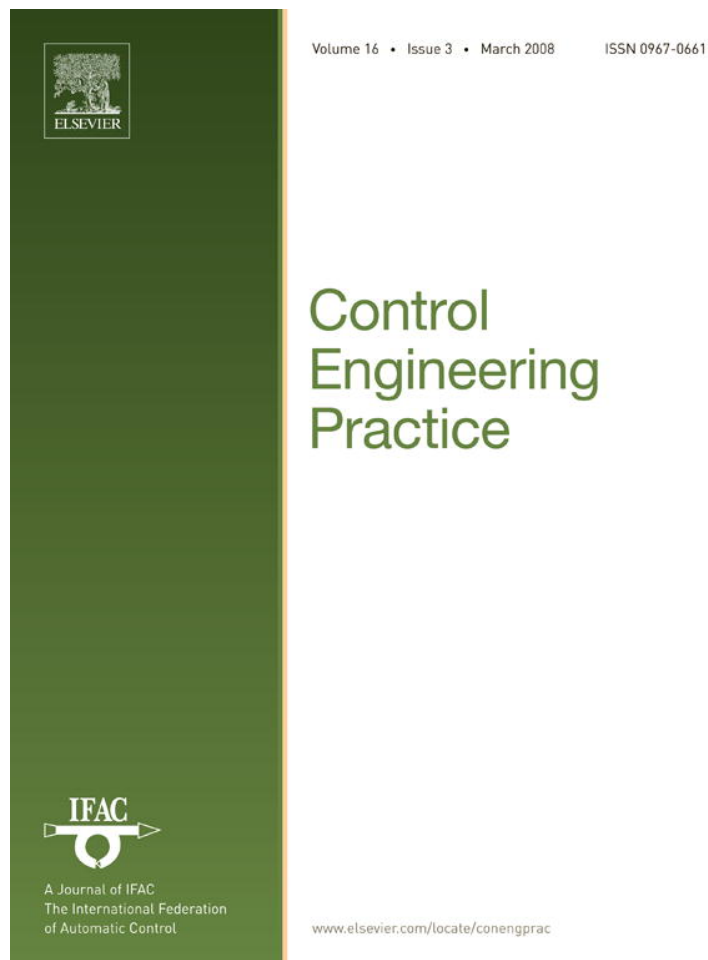




This article was published in an Elsevier journal. The attached copy is furnished to the author for non-commercial research and education use, including for instruction at the author's institution, sharing with colleagues and providing to institution administration.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



Monitoring a complex refining process using multivariate statistics

Ashraf AlGhazzawi, Barry Lennox*

Control Systems Group, The University of Manchester, UK

Received 19 September 2006; accepted 26 April 2007

Available online 27 June 2007

Abstract

Over the past decade, multivariate statistical process control (MSPC) methods have been proven, in the process industries, to be an effective tool for process monitoring, modelling and fault detection. This paper describes the development of a real-time monitoring solution for a complex petroleum refining process with an installed multivariable model predictive controller. The developed solution was designed to track the time-varying and non-stationary dynamics of the process and for improved isolation capabilities, a multiblock approach was applied. The paper highlights the systematic and generic approach that was followed to develop the monitoring solution and stresses the importance of exploiting the knowledge of experienced plant personnel when developing any such system.

© 2007 Elsevier Ltd. All rights reserved.

Keywords: Condition monitoring; Multivariate statistical process control; Model predictive control; Principal component analysis; Recursive PCA and multiblock PCA; Condensate fractionation

1. Introduction

In today's competitive oil and gas industry, the pressure to improve the performance of processing facilities is intense. Unplanned outages and equipment failures can have a significant impact on plant economics and result in substantial loss of revenues. Real-time performance monitoring and the early detection of degraded process performance and equipment failure is becoming a fundamental prerequisite to sustain both plant production ability and profitability (Nimmo, 1995). One approach to avoiding or better dealing with such situations is to make more use of the data that is routinely collected from process plant. The advent of modern process measurement, automation, and information systems has resulted in a significant increase in the amount of process data available to plant operators and engineers. Unfortunately, it is often very difficult to monitor such large amounts of data and assess the condition of a processing facility.

Venkatsubramanian, Rengaswamy, Yin, and Kavuri (2003) reviewed a variety of techniques that have been

proposed to exploit the large amounts of data available on process plants. Of these techniques, multivariate statistical process control (MSPC) methods, and principal component analysis (PCA) in particular, have been demonstrated to provide a powerful approach for the detection and isolation of abnormal conditions. MSPC methods are data driven techniques that reduce the dimension of process data and extract key features and trends in the process data that are of interest to plant personnel. In reducing the dimension of the process data, MSPC techniques are able to provide a robust approach to process modelling and real-time condition monitoring. One of the earliest articles describing the use of MSPC techniques in the process industries was presented by Kresta, MacGregor, and Marlin (1991), who described a basic methodology for using MSPC techniques to detect abnormal conditions in continuous processes. Since this time there have been many industrial applications of MSPC reported. Marjanovic, Lennox, Sandoz, Smith, and Crofts (2006), for example, successfully applied partial least squares (PLS) to monitor the progression of a batch reactor, Lane, Martin, and Morris (2001) applied PCA to a multirecipe manufacturing process and detected when overdosing had occurred and Dayal, MacGregor, Taylor, Kildaw, and Marcikic (2004)

*Corresponding author.

E-mail address: barry.lennox@manchester.ac.uk (B. Lennox).

applied PLS to monitor an industrial digester. In the petrochemical industry Geng and Zhu (2005) demonstrated how a non-linear PCA technique could be applied to detect fault conditions in an ethylene process and Badcock, Bailey, Jonathan, and Krzanowski (2005) successfully applied PCA to monitor a fluidised catalytic cracking unit.

MacGregor and co-workers have recently extended the scope of MSPC applications in the process industries by using PCA as an image analysis tool. Their work has included the detection of abnormal conditions in the flavouring of snack foods (Yu & MacGregor, 2003) and the detection of lumber defects (Bharati & MacGregor, 2003). Further details on the application of MSPC techniques to industrial data can be found in Kourti (2005), Martin, Morris, and Lane (2002) and Eriksson et al. (2005).

Although there have been a large number of industrial applications of MSPC reported in the literature, there have been far fewer documented cases where MSPC systems have been applied in real time with their results interpreted by plant operators, rather than MSPC experts. The development of real time MSPC applications introduces many issues which are frequently ignored in off-line studies. For example, if plant personnel are to adopt the system then it is essential that the system is accurate and simple to interpret. Inaccurate information and in particular false alarms are often introduced when the dynamics of the process change with time. Such changes are routine in process systems and hence any developed monitoring solution must be able to cope with this.

A limitation with MSPC techniques is that there are, as yet, no clear methodologies for developing robust monitoring systems for process applications. An analogy can be made here with model predictive control (MPC) systems. Clear methodologies for applying MPC were developed in the 1980's and these have resulted in the widespread application of this technology throughout the process industries (Qin & Badgewell, 2003). By developing similar design methodologies for MSPC systems, it is anticipated that take-up of this technology in industry will improve considerably.

A particularly successful real-time application of MSPC was reported by Miletic, Quinn, Dudzic, Vaculik, and Champagne (2004) and Zhang and Dudzic (2006), who developed a real-time monitoring system for a continuous slab caster. This system was able to identify several abnormal conditions, saving money and increasing operator confidence in running the equipment. A further real-time application was reported by Lennox, Montague, Hiden, Kornfeld, and Goulding (2001), who described the successful application of PCA to monitor the integrity of a copper smelting furnace. The developed system was sufficiently accurate for automated plant shutdown procedures to be configured which acted on the information provided by the PCA monitor. The developed system was shown to extend the campaign life of the furnace significantly.

The focus of the work described in this paper was the development of a real-time condition monitoring system, based on PCA, for identifying abnormal conditions of a refining process. This process was considered to be complex in terms of its dynamic characteristics and also its size with respect to process variables and disturbances. Unlike many other monitoring applications which are designed to detect and isolate specific fault conditions, the primary aim for the system developed in this study was to identify operating conditions that would indicate that the process, or the MPC system applied to this process, was not behaving in an optimal manner. The methodology adopted in developing the condition monitoring solution is considered to be generic and should therefore be applicable to many different applications.

In the following section of this paper the MSPC techniques applied in this work are described. Section 3 provides a brief overview of the refining process that was studied in this work. The results from the study are presented in Section 4 and the main conclusions from this study are reported in Section 5.

2. Multivariate statistical process control

Although there may be hundreds of plant variables that are measured in any given process, there tend to be only a small number of underlying characteristics that actually drive the process. The purpose of PCA is to identify a new set of variables that reflect these characteristics. These new variables, termed scores or latent variables are linear combinations of the original process variables. The expectation is that there will be fewer scores than plant variables and therefore the plant can be monitored with much greater ease by simply analysing these new variables. For full details of PCA the reader is referred to Jackson (1991). The following sections describe the monitoring statistics that are typically applied when using PCA and provide a brief description of multiblock and adaptive extensions to PCA.

2.1. Condition monitoring using MSPC

Whilst the scores obtained using PCA can themselves be monitored, using two and three-dimensional charts, it is often more appropriate to monitor two univariate statistics that can be identified when PCA is applied. These statistics have been defined to simplify the description of a new data point in terms of the pre-defined PCA model and are referred to as T^2 (or Hotelling's) and SPE (squared prediction error) and are defined as follows:

$$T^2 = \sum_{k=1}^{np} \mathbf{t}_k \sigma_k^{-2} \mathbf{t}_k^T, \quad (1)$$

where σ_k is the standard deviation of the k th t -score.

$$SPE = \|\mathbf{E}\|_2^2. \quad (2)$$

Relative to the data used to develop the PCA model, the T^2 and SPE statistics indicate how far the plant is currently operating from normal operating conditions and whether the relationships between the process variables have changed significantly. Control limits can be determined for these two statistics such that their violation indicates abnormal operating conditions. Following the detection of abnormal conditions, the contribution that each variable makes to the two statistics can be determined and possible causes of the condition can be identified.

2.2. Multiblock PCA

The concept of multiblock PCA was introduced by Wold, Esbensen, and Geladi (1987). The technique was proposed as a tool to compare several blocks of variables that were measured on the same object. For example, in a chemical process, the measurements recorded in each subsystem may be classified as a separate block. Since its original development, several multiblock algorithms have been proposed, consensus PCA and hierarchical PCA being the most commonly applied methods (Qin, Valle, & Piovoso, 2001; Westerhuis, Kourti, & MacGregor, 1998). In this study consensus PCA was applied.

In consensus PCA, scores and loadings are identified for each of the individual blocks and then the information from these blocks are combined into what are termed super scores. When monitoring a process using multiblock PCA, T^2 and SPE charts can be determined for each of the individual blocks and also for the 'super level'. Westerhuis et al. (1998) proved that the super scores of consensus PCA are identical to the scores of regular PCA and hence the super level statistics do not provide any information beyond what is obtained through regular PCA. However, by analysing the T^2 and SPE statistics of the individual blocks, it is often possible to isolate the cause of abnormalities to particular sections of the process.

2.3. Recursive PCA

Recursive, or adaptive, MSPC algorithms have been proposed in recent years to cater for the time varying nature of many process systems. Wold (1994) developed a recursive technique which used exponentially weighted moving average (EWMA) filters in conjunction with PCA. Li, Yue, Valle-Cervantes, and Qin (2000) argued that this was not a complete recursive PCA (RPCA) scheme and identified a computationally efficient algorithm for updating the PCA model whenever new process data became available. In the case study investigated in this work, the sample time was such that computation efficiency was not an issue and therefore recursion of the model was achieved by re-calculating a PCA model at each sampling instant using a moving window of process data.

With appropriate choice of parameters, the recursive algorithms proposed by Wold (1994) and Li et al. (2000) will produce the same results as the recursive algorithm

employed in this work. However, the techniques proposed by Li et al. (2000) and Wold (1994) allow forgetting factors to be used in the adaptation of the model so that the model is weighted to model recent data with greatest accuracy. Such techniques are not recommended in adaptive process modelling (Sandoz, 2003) and in condition monitoring can result in the model adapting more readily to abnormal conditions (Wang, Kruger, & Lennox, 2003).

A further issue when applying any recursive algorithm is that many processes will produce the occasional measurement *spike*. To prevent the PCA model adapting to these measurements a filter should be used to detect and eliminate them. In this work the Hampel filter (Pearson, 2002) was found to be suitable. It is also important that the PCA model does not adapt to abnormal conditions and hence whenever an abnormal event is confirmed, the data collected during this period should not be used to adapt the model.

3. The condensate fractionation process description

The process that was investigated in this study was a condensate fractionation process. Fig. 1 shows a simplified process flow diagram of the process. This process separates a hydrocarbon feed into a mix of products for further processing or product blending. Cold feed is pumped through the pre-heating train to heat the feed up to the required desalting temperature. The feed then enters the desalter, and the desalted feed is fed to a pre-flash drum where light products and any remaining free water are removed. The pre-flash drum product is then heated by a set of heat exchangers before entering a pre-flash distillation column. In this column, feed is distilled into light, medium and heavy products, which are removed from the top, middle, and bottom sections of the column, respectively. Heavy product from the bottom of the pre-flash distillation column is further heated by two furnaces before entering the main fractionating distillation column. In the main column, feed is distilled into multiple products based on the difference in their boiling temperatures. The distilled products are collected at the column's top, middle, and lower sections and are further processed to produce motor gasoline, kerosene, and diesel oils.

This refining process is characterised by its relatively slow dynamics (settling time is approximately 1 h), large number of process variables and interactions between the various sections of the process due to recycle flows and heat integration. An installed MPC system provides accurate control of critical product specifications, minimises the impact of process disturbances, and drives the process to operate closer to its economic optimum point. Six inferential models (soft sensors) are implemented in the process to provide timely prediction of critical product specifications and to enable real-time control of these product specifications. The controller has 30 controlled variables, 17 manipulated variables and six disturbance variables, and is considered reasonably large with respect

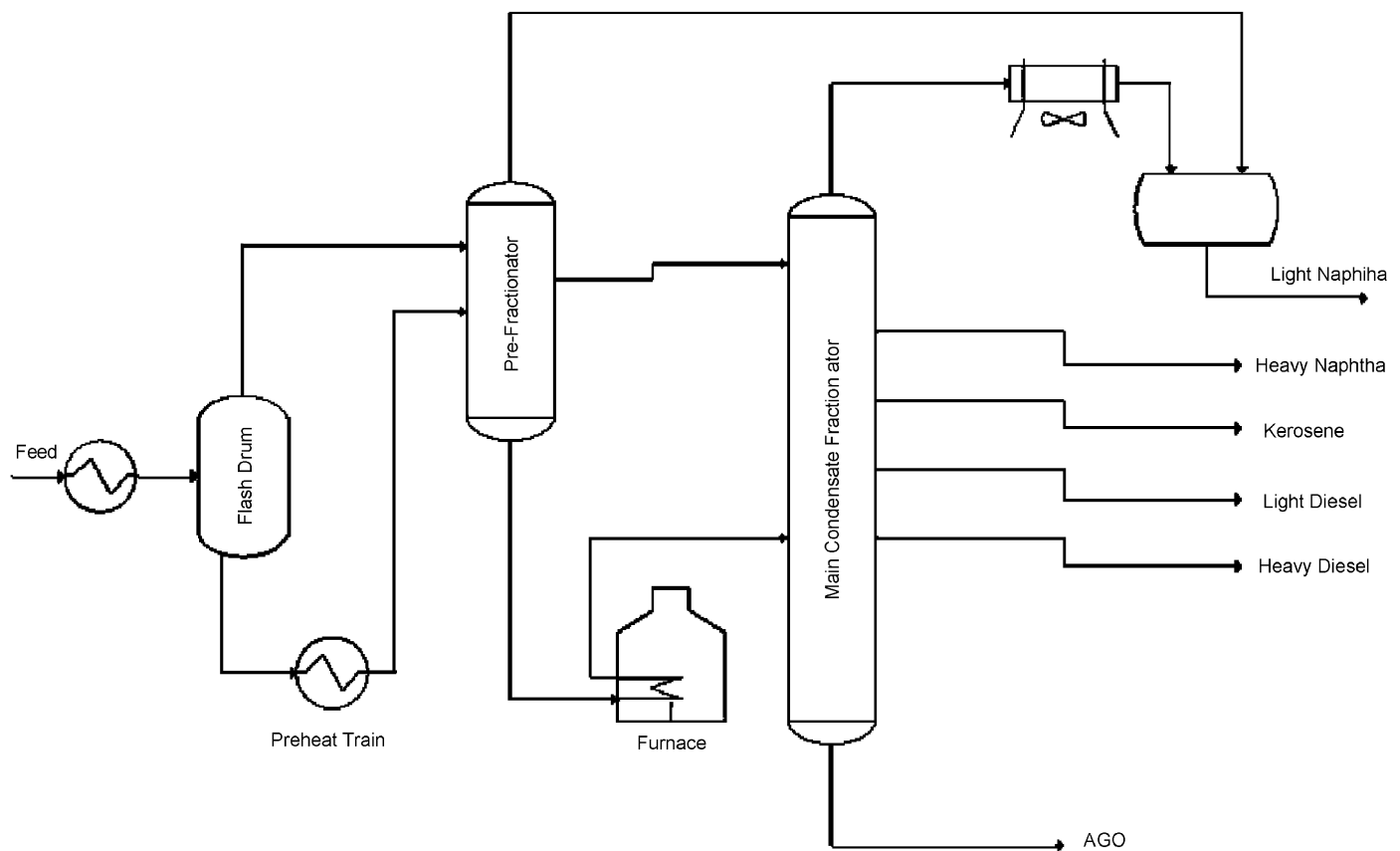


Fig. 1. Simplified flow diagram of the process.

to industrial standards. The primary controlled variables are the process feed flow, furnace outlet temperature, and several other key product specifications. Soft sensors are used to predict important product properties at the top of the pre-fractionator column, as well as top and side draw of the main condensate fractionator. The main objective of the MPC is to maximize valuable products production while maintaining product specifications. Typical process disturbances include changes in feed quality, blending flows between products and ambient temperature.

4. Methodology

This section will introduce the methodology followed in developing the condition monitor for the fractionation process. The basic methodology was based on that presented by Miletic et al. (2004), however, further recommendations are offered here that would help in developing industrial MSPC-based monitoring applications.

The application development methodology involved the following basic steps:

1. application design,
2. data collection and analysis, and
3. model development and proof of concept.

These steps are presented in detail below.

4.1. Application design

This step involved the specification of the objectives and scope of the study, in addition to developing a detailed understanding of the process operation. The primary aim of this study was to design and configure a general condition monitoring system for the fractionation process. It was intended that the monitoring system would provide plant operators with a single screen display which would indicate when the plant was behaving normally and detect and isolate any abnormality in process operation. Since the process is controlled by a multivariable model predictive controller, it was the intention that the monitoring system should be capable of monitoring the general operation of the MPC controller in addition to the process as a whole.

To develop an in-depth process understanding, flow diagrams and drawings were reviewed and several interviews were conducted with key facility staff. This step was considered to be critical for the project as it helped to identify the following objectives for the monitoring system:

- It should identify conditions that prevent the plant from meeting product specifications and maximising plant yield.
- It should provide an early indication of off-spec production as this has an impact on downstream process units.

The interviews also identified the key disturbances on the process and the frequency that these disturbances occur.

Following the interviews, a survey of the existing instrumentation on the process was carried out to identify the measurements that were available for use by the real-time condition monitoring system. Two approaches have been suggested in the literature to determine which of the available measurements should be utilised in the models (Miletic et al., 2004). These approaches recommend either including all process variables in the model or selecting a subset of variables that are considered to be important when assessing the overall performance of the process.

In this work, the second approach was adopted, where from an understanding of process operation, detailed interviews, and the survey of existing measurements, a subset of variables to be used in the development of the MSPC models was identified. In doing so, variables were classified as 'critical', 'important', and 'unimportant'. Only critical and important variables were chosen for use in the models. Identifying the process variables in this way was considered to be important because when all the variables were used in the model there were frequent occasions when the monitor would identify an abnormality, which could be traced to an unimportant condition. Such an alarm would be considered a false alarm by plant personnel and therefore undesirable. For example, measurements on flow lines leading to storage tanks, or blending streams are unlikely to have any significant effect on the plant operation and hence abnormal measurements here will not be important to plant personnel. By using only those variables that are considered to be important, or critical, it is more likely that any alarms raised by the monitoring system would be of interest to process operators and engineers.

In addition to the critical and important variables identified above, the cause and effect variables used in the MPC algorithm and the estimates made by several linear, soft sensors that were used in the controller were also included in the monitoring system. The inclusion of the MPC variables allowed a unified performance monitoring application to be developed that would monitor both the process and the general operation of the MPC system. It was found that by including many of the implicit variables from the MPC system, such as constraint limits and modelling errors, that a monitoring system could be constructed that provided detailed information relating to the performance of the MPC system. Unfortunately, interpretation of this system was considered to be too complex for operations staff and hence only the explicit MPC variables, i.e. controlled variables and manipulated variables, were used in this monitoring system. A monitoring system for use by control engineers is however being developed. This system will use all the available control system variables, together with the important process variables.

Since the soft-sensor models are linear combinations of other variables included in the monitoring system, they will

provide limited information. However, their inclusion was thought to be necessary as it might aid in the isolation of abnormal conditions. In total 164 process variables were used in the models developed in this study. This included the measurements used in the MPC system plus many other measurements that were not employed in the controller.

4.2. Data collection and analysis

Process data is the cornerstone of all data driven modelling techniques, thus the quality of process data used in developing the MSPC condition monitoring model is crucial to the success of the application. It is useful to consider the following questions when collecting data for modelling purpose:

- What is considered 'normal' process operation?
- How much data is available and how much should be used?
- What is the quality of the process data?
- Is the data collected from data historians compressed? If so, is it useable?
- How noisy is the collected process data?

Identifying what constitutes normal operation for a process is a challenging task, however, it can be simplified somewhat if accurate records are available of any major process upsets that affected the process during the period over which data is available.

Complex process units such as the one considered in this study have several operating modes, produce multiple products with various specifications and undergo routine maintenance tasks. Maintenance tasks are often carried out on critical process equipment while the process is still in operation. In this application, such tasks involved the removal of a section of heat-exchangers, which were taken out of service for cleaning and maintenance. Normal operation in the fractionator was maintained during this period through the use of a second heat-exchanger section. Since maintenance practices, such as this occur frequently, data collected during these operations should be considered to be 'normal'.

A further complication with identifying normal operating data is that the definition of normal operation can be interpreted differently by operating personnel thus it is recommended that this task be addressed during the personnel interviews to arrive at a common understanding of what should be considered 'normal' operation.

Another important factor to consider that is often overlooked when collecting process data for analysis is to identify periods when control valves are saturated, i.e. fully open or fully closed. Valve saturation will introduce non-linearity into the process behaviour and hence will affect the ability of the PCA model to identify relationships within the data. To eliminate any problems associated with saturated valves, all periods of data during which any of the control valves, identified as being critical, were outside

of their linear range, i.e. open less than 20% or more than 80%, were eliminated from the data. This will have the effect that the developed monitor will identify any periods of data when the valve is saturated as abnormal. This is desirable in this application as it warns process operators that the MPC system may be constrained and hence may not be performing optimally. A further consideration that was made when identifying normal operation was to remove sections of data when one or more control variables were outside of their upper and lower soft constraint limits. By using this data to identify the model, the monitoring system will detect when the control system is not behaving optimally, i.e. when the control variables are exceeding their limits.

Once the period of normal operation has been identified, data is collected from plant historians for the entire period of normal operation. The rate at which data is sampled is an important factor to consider. A high sampling frequency will enable abnormal conditions to be identified very quickly, however, it may also necessitate the need to consider serial correlations in the PCA model. Such correlations and how they impact on PCA applications is still the subject of research (see Chiang, Russell, & Braatz, 2002, for example). In the application studied here, discussions with plant personnel indicated that the detection of abnormal conditions during steady state, rather than transients was important and hence a 1-h sampling rate was considered to be appropriate. This sampling rate is comparable to the settling time of the system.

Following the identification of normal operating conditions, approximately 8 months, or 6000 h of process data was available for this study. Discussions with plant personnel indicated that disturbances to this process and changes in operating conditions are such that data collected from a period of approximately 4–6 months would be likely to cover all the typical operating conditions for this plant. Eight months of data is therefore sufficient to both develop and test a monitoring system. In many applications, effects such as those from changes in the season will often mean that significantly more data is required to thoroughly test any developed solution.

The process data used in this study was collected from a plant data historian. An issue that often affects data analysis and data-driven modelling techniques is the compression of data in plant historians. Although it is recommended, whenever possible, that uncompressed data be used for process modelling and analysis, practical limitations in computer and control systems data storage space dictates the use of reconstructed historical data obtained from historians. Commercial data historians use compression to facilitate the storage of large amounts of historical data. However, severe data compression causes a substantial loss of information from the data and may render the reconstructed data unusable for modelling purposes. The historian used in this study utilised the swinging door compression method (Matthew, Liakopoulos, Dragana, & Georgakis, 1998). Using this method, the

compression factor (CF) can be defined as the ratio between original data storage requirement and that of the archived data (Thornhill, Choudhury, & Shah, 2004). The degree of compression is set by the users, but can be estimated from the reconstructed data set. As a rule of thumb, Thornhill et al. (2004) suggests that a CF of 3 or less ($CF \leq 3$) is acceptable. All the data collected and used in this exercise had a CF of less than 1.5 ($CF \leq 1.5$) with the exception of several set-points that were rarely changed. Finally, process noise and outliers can also be a factor that affects data driven modelling techniques. In this work first order filters were used to remove the noise from some signals and the Hampel filter, by Pearson (2002) was found to be suitable for the detection and removal of outliers.

4.3. Model development and testing

The last step in developing the prototype monitoring system is the identification of the model and its subsequent off-line analysis. This step typically involves the investigation of how well the developed models are able to track process conditions and their ability to identify significant abnormalities and excursions. Control limits for both the SPE and T^2 charts should also be examined in this step so that appropriate levels can be specified to ensure that only significant abnormalities exceed the limits.

5. Condition monitoring application analysis and results

This section presents, in detail, the results obtained when several PCA techniques were developed and compared for the proposed monitoring application.

5.1. Static PCA models

For the condition monitoring application, a static PCA model was initially developed for the process. This model was developed using approximately 4500 sample points from the normal operating data, with the remaining 1500 being used for testing and validation. The model was identified using the algorithms and equations presented in Section 2. Fig. 2 shows the cross validation curve. This figure shows the error that is produced when the PCA model is used to reconstruct the values of unseen process data. In this example, the normal operating data was divided into 10 separate blocks, with nine blocks used to identify the model and one used to determine the reconstruction error. This process is repeated until the error has been determined over each of the 10 blocks and then these errors are summed together. Further details of cross validation techniques can be found in Wise and Gallagher (1996). The recommendation of Wise and Gallagher (1996) would be to use either five or seven principal components, as there is very little reduction in the error when more than this number of components was used. Many of the results obtained in this section were repeated using a variety of principal components. These

results indicated that when more than approximately 10 components were used, the monitor became very sensitive and identified many conditions, which although could be considered to be abnormal were of no interest to plant personnel. The results obtained with between four and eight principal components were found to be very similar. Discussions with plant personnel indicated that the seven component models provided marginally better results than the five component models. Hence all the models identified in this paper employed seven components. The fact that the monitoring results are robust to the number of principal components is encouraging as it is important that the

developed monitoring system should be robust to choices made by the developer.

Fig. 3 shows the T^2 and SPE charts that were produced when the normal operating data was passed through the identified PCA model. These charts show some violation of the control limits, for example, between sample points 200 and 400 on the T^2 chart and around sample point 2900 on the SPE chart. Investigation of each of the violations on these charts indicated operating conditions that would be considered to be abnormal. These conditions were not identified during the data collection and analysis step of this study and contained examples of the type of conditions that it was hoped the condition monitor would be able to detect. For example, Fig. 4 shows the contributions made by each of the process variables to the elevated SPE recorded at around sample number 2900. This chart highlights two variables as being the source of the problem, variables 46 and 47.

The raw measurement for variable 46, reflects the fuel gas heating value at the furnace. This value was found to increase significantly at sample number 2900. It is worth noting that the major contributing variables, as identified by the contribution chart, variable 47, is also a measurement of the fuel gas heating value, but this measurement is taken at a different process location. Significant variation in this value will act as a disturbance to the furnace and may ultimately affect other key process variables in the main fractionator. The early detection of this elevation in fuel gas heating value would help operators to provide manual feedforward control to compensate for this disturbance and hence improve the efficiency of the process. In this particular example, operations staff did

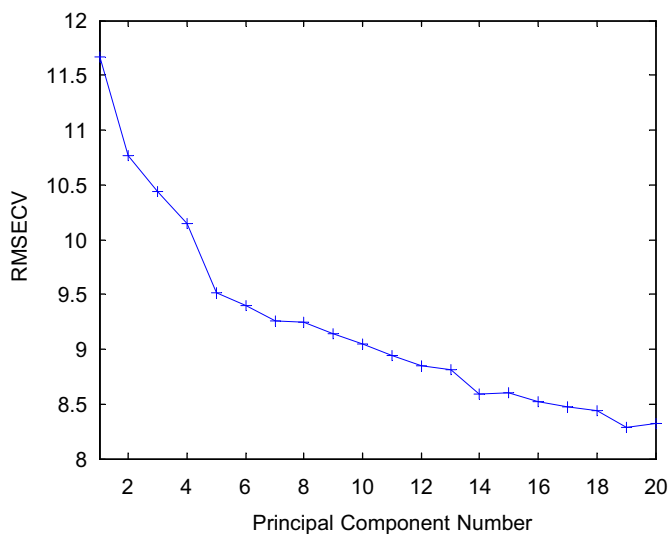


Fig. 2. Cross validation for static PCA model.

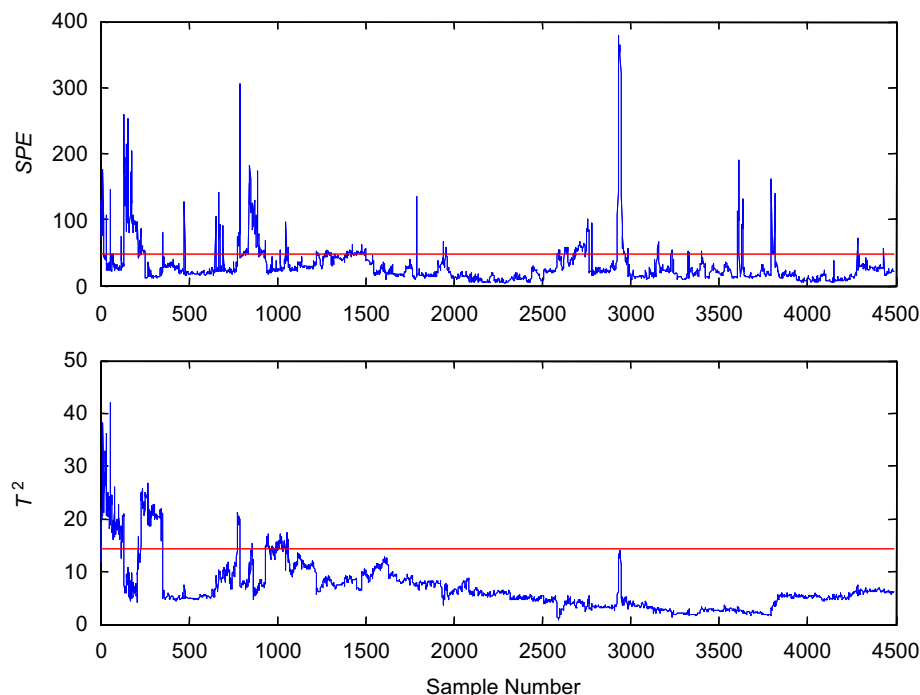


Fig. 3. MSPC charts for developed static PCA model for training data set.

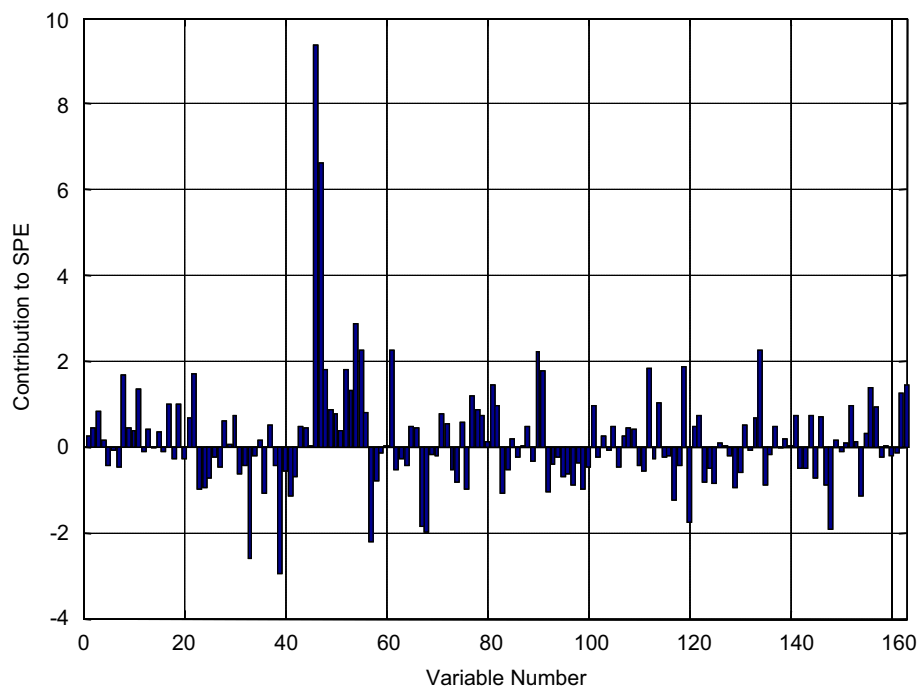


Fig. 4. Contribution plot.

detect and isolate the problem, however, the PCA model would have alerted the operations staff to the problem sooner.

Having identified several abnormal conditions, the data collected during these periods was removed and the PCA model re-calculated.

The above example indicates how the PCA model and data can be interrogated to reveal the likely cause of any abnormality. As discussed earlier in this paper, multiblock techniques have been developed to aid in the identification of the root cause of any abnormal condition and these techniques are considered appropriate if the process being monitored can be divided into specific processing units. For this application, the process can be divided into the pre-heat and pre-flash section and the main fractionator column sections and therefore a two-block PCA model was developed.

Multiblock PCA charts for blocks 1 and 2 are shown in Figs. 5 and 6 for the normal operating data used in the previous analysis. Note that the *super block* statistics are identical to those obtained using the static PCA model and are therefore not displayed. The advantage of the multiblock approach is that Figs. 5 and 6 give a clear indication to operators where the abnormality in the process occurs. Figs. 5 and 6 show that at sample number 2900, there is an increase in the *SPE* for block 2, but no significant increase in the *SPE* for block 1. Hence the process operators can quickly identify that there is an abnormal condition in block 2.

To test the suitability of using the PCA and multiblock PCA models as real-time condition monitors, they were each applied to the validating data, which contained data

collected under normal operating conditions. Fig. 7 shows the *SPE* and T^2 statistics that were determined for the static PCA model over this data set. The charts clearly show that although the T^2 statistic remains within its limits, the *SPE* statistic exceeds its limit significantly. Thorough analysis of this novel data set indicated that there were no particularly significant abnormalities present and further investigation showed that the reason for the elevated *SPE* value was because the model was not able to track the time-varying and non-stationary characteristics of the process. Fig. 8 illustrates how the process mean varies considerably over the combined training and evaluation data sets (the first 4500 data points represent the training data with the remainder, the evaluation data). Furthermore, the dynamics of the process change considerably with time. For example, when a heat exchanger is removed from service, or when there is a change in the characteristics of the feed stock to the process.

5.2. Recursive PCA

To address the limitation of the static PCA model developed in the previous section, a condition monitor based on RPCA techniques was developed. The RPCA model was based on the approach presented in Section 2.4. By updating the parameters of the PCA model, together with the corresponding control limits for the T^2 and *SPE* charts, changes in the correlation structure of the process can be accommodated by the model.

As discussed in Section 2.4, the RPCA algorithm used in this work calculates a new PCA model at each sampling instant using a moving window of data. The length of this

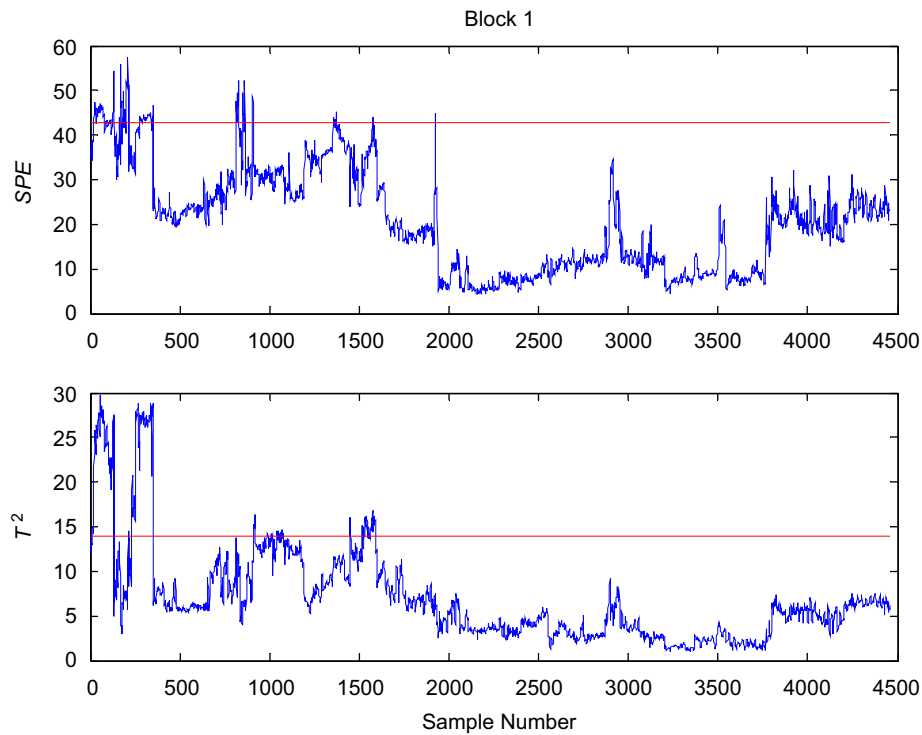


Fig. 5. Block 1 MBPCA charts.

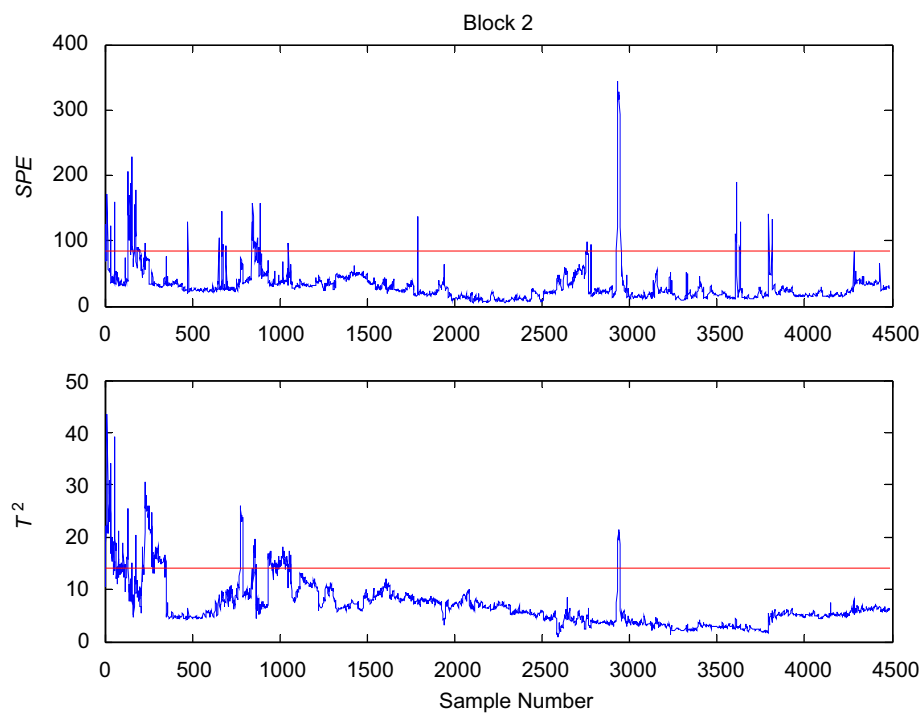


Fig. 6. Block 2 MBPCA charts.

window determines how quickly, or slowly the model will adapt to new process conditions. The performance of the RPCA algorithm was tested on the same data set used to develop and test the static PCA models in the previous section. The performance of the RPCA model was

examined using moving windows of data with varying lengths.

Figs. 9 and 10 show SPE and T^2 charts produced by the RPCA model using moving windows of length 5000 and 3000 samples, respectively. Note that the data set that was

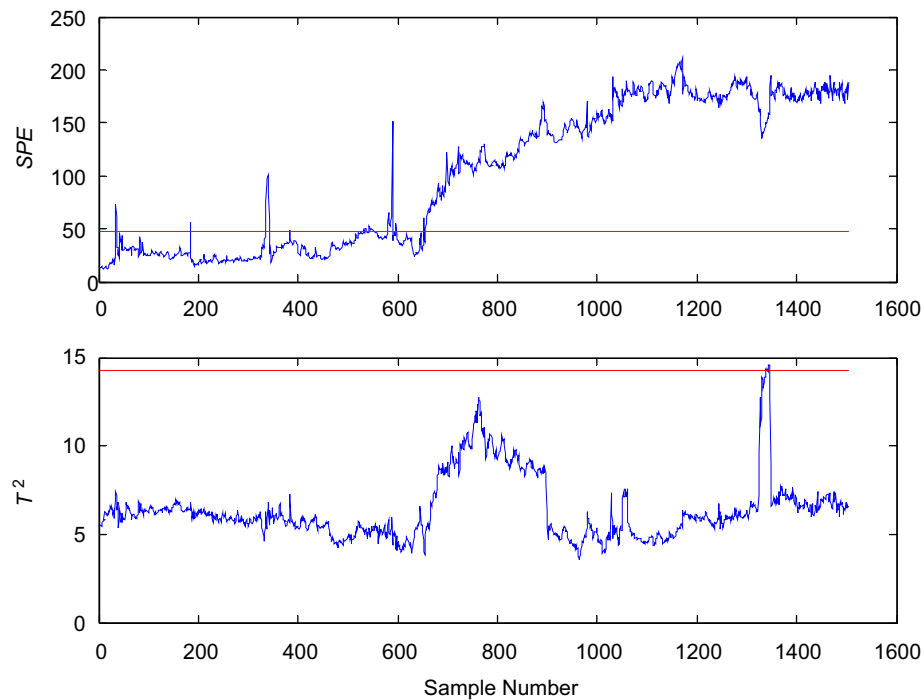


Fig. 7. Novel data set projected onto the static PCA model.

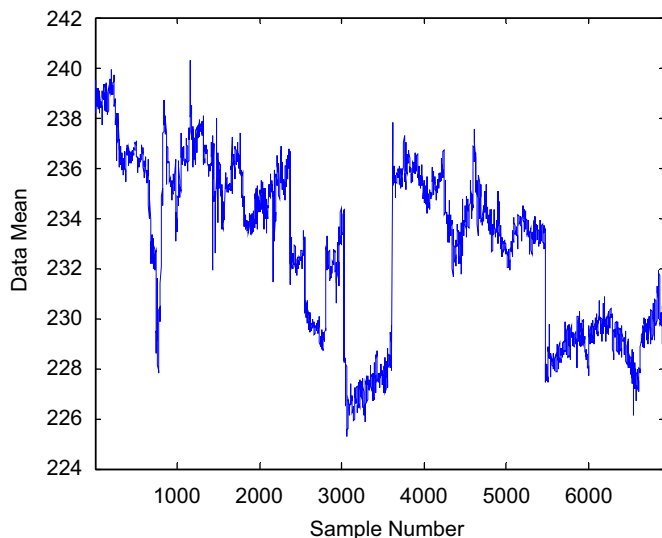


Fig. 8. Process data mean for the normal operating conditions.

used here was the combined training and evaluation data sets that were used in the static PCA tests, hence there are 6000 data points. With a sampling interval of 1 h, 5000 and 3000 samples corresponds to approximately 208 and 125 days of data, respectively. These can be considered to be relatively large moving windows, however, they were thought to be consistent with the rate at which the feed stock and other disturbances to the process occur.

The charts displayed in Figs. 9 and 10 illustrate that changing the length of the moving window had no significant impact on the charts. However, when the length of the moving window was specified to be much smaller

than 3000, 1000, for example, then the model adapted very quickly to any abnormal condition, and there was then the chance that it would be undetected. This issue is described in more detail in Wang et al.(2003).

Comparison of Figs. 9 and 10 with Figs. 3 and 7 indicates that the RPCA model is able to identify the abnormal operating conditions that were detected in the normal operating data, and unlike the static PCA model, it is able to provide a useful monitoring statistic for the novel data set. For example, the peak which occurs in the *SPE* statistic at approximately sample number 5200 was caused by the two variables shown in Fig. 11. The variable displayed in the left-hand side of Fig. 11 has moved into a slightly lower operating region, which is not a problem. However, the variable displayed in the right-hand side of Fig. 11 indicates a disturbance in the level at the base of the column, which would be of interest to the operators.

Finally, a recursive multiblock PCA (RMBPCA) model was developed and compared with the RPCA approach. Figs. 12 and 13 show blocks 1 and 2 charts respectively for this model when applied to the training and evaluation data sets. The RMBPCA model was able to identify the abnormality shown in the *SPE* charts of the RPCA model that occurred at sample 5200, and isolated block 2 as the source of the condition. Also, the RMBPCA model was able to identify and isolate other abnormalities occurring outside of the training data set such as that at sample 4600 in block 1 and sample 5800 in block 2. The key contributing variables for both these peaks were identified from contribution charts. Analysis of these variables showed that between sample numbers 4100 and 5100

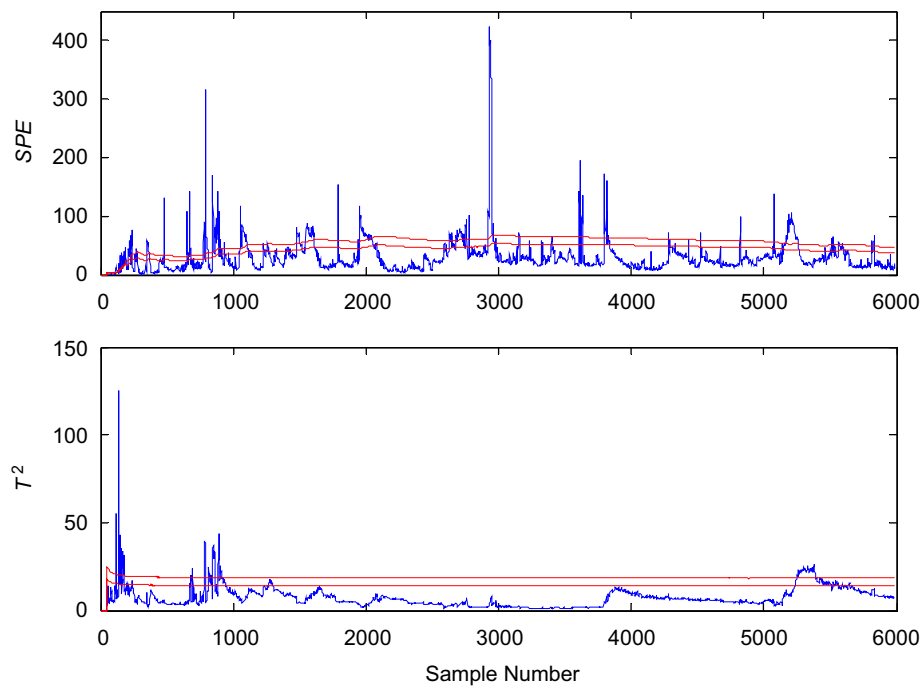


Fig. 9. RPCA model charts using a moving window width of 5000.

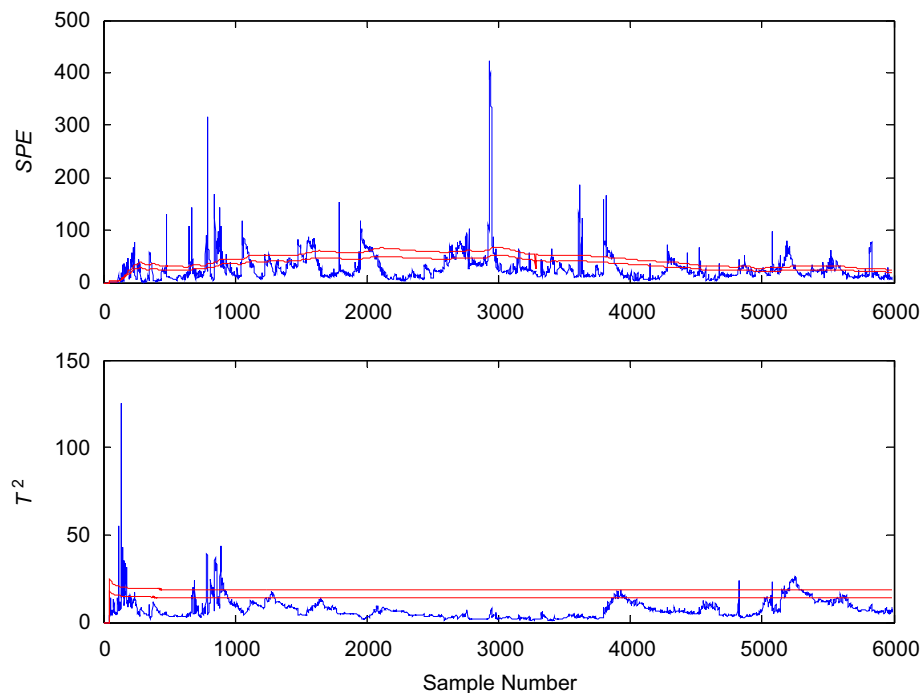


Fig. 10. RPCA model charts using a moving window width of 3000.

variable 14 reduced to an abnormally low value and variable 36 dropped significantly at approximately sample number 5800. Detailed analysis of the RMBPCA charts showed that it highlighted conditions throughout the data set that would have aided process operators in maintaining the operation of the unit at optimal conditions.

The results of this study are summarised in Table 1.

6. Conclusions and future work

This paper has presented the results from a detailed study into the development of a condition monitoring solution for application to a fractionation process. This process is a large and complex system with almost 200 process measurements and dynamics that vary with time.

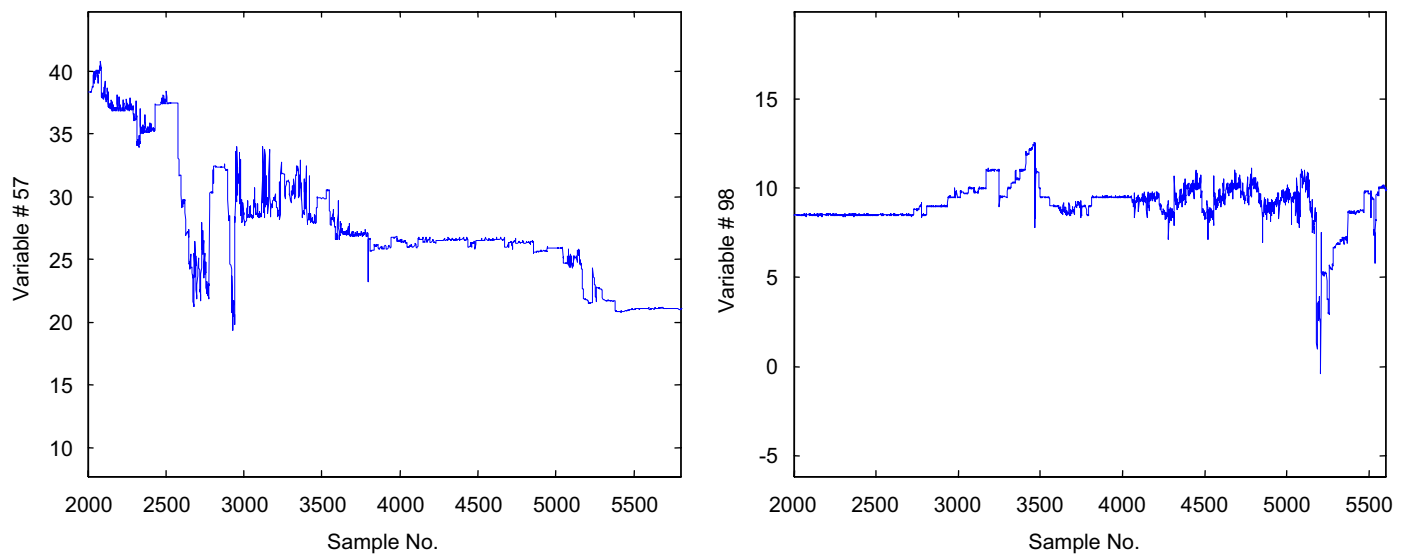


Fig. 11. Raw data.

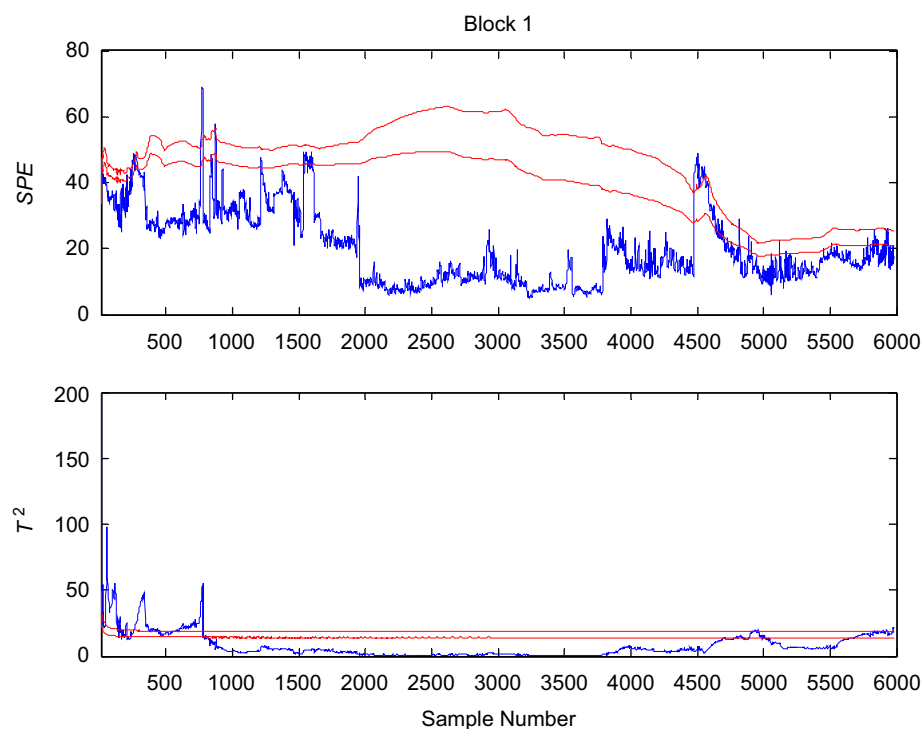


Fig. 12. RMBPCA for block 1.

The overall aim of the study was to develop a simple process monitoring system that would enable process operators to quickly and easily identify any sources of abnormality in the process.

The study showed that static PCA techniques were not suitable in this application as the process contained time-varying dynamics. Recursive techniques, however, were able to provide an accurate method for identifying abnormal conditions. Although it was possible for operators to identify these conditions without the aid of

the PCA monitoring system, the complexity of the process means that the monitoring system would certainly make it easier for operators to identify when the process is not behaving optimally.

As the process can be divided into two logical sections, multiblock and recursive multiblock techniques were also examined in this paper. The multiblock PCA models helped to quickly identify and isolate the section of the plant where the abnormality occurred. This approach is recommended for complex process units that can be

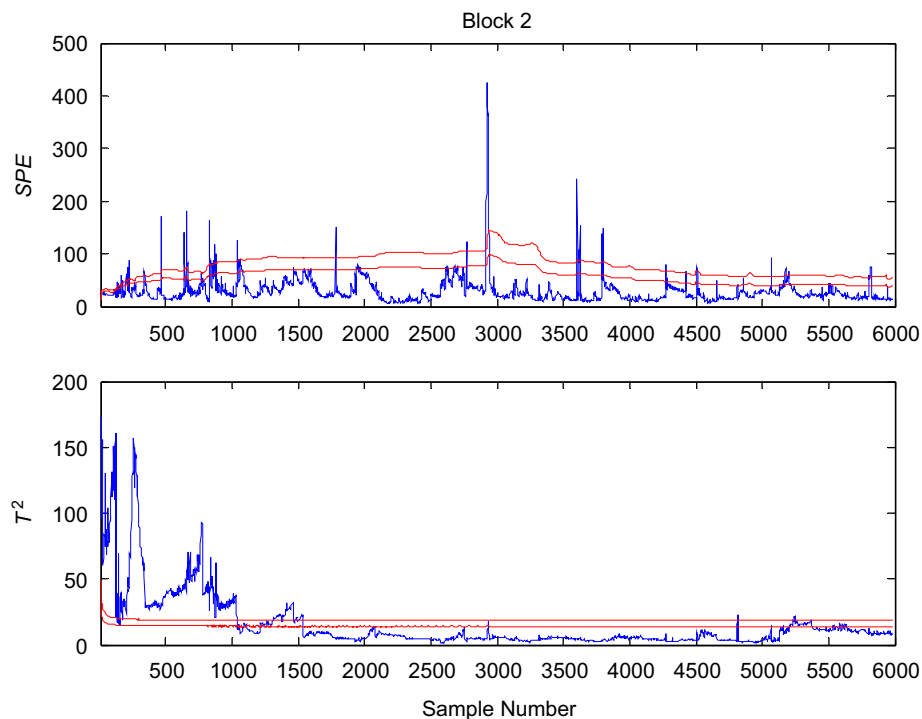


Fig. 13. RMBPCA for block 2.

Table 1
Summary of different modelling structures

Algorithm	Advantages	Disadvantages
Static PCA	Contains few user defined parameters	Unable to cope with time-varying characteristics of the process and hence unsuitable for the fractionation process
Static multiblock PCA	Allows the root cause of any abnormality to be isolated to the relevant section of the process	Unable to cope with time-varying characteristics of the process and hence unsuitable for the fractionation process
Recursive PCA	Able to adapt to time-varying characteristics	Requires that process variables be divided into separate blocks which may in some situations be difficult Requires an extra 'tuning' parameter (length of the window) to be specified by the user
Recursive multiblock PCA	Allows the root cause of any abnormality to be isolated to the relevant section of the process Able to adapt to time-varying characteristics	Requires that process variables be divided into separate blocks which may in some situations be difficult Requires an extra 'tuning' parameter (length of the window) to be specified by the user

divided into separate logical sections, and when the number of process variables included in the condition monitoring application is large. As with the standard PCA

approach, multiblock PCA was found to be unable to track the time-varying characteristics of this process and hence a recursive multiblock PCA approach was deemed to be the most suitable algorithm to apply to the process studied in this work.

The results of this study are such that Saudi Aramco now has confidence that the monitoring system can provide benefits in real time. Further work is now being undertaken to identify a software package that can be configured to interface with existing plant systems, so that the monitoring solution can operate in real time.

A complex issue which arose from this study was in monitoring the performance of the MPC system. The use of multivariate statistical techniques to provide a monitoring solution for this controller is the subject of current research.

References

- Badcock, J., Bailey, T. C., Jonathan, P., & Krzanowski, W. J. (2005). To projection methods for use in the analysis of multivariate process data with an illustration in petrochemical production. *Technometrics*, 46(4), 392–403.
- Bharati, M., & MacGregor, J. F. (2003). Softwood lumber grading through on-line multivariable image analysis. *Industrial and Engineering Chemistry Research*, 42, 5345–5353.
- Chiang, L. H., Russell, E. L., & Braatz, R. D. (2002). *Fault detection and diagnosis in industrial systems*. London: Springer.
- Dayal, B., MacGregor, J., Taylor, P., Kildaw, R., & Marcikic, S. (2004). Application of feedforward neural networks and partial least squares regression for modelling Kappa number in a continuous Kamyr digester. *Pulp & Paper Canada*, 95(1), 26–32.
- Eriksson, L., Johansson, E., Kettaneh-Wold, N., Trygg, J., Wikstrom, C., & Wold, S. (2005). *Multi- and megavariate data analysis*. Umetrics.

- Geng, Z., & Zhu, Q. (2005). 'Multiscale nonlinear principal component analysis and its application for chemical process monitoring. *Industrial and Engineering Chemistry Research*, 44(10), 3585–3593.
- Jackson, J. (1991). *A user's guide to principal components*. New York: Wiley.
- Kourti, D. (2005). Application of latent variable methods to process control and multivariate statistical process control in industry. *Journal of Adaptive Control and Signal Processing*, 19(4), 213–246.
- Kresta, J. V., MacGregor, J. F., & Marlin, T. J. (1991). Multivariate statistical monitoring of processes. *Canadian Journal of Chemical Engineering*, 69(1), 35–47.
- Lane, S., Martin, E. B., & Morris, A. J. (2001). Performance monitoring of a multi-product semi-batch process. *Journal of Process Control*, 11(1), 1–11.
- Lennox, B., Montague, G., Hiden, H., Kornfeld, G., & Goulding, P. (2001). Process monitoring of an industrial fed-batch fermentation. *Biotechnology and Bioengineering*, 74(2), 125–135.
- Li, W., Yue, H., Valle-Cervantes, S., & Qin, S. (2000). Recursive PCA for adaptive process monitoring. *Journal of Process Control*, 10, 471–486.
- Marjanovic, O., Lennox, B., Sandoz, D., Smith, K., & Crofts, M. (2006). Real-time monitoring of an industrial batch process. *Computers and Chemical Engineering*, 30(10–12), 1476–1481.
- Martin, E., Morris, J., & Lane, S. (2002). Monitoring process manufacturing performance. *IEEE Control Systems Magazine*, 10(5), 26–39.
- Matthew, J. W., Liakopoulos, A., Dragana, B., & Georgakis, C. (1998). A practical assessment of process data compression techniques. *Industrial and Engineering Chemistry Research*, 37, 267–274.
- Miletic, S., Quinn, S., Dudzic, M., Vaculik, M., & Champagne, M. (2004). An industrial perspective on implementing on-line applications of multivariate statistics. *Journal of Process Control*, 14, 821–836.
- Nimmo, I. (1995). Adequately address abnormal operations. In I. Nimmo (Ed.), *Chemical engineering progress*, September 1995 (pp. 36–45).
- Pearson, R. (2002). Outliers in process modeling and identification. *IEEE Transactions on Control Systems Technology*, 10(1), 55–63.
- Qin, J., & Badgewell (2003). A survey of industrial model predictive control technology. *Control Engineering Practice*, 11(7), 733–764.
- Qin, S. J., Valle, S., & Piovoso, M. J. (2001). On unifying multiblock analysis with application to decentralized process monitoring. *Journal of Chemometrics*, 15(9), 715–742.
- Sandoz, D. J. (2003). The exploitation of adaptive modeling in the model predictive control environment of Connoisseur. In V. VanDoren (Ed.), *Techniques of automatic control*. Butterworth.
- Thornhill, N., Choudhury, M., & Shah, S. (2004). The impact of compression on data-driven process analyses. *Journal of Process Control*, 14, 389–398.
- Venkatsubramanian, V., Rengaswamy, R., Yin, K., & Kavuri, S. N. (2003). A review of fault detection and diagnosis; part I: Quantitative model-based methods. *Computers & Chemical Engineering*, 27, 293–231.
- Wang, X., Kruger, U., & Lennox, B. (2003). Recursive partial least squares algorithms for monitoring complex industrial processes. *Control Engineering Practice*, 11(6), 613–632.
- Westerhuis, J., Kourti, T., & MacGregor, J. (1998). Analysis of multiblock and hierarchical PCA and PLS models. *Journal of Chemometrics*, 12, 301–321.
- Wise, B., & Gallagher, N. (1996). The process chemometrics approach to process monitoring and fault detection. *Journal of Process Control*, 6, 328–348.
- Wold, S. (1994). Exponentially weighted moving principal components analysis and projection to latent structures. *Chemometrics and Intelligent Laboratory Systems*, 23, 149–161.
- Wold, S., Esbensen, K., & Geladi, P. (1987). Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 2, 37–52.
- Yu, H., & MacGregor, J. F. (2003). Multivariate image analysis and regression for prediction of coating content and distribution in the production of snack foods. *Chemometrics and Intelligent Laboratory Systems*, 67, 125–144.
- Zhang, Y., & Dudzic, M. S. (2006). Industrial application of multivariate SPC to continuous caster start-up operations for breakout prevention. *Control Engineering Practice*, 14(11), 1357–1375.