

5. INDUSTRIAL APPLICATION

5.1. Process Description

The industrial process selected to demonstrate the detection and diagnostic capabilities of the multi-group and multi-group multi-block methodologies is a batch production process in a manufacturing sense. The process includes the two main characteristics of batch operation, flexibility and finite duration. However, in a statistical sense the data matrix is only two-dimensional with a single, between batch source of variability. This is as a result of the semi-discrete nature of this particular manufacturing process, with the process variables being measured only once during each batch.

The process produces a wide range of household products to meet the demands of a rapidly changing and evolving market. It comprises a sequence of individual production steps involving the sequential dosing and mixing of a number of raw materials. During each raw material addition (dosing) a number of process measurements are recorded including operations and dosing times, flow meter measurements and load cell dose weights, batch temperature and temperature of the hot and cold process water added to the batch.

There are two types of product produced on the plant. Each formulation can be subdivided into a number of different recipes, which in-turn can be further subdivided into a number of different varieties. The formulations are manufactured in separate areas of the plant, with each area having two mixers that are used simultaneously. For some recipes there is a pre-mixing stage which is carried-out in parallel with the main mixing process, prior to being added to the main mix. Off-line analysis of several historical data sets showed that there were distinct differences between the mixers (Lane *et al*, 2001). As a consequence a separate model would be required for each mixer as well as for each variety of recipe, which is not a practical solution.

5.2. *Multi-group PLS*

To demonstrate the application of multiple group PLS the production of four recipes is considered. Each process mixer is also considered as a separate 'recipe' and thus the process model contains four distinct groups as shown in Table 1. It is observed that recipes 3 and 4 have fewer raw materials and process variables than recipes 1 and 2.

Table 1: Composition of the multi-recipe data sets

Recipe	**Mixer**	**Batch's**	**Raw Mat's**	**Proc Var's**	**Qual Var's**
1	1	19	23	120	1
2	2	21	23	120	1
3	1	20	17	90	1
4	2	29	17	90	1

The pooled variance-covariance matrix is constructed as a weighted sum of the individual elements of the individual variance-covariance matrices. To address recipes containing different numbers of raw materials and process variables, the pooled variance-covariance matrix is constructed from the individual elements of each of the individual variance-covariance matrices:

$$\mathbf{R}_p(k,j) = \frac{\sum (n_i - 1)\mathbf{R}_i(k,j)}{\hat{N} - \hat{g}} \qquad (17)$$

where $\mathbf{R}_p(k,j)$ is the $(k,j)^{\text{th}}$ element of the pooled covariance matrix, $\mathbf{R}_i(k,j)$ is the $(k,j)^{\text{th}}$ element of the individual variance-covariance matrix for recipe i, $\hat{g}$ is the number of recipes containing both the i^{th} and j^{th} variables and $\hat{N}$ is the total number of observations for the groups containing both the i^{th} and j^{th} variables. If either or both of the variables are not present in a recipe then there is no contribution to the pooled variance-covariance matrix.

Fig. 2 shows the reference model constructed using the pooled variance-covariance matrix of the four groups to calculate the latent variable loadings. The ability to handle data sets containing different numbers of variables is a powerful advantage of the pooled variance-covariance approach. However, care needs to be taken when considering recipes containing different numbers of variables because the inclusion/exclusion of variables may affect the covariance structure of the data set. Cross-validation indicated that the first ten latent variables, explaining 60% of the process variation and 70% of the quality variation were sufficient to monitor the process.

To demonstrate the detection and diagnostic capabilities of the multiple-group PLS model, a batch exhibiting an abnormal product temperature profile was projected onto the reference model, Fig. 3. The bivariate scores plot of latent variables 3 and 4 clearly identified the batch as being abnormal. In this particular example, by examining the loadings, it was identified that latent variable 3 was associated with the product temperature and latent variable 4 the temperature of the process water added to the batch.

The monitoring of process performance takes place on the completion of each raw material dosing step. It can be seen that the observations move away from the centre of the control region following the first raw material dosing. Subsequent observations continue to drift away from the control region and an "out-of-statistical-control" signal occurs after completion of the third raw material dosing.

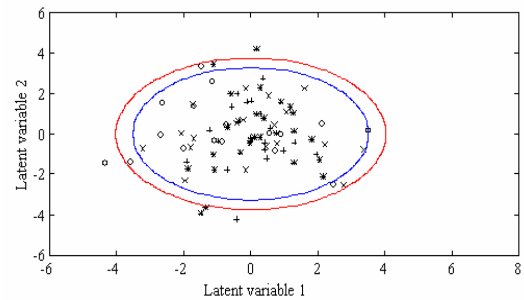

Fig. 2. Bivariate Scores Plot (Pooled) (Recipe 1 'x', Recipe 2 'o', Recipe 3 '+', Recipe 4 '*')

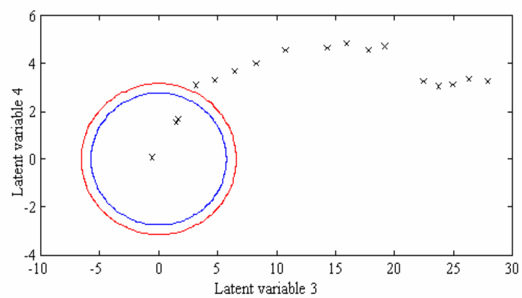

Fig. 3. Bivariate scores plot for latent variable three versus latent variable four.

The contribution plot for latent variable three (Fig. 4), for the first observation outside the action limits, shows three variables making larger contributions to the latent variable scores - variable 4 (cold process water temperature), variable 5 (product temperature following the first raw material dosing) and variable 14 (product temperature following third raw material dosing).

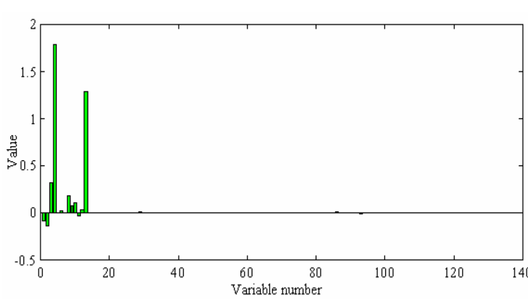

Fig. 4. Contribution Plot for latent variable three

The problem was identified to be due to the failure of the process water cooling system and water entered the batch above the required temperature, which led to the product temperature being abnormally high. As the product temperature is monitored following each raw material dosing, the latent variable scores continue to move away from the control region. A second dosing of process water also occurred later on in the batch which can be observed by the small change in the direction of the scores following the seventh raw material dosing (Fig. 3). At present no remedial action is taken during the production process and one of the motivations for the on-line application was to allow operators to be more proactive when a processing problem occurs.

6. MULTI-GROUP MULTI-BLOCK PLS MONITORING

Several of the recipes in the manufacturing process undergo a pre-mixing as well as a main mixing stage during their manufacture. This makes multi-block PLS a particularly attractive methodology for building a monitoring model as there is a natural blocking of the variables. To demonstrate the multiple group multi-block PLS methodology three recipes were selected. This resulted in six data sets, since each recipe can be produced in either of two mixing areas. The composition of the data set is summarised in Table 2.

Table 2: Composition of the process data sets

Recipe	Mixing Area	Batch's	Quality Variables
1	1	48	3
1	2	43	3
2	1	16	3
2	2	27	3
3	1	51	3
3	2	31	3

Pre-mixer		Main-mixer	
Raw Materials	Process Variables	Raw Materials	Process Variables
10	41	17	84
10	41	17	84
10	41	17	84
10	41	17	84
10	41	17	84
10	41	17	84

In this case each recipe contains the same number of raw materials. This was because the recipes were very similar and only differed in terms of colouring. In this application the individual colourings were treated as being the same raw material because the colouring was only considered as being a "minor" raw material and was perceived to have no influence on product quality.

As in the previous study, a set of monitoring charts were constructed using data collected when the process was manufacturing acceptable quality product and there were no assignable causes of variation present in the data. In this example three sets of monitoring charts were required one for each individual block and one for the overall process, i.e. the consensus model.

The fault detection and diagnostic capabilities of the multiple group multi-block PLS model, are demonstrated using a batch where a raw material overdosing took place. As with the previous example, each latent variable score "x" represents the status of the current "on-line" batch following each successive raw material dosing. The consensus scores plot for monitoring the overall process (Fig. 5) shows an "out-of-statistical-control" signal at the sample point following the overdosing. Unlike the previous example, the latent variable scores cluster in the center of the control region indicating "good" operation until the overdosing occurs. Following the overdosing there is an abrupt departure from the control region.

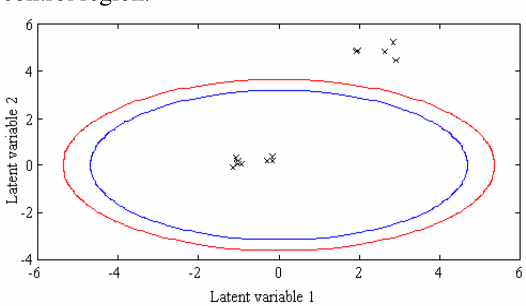

Fig. 5. Latent variable scores plot (Consensus chart)

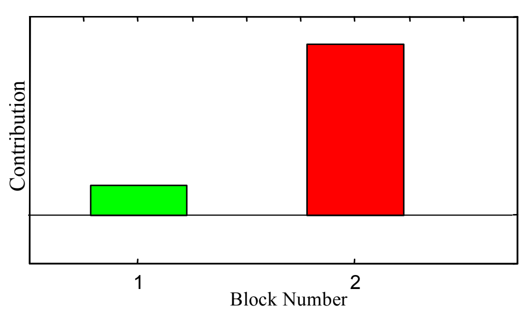

Fig. 6. Contribution plot for latent variable 2 (Consensus chart)

Examination of the contribution plot for latent variable 2 (Fig. 6) identified the pre-mixer as being the process subsection where the overdosing occurred. Inspection of the monitoring chart for the pre-mixer (Fig. 7) also identified the "out-of-statistical-control" signal at the sample time following the overdosing. In contrast no out-of-control signal was identified from the latent variable scores plot of the main mixer (Fig. 8).

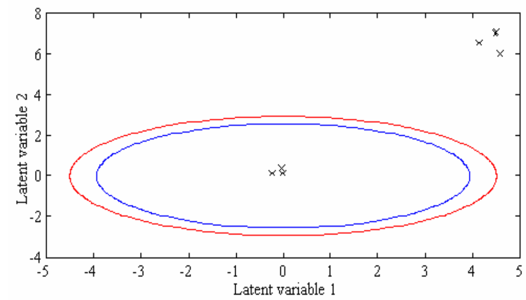

Fig. 7.Latent variable scores plot (Pre-mixer chart)

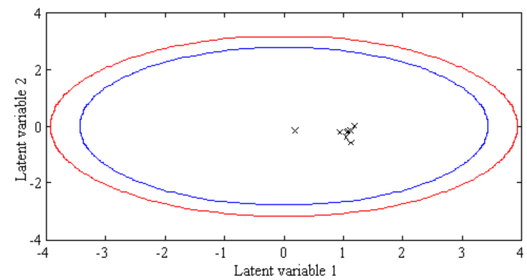

Fig. 8.Latent variable scores plot (Main mixer chart)

For this particular fault both the consensus and the pre-mixing monitoring charts detect the fault at the same time. This is due to the small number of variables included in each block and the abrupt impact that the fault had on the latent variable scores. This is a direct consequence of the individual block monitoring charts only monitoring the deviation of the variables contained within that particular block and not the whole process, which is the case with the consensus monitoring charts.

Investigation of the contribution plot for latent variable 2 (Fig. 9) identified variables 9 (dosing time), 10 (load cell dose weight), 11 (flow meter dose weight) and 12 (product temperature) as being responsible for the "out-of-statistical-control" signal.

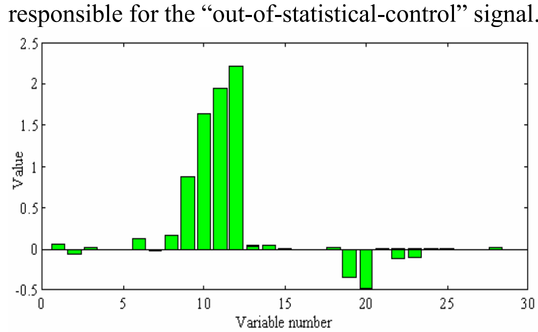

Fig. 9. Contribution plot for latent variable 2 (pre-mixer).

From this information, it was concluded than an overdosing of raw material 3 had occurred. A process engineer later confirmed that the overdosing was caused by a malfunction of a dosing valve that controls the amount of raw material being added into the batch. Prior to the application of MSPC, a similar problem had caused the manufacture of a number of "out-of-specification" batches before the problem had been identified, by chance.

In such manufacturing situations only a few "key" variables may be monitored and faults impacting on the remaining variables can be missed. The study has demonstrated that such processes could be monitored effectively using multi group multi-block MSPC. In contrast to the pre-mixing monitoring chart, the latent variable scores for the main mixer remain clustered in the center of the control region throughout the manufacture of the entire batch (Fig. 8). This is because the fault detected in the pre-mixer has no impact on the variables being monitored in the main mixer. However, as the product from the pre-mixer is added to the main mixer, the product in the main mixer will be affected by the pre-mixer fault unless the process is halted when the fault occurs.

7. CONCLUSIONS

The industrial application has demonstrated the extension of PLS based MSPC methodologies to processes where different products or recipes are produced with a different number of variables being monitored. In the manufacturing application, the eigenvectors of the pooled variance-covariance matrices gave a good representation of the production process even though in some cases the recipes contained a different number of raw materials and as a consequence the data sets contained different numbers of variables.

The results showed that for both multi group and multi group multi-block PLS the models had good detection and diagnostic properties. The latent variable scores plots detected batches which were out-of-statistical-control and the corresponding scores contribution plots identified the variables responsible for the out-of-control signal. In this application the contribution plots gave explicit information about the process abnormalities.

8. ACKNOWLEDGEMENTS

The authors acknowledge the EU ESPRIT Project 'Process Diagnostics for Plant Performance Enhancement' (PROGNOSIS No. 22281) and in particular the contribution to the analysis by Dr Lane.

9. REFERENCES

Flury B.N. (1987). Two generalisations of the common principal component model, *Biometrika*, **74**, pp. 59-69.

Garthwaite P.H. (1994). An interpretation of partial least squares. *Journal of the American Statistical Association,* **89**, pp. 122-127.

Geladi P. and B.R. Kowalski (1986). Partial least-squares regression: A tutorial. *Analytica Chimica Acta*, **185**, pp. 1-17.

Kosanovich, K.A., S. Dahl and M.J. Piovoso (1996). Improved process understanding using multi-way principal component analysis, *Ind. Eng. Chem. Res*., **35**, pp. 138-146.

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,* **5**, pp. 277-284.

Kourti T. and MacGregor J. F. (1996), Multivariate SPC Methods for Process and Product Monitoring, *Journal of Quality Technology*, **28 (4)**, pp. 409-426.

Krzanowski W.J. (1984), Principal Component Analysis in the presence of Group Structure, *Applied. Statistics*, **33(2),** 164-168.

Lane S., E.B. Martin, R. Kooijmans and A.J. Morris (2001). Performance monitoring of a multi-product semi-batch process. *Journal of Process Control*, **11(1)**, pp. 1-11.

Lindgren, F., P. Geladi and S. Wold (1993). The kernal algorithm for PLS. *Journal of Chemometrics*, **7**, pp. 45 – 59.

MacGregor J.F., C. Jaeckle, C. Kiparissides and M. Koutoudi (1994). Process monitoring and diagnosis by multiblock PLS methods. *AIChE*, **40**, pp. 826-838.

Martin, E.B., A. Bettoni and A.J. Morris (2002). Manufacturing performance monitoring of a mixed batch continuous process through resampling. *Journal of Quality Technology,* **34(2)**, pp. 171-186.

Rius A., M. P. Callao and F. X. Rius (1997). Multivariate statistical process control applied to sulfate determination by sequential injection analysis. *The Analyst,* **122**, pp. 737-741

Simoglou, A., E. B. Martin and A. J. Morris (2000). Multivariate statistical process control of an industrial fluidised-bed reactor. *Control Engineering Practice*, **8**, pp. 893-909.

Weighell, M., E.B. Martin and A.J. Morris (2001). The statistical monitoring of a complex manufacturing process, *Journal of Applied Statistics,* **28(3&4)**, pp. 409-425.

Westerhuis J.A. and P.M.J. Coenegracht (1997). Multivariate modelling of the pharmaceutical two-Step process of wet granulation and tableting with multiblock partial least squares, *Journal of Chemometrics*, **11**, pp. 379-392.

Wise B.M. and N.B. Gallagher (1996). The process chemometrics approach to process monitoring and fault detection, *Journal of Process Control, **6***, pp. 329-348.

Wold S., N. Kettaneh and K. Tjessem (1996). Hierarchical multiblock PLS and PC models for easier model interpretation and as an alternative to variable selection. *Journal of Chemometrics*, **10**, pp. 463-482.