COMPARISON OF MULTIVARIATE AND UNIVARIATE STATISTICAL PROCESS CONTROL AND MONITORING METHODS

R.P. LEGER[†], WM.J. GARLAND[†], J.F. MACGREGOR[‡], W.F.S. POEHLMAN[*]

[†] Department of Engineering Physics, McMaster University, Hamilton, Ontario, Canada, L8S 4M1
[‡] Department of Chemical Engineering, McMaster University
[*] Department of Computer Science & Systems, McMaster University

Work in recent years has lead to the development of multivariate process monitoring schemes which use Principal Component Analysis (PCA). This research compares the performance of a univariate scheme and a multivariate PCA scheme used for monitoring a simple process with 11 measured variables. The multivariate PCA scheme was able to adequately represent the process using two principal components. This resulted in a PCA monitoring scheme which used two charts as opposed to 11 charts for the univariate scheme and therefore had distinct advantages in terms of both data representation, presentation and fault diagnosis capabilities.

Key Words: statistical process control, principal component analysis, cumulative summation control chart, covariance matrix, fault detection and diagnosis

1. INTRODUCTION TO STATISTICAL PROCESS CONTROL

The reliability of a manufacturing process is becoming an increasingly important element of the operation as efforts are made to reduce cost. Key to improving the reliability of a process is promptly detecting and diagnosing faults in the process which cause the process variables to move away from their desired values. One methodology used to detect process faults is Statistical Process Control (SPC). SPC involves setting up control charts which are used to monitor the process variables for faults. The control charts use control limits which are based on the inherent or "common cause" variation which affects the process variables at all times. This inherent or natural variability is considered a natural part of the process which cannot be eliminated. The task of the control charts is to distinguish between the natural variation in the process which cannot be avoided and faults which have an assignable cause. The criteria for measuring the performance of the control charts is based on two types of errors. If the control chart indicates a fault is present when the process is in control, a false alarm has occurred. This type of error is known as a Type I error. If the control chart fails to detect a fault which is actually present, a Type II error has occurred. The ideal control chart scheme will minimize both types of errors.

When SPC was first introduced in the 1930's, typically, there were very few variables being measured on the process. The measured variables were usually product quality variables and the assumption was made that the variables were independent of each other. Hence, univariate control charts were set up for each of the measured variables. The introduction of computers and sophisticated, high speed data acquisition systems on the shop floor has brought about a major change in the data available for SPC techniques. Now, data are available on hundreds or perhaps thousands of process variables as well as the product quality variables. Faults will influence this process data set as well as the quality data. The process data set has many useful characteristics for fault detection and diagnosis including being measured very frequently and precisely in an on-line manner. However, the process data set also has drawbacks. First, the fact that hundreds of variables are being measured very frequently causes the dimensionality of the problem to become unmanageable. Also, all the measured process variables are not independent. Typically, there are only a few underlying events driving the process and each measured variable gives a little different information on the events. This causes the rank of the data matrix to be less than the number of variables and causes computational difficulties.

One method for dealing with this large number of correlated variables is to reduce the dimension of the problem using new multivariate SPC methods. Two multivariate methods which have received much attention recently are Principal Component Analysis (PCA) and Partial Least Squares (PLS) analysis. These methods break the data set down into uncorrelated variables, or principal components, which are monitored for assignable cause events.

Regardless of whether univariate and multivariate SPC methods are used to monitor the process, the general method used to develop the monitoring scheme is the same. First, historical data is collected from the process when operating normally. It is important at this step to remove any data which represent faults that should be detected in the future. Therefore, the data used to develop the monitoring scheme should contain only inherent variability. Next, a statistical model (univariate or multivariate) is developed which accurately describes this process data. Finally, new data can be compared to the model to determine if the process is continuing to operate normally or if there is a fault present.

This paper will present a comparison of univariate and multivariate SPC methods applied to a simple process. Section 2 will describe the process used for the research. Section 3 will describe the traditional univariate approach which was applied to the process. Section 4 will provide a brief description of the method of PCA, how PCA can be applied to process monitoring and how the method was applied to the process. Section 5 will compare the results of the two monitoring procedures and Section 6 will give some conclusions and possible areas for future work.

2. EXPERIMENTAL PROCESS

In order to do a comparison of univariate and multivariate monitoring methods, a simple test bed process was required. The process used was a simple model of the heat transport system of a CANDU nuclear reactor. A diagram of the model is shown in Figure 1.

The loop can be divided into two sections, primary side and secondary side, similar to a nuclear reactor. The primary side consists of water being pumped in a figure of eight loop. As the water flows through the core it is heated by the pipes electrically. The water then flows up through one cooling tower around the U-tube and down through the second cooling tower. It then flows through a pump and into a second core section. As observed from Figure 1, the flow through each of the two core sections is in opposite directions and the loop is symmetric. The primary side flow is identified as the dark solid lines. The secondary side is defined as the cooling water side. The cooling water enters the bottom of four cooling towers, flows upwards removing heat from the primary side and exits at the top of each tower. The secondary side flow is identified as the light dashed lines in Figure 1.

There were a total of 11 variables measured on the model. On the primary side, six variables were measured, four temperatures and two flow rates. T2(x_2) and T4(x_4) measure the temperatures at the core inlets and T3(x_3) and T5(x_5) measure the temperatures at the core outlets. There are two flow orifices and pressure transducers, F1(x_{10}) and F2(x_{11}), located immediately after the two pumps. On the secondary side, there were five variables measured, all temperatures. T1(x_1) measures the inlet cooling water temperature. T6(x_6) and T7(x_7) measure the outlet temperatures from towers 1 and 2 respectively, while T8(x_8) and T9(x_9) measure the outlet temperatures from towers 4 and 3 respectively. The exact locations of all 11 sensors are shown in Figure 1.

The data acquisition system for the model was set up to collect measurements from the 11 sensors every 0.25 seconds. The measurement data was written to a file on a PC hard disk in binary format. The program created a new data file every five minutes.

3. UNIVARIATE METHODOLOGY

3.1 Types of Univariate Control Charts

There are several different choices for univariate control charts, based on the type of data available from the process. Two of the more common types for used for monitoring the variable mean or target are the Shewhart Control chart and the cumulative summation (CUSUM) control chart. The Shewhart chart plots successive sample averages and typically has control limits set at $\bar{\bar{x}} \pm 3\sigma_{\bar{x}}$, where $\bar{\bar{x}}$ is the overall average of the sample averages and $\sigma_{\bar{x}}$ is an estimate of the standard deviation of the sample averages. These charts are effective in quickly detecting large mean shifts, on the order of 1.5 to approximately 2 standard deviations. However, they are relatively insensitive to persistent moderate shifts in the mean, on the order of 1-standard deviation.[1] Quite often, these types of shifts are common and are a first indication that a fault has occurred. Therefore, it is desirable to detect these shifts promptly and accurately. A popular type of chart which is sensitive to moderate persistent changes is the cumulative sum (CUSUM) control chart. This chart was first introduced by E.S. Page in 1954.[2] As the name implies, this type of chart cumulates deviations of the sample averages from the target or desired value. Once these cumulations reach either a high or low limit, an out-of-control signal is given. The ability of CUSUM charts to detect moderate faults provided the justification for their use as the univariate method for this research. The next section will cover the basic CUSUM chart scheme.

3.2 CUSUM Control Chart Scheme

A typical CUSUM control chart scheme is shown in Figure 2. As observed, it consists of two charts, a run chart plotting the successive differences between the sample average and target, (x-μ), and the control chart. The parameters shown on the control chart, are defined as follows:

k : the threshold for cumulation, which can be defined as the minimum difference between sample average and target that will cause the cumulation to begin. This value is also sometimes referred to as the allowable slack in the process. Typically, k will be set equal to one half of the deviation from target which is to be detected quickly.[3]

SH_i and SL_i : the high side and low side cumulation terms. Sample averages which are above k are added to the cumulation terms. Sample averages which are below k are subtracted unless the cumulation terms are already zero.

h : the control limit. If either SH or SL cumulate above h, intervention in the process is required.

The control limit, h, is determined by minimizing the number of false alarms, Type I errors and minimizing the time required to detect the deviation from target which should be detected quickly.

4 MULTIVARIATE METHODOLOGY

4.1 Introduction to Principal Component Analysis (PCA)

PCA is a technique for transforming a group of correlated variables via linear transformations into a new group of uncorrelated variables. PCA can also used to reduce the dimension of a data matrix. The purpose of this section will be to introduce the basics of PCA. This will be done by reviewing a simple 2-dimensional example taken from Jackson.[4] The two variables, X_1 and X_2, are plotted in Figure 3.

A typical analysis used to describe this data would be a least squares linear regression. The two lines associated with the least squares fit are shown in Figure 3. However, one may want to do the prediction in either direction, that is, consider the two variables as interchangeable. In this case, an orthogonal regression line is required. An orthogonal regression line minimizes the deviations perpendicular to the line itself. This line is also shown in Figure 3 and is known as the first principal component of the data set. The position of this line is calculated by examining the covariance matrix of the data set. The covariance matrix is used because it is a measure of the variability in the data set which the principal component is attempting to explain. This calculation will be discussed in detail below.

The method of PCA is based on the matrix result that a symmetric, nonsingular matrix, such as a covariance matrix, can be reduced to a diagonal matrix, as follows:

$$U^T S U = L \tag{1}$$

where: S - covariance matrix $= \begin{bmatrix} s_{11}^2 & s_{12}^2 \\ s_{21}^2 & s_{22}^2 \end{bmatrix}$

where: s_{11}^2 = variance of variable 1 $= \frac{1}{n-1}\sum_{i=1}^{n}\left(x_{1,i} - \bar{x}_1\right)^2$

s_{12}^2 = covariance between variable 1 and variable 2 $= \frac{1}{n-1}\sum_{i=1}^{n}\left(x_{1,i} - \bar{x}_1\right)\left(x_{2,i} - \bar{x}_2\right)$

L - diagonal matrix containing the eigenvalues of S

U - columns of U are the eigenvectors of S

U^T - transpose of U

The elements of U are also the directional cosines relating the new axis to the old axis. In the case of the test data:

$$S = \begin{bmatrix} 0.7986 & 0.6793 \\ 0.6793 & 0.7343 \end{bmatrix} \tag{2}$$

$$U = \begin{bmatrix} 0.7236 & -0.6902 \\ 0.6902 & 0.7236 \end{bmatrix} \tag{3}$$

Figure 4 shows the lines representing the two principal components and the angles defined by the cosines in U.

The position of each data point on the new principal component axis can be calculated by:

$$t = \left[X - \bar{X}\right] * U = \begin{bmatrix} (x_{1A} - \bar{x}_1) & (x_{2A} - \bar{x}_2) \\ (x_{1B} - \bar{x}_1) & (x_{2B} - \bar{x}_2) \\ . & . \\ . & . \\ . & . \end{bmatrix} * \begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{bmatrix} = \begin{bmatrix} t_{1A} & t_{2A} \\ t_{1B} & t_{2B} \\ . & . \\ . & . \\ . & . \end{bmatrix} \tag{4}$$

where A and B represent new observations of all variables. The individual t values are referred to as scores and are calculated as:

$$t_{i,A} = \left(x_{1A} - \bar{x}_1\right) * u_{1i} + \left(x_{2A} - \bar{x}_2\right) * u_{2i} \qquad \text{where } i=1,2 \tag{5}$$

Rearranging equation 4 by post multiplying by U^T shows that each original variable is made up of a linear combination of the principal components :

$$X = t * U^T + \bar{X} \quad \text{or} \quad \hat{x}_{1A} = \left[t_{1A} * u_{11}^T + t_{2A} * u_{21}^T\right] + \bar{x}_1 \tag{6}$$

The calculation of x_{1A} will only be an estimate if not all of the principal components are used in the calculation. This leads to a squared prediction error for the A^{th} multivariate observation, calculated as follows:

$$SPE = \sum_{i=1}^{\#var}\left(x_{i,A} - \hat{x}_{i,A}\right)^2 \tag{7}$$

PCA has some interesting properties which can be used in conjunction with process monitoring. Two of these properties are:

1. Trace(S) = Trace(L)

2. l_1 = first element in the diagonal matrix L = variance of PC1.

If the sum of the variances of all the variables is used as a measure of the overall variability of the data set, the eigenvalues of the data set may be thought of as variance components. In this respect, the ratio $\frac{l_1}{\sum l}$ = portion of the total variability accounted for by PC1. One note should be made with regard to property 1. If the data matrix contains variables in different units, for example temperatures and pressures, then the trace of S will have no meaning because numbers with different units are being added. Typically, process data matrices will contain different units. The solution to this problem is to make the numbers unitless. This is done by dividing all values for each variable by the standard deviation of that variable. This transforms the covariance matrix into the correlation matrix. If the correlation matrix is used instead of the covariance matrix, the sum of the variances will equal the number of variables. The next section will outline how these properties are used for process monitoring.

4.2 PCA Methodology for Multivariate SPC

The main characteristic of PCA used to monitor processes in a multivariate fashion is its ability to adequately represent the data in a reduced dimension space. As stated in Section 1, the monitoring scheme, and hence PCA model, is based on historical data collected when the process was operating in control. Therefore, the principal components will be modeling common cause variation. However, typically the first few eigenvalues will be large, well separated and account for the greater part of the variability. These eigenvalues and principal components represent variability which can be attributed to natural correlations which are present in the data. The remaining eigenvalues are usually small and close to the same value. There are several different tests available to determine how many principal components should be retained in a model. These include Barlett's significance test, and cross validation methods.[4] The process can be monitored by tracking two items; the movement of the process in the plane (or hyperplane if more than two principal components are used) defined by the principal components used in the model and the SPE as calculated by equation 7. A typical example of these charts is shown in Figure 5. The generation of the t-plot shown in Figure 5 is sometimes referred to as a projection because the original variables are "projected" on to the lower dimensional plane defined by the principal components.

The control contour for the t-plot is defined by Hotelling's T^2 statistic based on the principal components retained in the model[15]:

$$T_P^2 = \sum_{i=1}^{P} \frac{t_i^2}{l_i} \qquad (8)$$

where: P = number of retained principal components

Equation 8 represents an ellipse when two principal components are used. For a given desired confidence level for Type I errors (α), the control limit for T_P^2 based on the F-distribution is calculated as follows[14]:

$$T^2_{\alpha,P,n} = \frac{P(n-1)}{n-P} * F_{\alpha,P,n-P} \qquad (9)$$

where: α = desired confidence level
n = number of measurements used to develop the model

P = number of retained principal components

The upper limit for the squared prediction error chart is calculated using a method described in Jackson and is often referred to as a Q-statistic.[4] The method uses the sum, sum of the squares and sum of the cubes of the eigenvalues of the principal components not included in the model. Note that if all the principal components are used, there will be no SPE for the development data. This is similar to fitting N data points with an N^{th} order polynomial.

Using the t-plot and SPE plot in combination can provide an effective monitoring scheme. There have been several papers written on this subject.[6,7,8,9] If the fault is a new event which was not included in the development data set, the relationship between the variables will be changed and the covariance structure will be changed. This will cause the new observation to move away from the defined plane and will be detected by a high value of the SPE. If the fault causes larger than normal variations in the principal components used in the model but the basic relationship between the variables described in the development data does not change, it will be detected in a shift in the t-plot.

In the next section, the results of developing univariate CUSUM chart control schemes and a multivariate PCA control scheme will be discussed and compared.

5.0 RESULTS

5.1 Model Development

5.1.1 Data Collection

The first step in developing a statistical model is to collect some historical data from the process. As stated above, it is very important that the historical data contain only inherent process variability. In this study, historical data was collected from the model loop on four different days. The data was collected after the loop was allowed to reach steady state for a given power level. In total, steady state data was collected for 75 minutes. It was decided that the data acquisition rate of every 0.25sec. was quicker than required for this study. Therefore, the statistical models were based on a new data sample every 10 sec. The raw data collected during the 10 second intervals was averaged to provide one data point. Thus, the 75 minutes of steady state data provided 450 data points which were used to build the statistical models.

5.1.2 CUSUM Control Scheme

The CUSUM control schemes for this data were presented by Leger at the CNS Simulation Symposium held in Oct/95.[10] The results will be reviewed briefly here.

Three quantities are required to design a CUSUM scheme: the target values for each analyser, μ, the standard deviation for each analyser, σ, and the shift in mean which is desired to be detected quickly, Δ. The first two parameters, means and the standard deviations for each analyser, were calculated from the steady state data. Their values are shown in Table 1. The third parameter required, Δ, was set equal to $2*\sigma$.

The CUSUM schemes where then set up according to the four steps listed below:

1. threshold = $k = \Delta/2 = \sigma$
2. $k^* = k/\sigma = 1.0$
3. chose $h^* = 3.5$ to give ARL(0) = 2670, ARL(Δ) = 4.25
4. control limit = $h = h^* * \sigma$

In the above steps, the ARL stands for Average Run Length which is defined as the average number of samples taken before the control chart gives an "out of control" signal indicating that an intervention must be made in the process. Ideally, ARL(0), which is the ARL between false signals should be as large as possible. Also, ARL(Δ), which is the ARL for a process shift of Δ, should be as small as possible. The ARL values given above were obtained from tables found in Marquardt.[1] Given these ARL values and that data points were available every 10 seconds, one would expect a false alarm when the process is exactly on target every 7.4 hours and a shift of Δ should be detected in 42.5 seconds.

The above schemes were tested with the steady state data to determine the number of false alarms which would occur. Considering that there was only 75 minutes of steady state data, at most, one false alarm for each analyser would be expected. When the schemes were applied to the data, over 600 false alarms were encountered. The source of the false alarms was determined to be shifts in the average day to day mean values for each analyser which caused the data to be autocorrelated. CUSUM charts are only statistically justifiable for independent, normally distributed observations. Harris has shown that when faced with autocorrelated data, the ARL(0) drops rapidly as the autocorrelation increases[16]. However, the CUSUM chart can still be used as a monitoring method by modifying its scheme to reduce the number of Type I errors to an acceptable level for the specific case at hand. Lucus suggested doubling the value of k as one possible modification[3]. This will also increase the ARL for a true shift. For this investigation, the threshold limits were calculated by adding the chosen value of k to the maximum daily mean and subtracting k from the minimum daily mean. This in effect widened the threshold for cumulation and therefore increased the value of Δ for each analyser. The actual control limits were also doubled to $7.0 * \sigma$. This reduced the number of false alarms to 14, 13 of which were associated with one specific variable, x_1. Based on this observation, the number of false alarms was considered acceptable. The final CUSUM schemes are summarized in Table 2.

5.1.3 PCA Scheme

A PCA model was developed for the same steady state data described above using the MACSTAT code developed at McMaster University. Two important aspects of this code are how it calculates the principal components and what stopping criteria is used. The code calculates the principal components of a data set using the NIPALS(Nonlinear Iterative Partial Least Squares) routine.[11,12] In this routine, each eigenvector is calculated sequentially using an iterative algorithm.[12] Secondly, cross-validation, due to Wold, is used to help the user determine the number of principal components to be retained.[13]

For this project, the data matrix contained temperatures and flowrates. Therefore, the correlation matrix was used in all calculations. Using MACSTAT, it was found that the steady state data could be adequately represented with 2 principal components. The first and second eigenvalues were found to be 5.27 and 1.65 respectively. The cumulative variability explained by the first two principal components was:

$$\frac{5.27}{11} + \frac{1.65}{11} = \frac{SS_X - SS_{res}}{SS_X} = 63.1\% \qquad (10)$$

The calculation using the sum of squares (SS) is required because the NIPALS routine does not calculate the eigenvalues. Thus, using two principal components, the process could be monitored using one t-plot (t1 vs t2) and the SPE plot. Using this model yielded 6 false alarms for the steady state data. The SPE plot and t plot for the steady state data are shown in Figure 5.

5.2 Model Testing

The statistical models were tested in two respects. First, they were tested using new steady state data to determine the number of false alarms. Then, their response to specific faults was tested. Both of these tests will be discussed below.

A new steady state data set was collected from the process for 105 minutes over four different days. For this data, the 11 CUSUM charts recorded 55 false alarms, all associated with x_8. This might indicate that the CUSUM scheme for x_8 was too sensitive. The PCA model recorded 16 false alarms, of which 10 were again associated with x_8. Table 3 summarizes the number of Type I errors for both the development data the testing data for both the CUSUM charts and the PCA model.

In order to test the response on the models to actual faults, some test faults were designed. It was decided to use six different faults, described below:
1. 10% power increase, FAULT1
2. 10% power decrease, FAULT2
3. cooling water shut off to all four cooling towers, FAULT3
4. cooling water shut off to cooling towers 1 and 2, FAULT4
5. right hand side re-circulating pump valve open (valve 1), FAULT5
6. left hand side re-circulating pump valve open (valve 2), FAULT6.
Fault detection was considered a success if there were no Type II errors and the faults could be detected in a reasonable amount of time. In order to test the models, each fault was initiated in the process and data was collected. Two separate tests were completed for each fault. The results from these tests are shown in Table 4. As observed, the were no Type II errors, that is, all faults were detected. Also, all detection times were relatively short, with the longest time being 3.5 minutes for the detection of FAULT2 by the CUSUM charts.

5.3 CUSUM and PCA COMPARISON

The results of the previous section would indicate that in terms of Type I errors, the PCA methodology has a slight performance advantage. It should be noted that the control limits for the PCA charts were also expanded due to the autocorrelation in the data. This can be partially observed in Figure 5 where it is seen that the steady state data is roughly grouped into three to four clusters. These clusters would repersent the data from the four diferent days. In order the make the monitoring schemes more sensitive to faults of smaller magnitude, the data could be prepocessed to remove the effect of the autocorrelation. This could be done by monitoring the differences between the various temperatures and the cooling water inlet temperature as opposed to the absolute values. Another methos for dealing with the correlated observations would be the smooth the data with an exponentially weighted moving average[16]. In terms of Type II errors and fault detection times, the CUSUM charts and PCA charts are basically equivalent for monitoring the process. However, the PCA charts have a clear advantage over the CUSUM charts in two areas; data representation and presentation and fault diagnosis capability. These two areas will now be discussed in detail.

The advantage of the PCA charts in terms of data representation and presentation will be examined first. The SPE and t plots for the second test of FAULT1 are shown in Figure 6. As observed from these charts, it is very clear that a fault has occurred from approximately observation #38 onward. In terms of on-line monitoring, the PCA representation is much more concise and comprehensible as compared to monitoring 11 individual CUSUM charts. Even if only the CUSUM charts which were signaling alarms were presented to the operator, the situation would quickly become confusing during a fault because several charts would be present. This is due to the correlated nature of the data, described early. Therefore, the ability of PCA to represent the data in a reduced dimensionality is clearly demonstrated here as an advantage over the univariate CUSUM charts.

The second advantage of the PCA methodology is it's capability for providing a starting point for fault diagnosis. This is done by examining the underlying PCA model at the point where a fault is detected.[5] The examination can be presented to the operator in the form of contribution plots of both the SPE and t values. Examples of the contribution plots for FAULT1 are shown in Figure 7. As observed from Figure 7, the contribution plots show the contributions of each variable to overall SPE and t values. FAULT1 affects both the SPE and t values. From Figure 7a, the largest contributions to the SPE occur from the primary side core outlet temperatures, x_3 and x_5, and the cooling water inlet temperature, x_1. This indicates that there is an inconsistency among these variables; that is, the relationship or correlation among these variables has been broken. The model prediction for x_3 and x_5 is low while the prediction of x_1 is high. It should be noted that a positive contribution to the SPE results from a low prediction from the model, as stated in equation 15. These observations indicate a high power fault because the core outlet temperatures would be higher than expected and the cooling water temperature would not be correspondingly rising, as expected by the model. Figure 7b indicates that the observed shift t1, as shown in Figure 6, is caused by the primary side temperatures, x_2-x_5. This indicates that there has been a larger than normal shift in these variables. Once more, this would indicate that there is a problem on the primary side, possibly a power fault. It would be more difficult to extract this information from the 11 individual CUSUM charts in an on-line manner.

A similar analysis for the 5 other test faults was completed. These results are summarized as follows:
FAULT2: The contributions to t1 and the SPE involved the same variables as FAULT1 but were reversed, as expected.
FAULT3: For FAULT3, the contributions to the SPE or t plots did not reveal a clear diagnosis. This could be explained by the fact that all the variables were affected by shutting off the cooling water.
FAULT4: There were large positive contributions to the SPE from x_2, x_6 and x_7, indicating the prediction was low. There were large negative contributions from x_4, x_5, x_8, x_9 and x_{10} which indicates the prediction was high. Variables x_2, x_6 and x_7 are on the right hand side of the loop while x_4, x_5, x_8, x_9 and x_{10} are on the left hand side. This would indicate that the correlation between these variables was broken, as would be expected by shutting off the cooling water to the towers on one side of the loop. The low predictions for x_6 and x_7 and the high predictions for x_8 and x_9 would indicate that the cooling water was turned off to towers 1 and 2.
FAULT5\FAULT6: There were large negative contributions to the SPE from x_{10} and x_{11}, meaning the model prediction was high. This would be expected from low flow faults.

6.0 CONCLUSIONS AND FUTURE WORK

Based on the above analysis and discussion, the following conclusions can be made:

(1) For the given simple process, the process monitoring scheme based on the multivariate PCA charts had a slight performance advantage over the univariate CUSUM control charts in terms of Type I errors. In terms of Type II errors and fault detection times, the two methods were basically equivalent.

(2) The multivariate PCA monitoring scheme showed a distinct advantage over the univariate scheme with respect to data representation and presentation. This was accomplished through reducing the dimension of the problem by using the PCA model.

(3) The multivariate PCA monitoring scheme also showed an advantage over the univariate scheme in the area of fault diagnosis. The PCA scheme can begin to diagnose the fault by providing information on which variables are contributing to the SPE and shifts in t scores when a fault is detected. This information can be presented to the operator in the form of contribution plots. The contribution plots were able to provide useful information for diagnosing 5 out of the 6 faults tested.

One area of future work for this project would be to streamline the fault diagnosis process for the PCA methodology. Currently, diagnosing the fault using the contribution plots takes time and may prove to be difficult to do in an on-line manner. One approach to overcome this could be to use an expert system to analyse the contribution plots. Also, previous work has shown that radial basis function neural networks can be used for fault diagnosis with the univariate scheme.[10] It may also be possible to develop a radial basis function neural network to be used in conjunction with the PCA monitoring scheme for fault diagnosis.

TABLE 1: STEADY STATE MEANS AND STANDARD DEVIATIONS

Analyser	Steady State Mean	Steady State Std.
T1(x_1)	10.26°C	0.06°C
T2(x_2)	42.92°C	0.21°C
T3(x_3)	50.36°C	0.30°C
T4(x_4)	41.91°C	0.21°C
T5(x_5)	49.02°C	0.26°C
T6(x_6)	26.30°C	0.24°C
T7(x_7)	28.93°C	0.20°C
T8(x_8)	26.30°C	0.19°C
T9(x_9)	26.12°C	0.21°C
F1(x_{10})	7.46V	0.03V
F2(x_{11})	7.14V	0.05V

TABLE 2 : FINAL CUSUM SCHEMES

Analyser	Mean	Δ	Upper Threshold Limit	Lower Threshold Limit	Control Limit (=7.0*σ)
T1(°C)	10.26	3.33σ	10.38	10.18	0.42
T2(°C)	42.92	3.48σ	43.27	42.54	1.47
T3(°C)	50.36	3.57σ	50.85	49.78	2.10
T4(°C)	41.91	3.29σ	42.26	41.57	1.47
T5(°C)	49.02	3.50σ	49.46	48.55	1.82
T6(°C)	26.30	4.16σ	26.80	25.80	1.68
T7(°C)	28.93	4.05σ	29.41	28.60	1.40
T8(°C)	26.30	3.16σ	26.67	26.07	1.33
T9(°C)	26.12	4.00σ	26.55	25.71	1.47
F1(V)	7.46	3.00σ	7.51	7.42	0.21
F2(V)	7.14	3.80σ	7.22	7.03	0.35

TABLE 3: NUMBER OF FALSE ALARMS GENERATED AT STEADY STATE

	Univariate Method	Multivariate Method
# False Alarms @ SS (Training Data, 450 points, 4 days)	14	6
# False Alarms @ SS (Testing Data, 630 points, 4 days)	55	16

TABLE 4: TIMES TO DETECTION OF A FAULT (RESULTS FROM TWO SEPARATE TESTS)

* Spurious alarms before fault

Fault Description	Analysis Method			
	Univariate Method (min:sec)		PCA Method (min:sec)	
	Test 1	Test 2	Test 1	Test 2
10% Power Increase	0:10*	1:20	0:10*	1:20
10% Power Decrease	3:30	2:00	1:10	2:50
Cooling Water Off	1:20	1:10	1:20	0:40
Cooling Water 1/2 Off	1:30	1:30	1:10	0:50
Right Re-Circ Valve Open	0:10	0:10	0:10	0:10
Left Re-Circ Valve Open	0:10	0:10	0:10	0:10

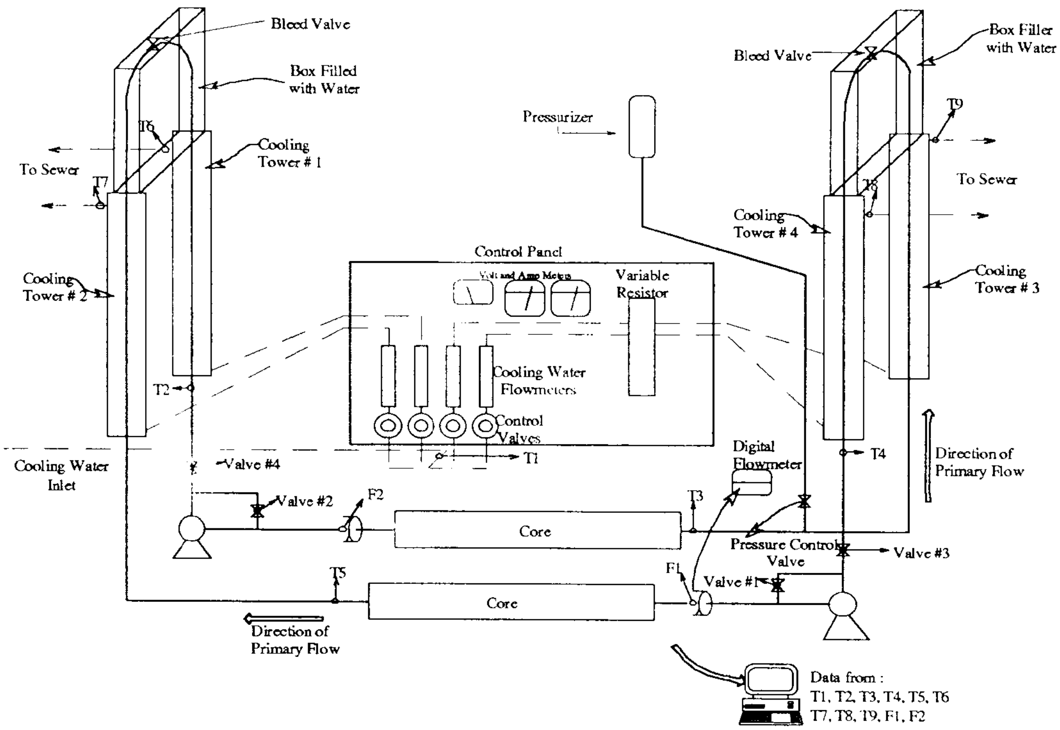
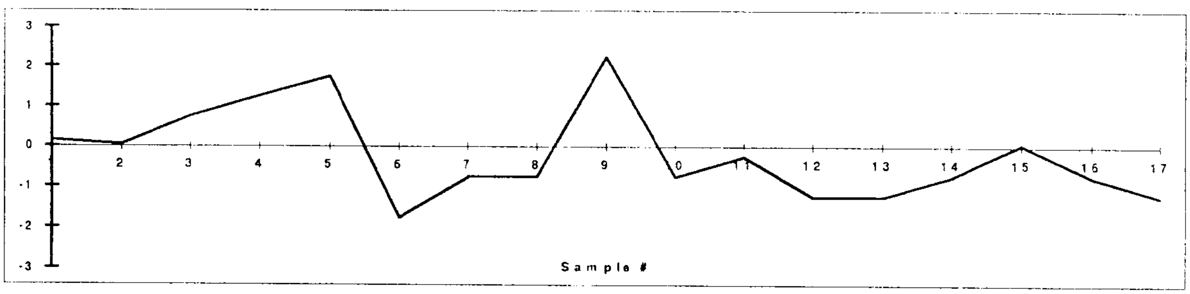


FIGURE 1 : DIAGRAM OF MODEL HEAT TRANSPORT SYSTEM



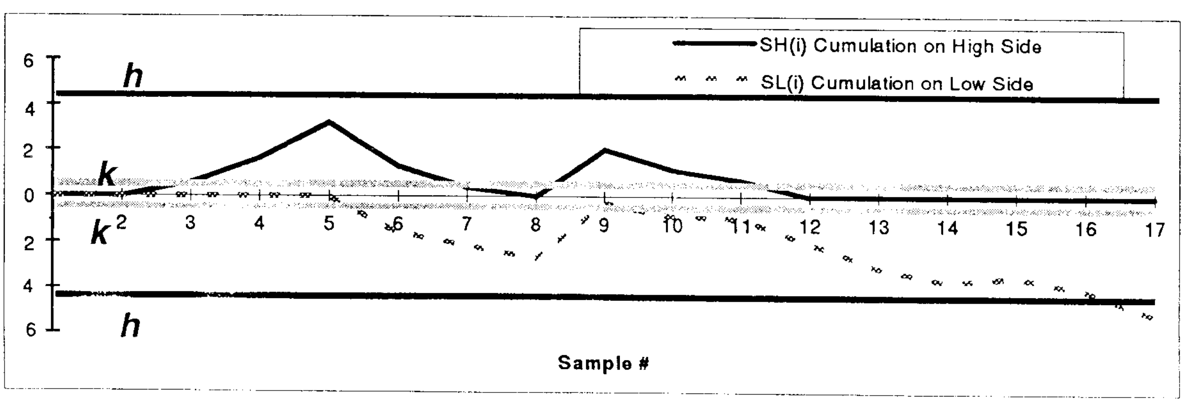


FIGURE 2 : TYPICAL CUSUM SCHEME

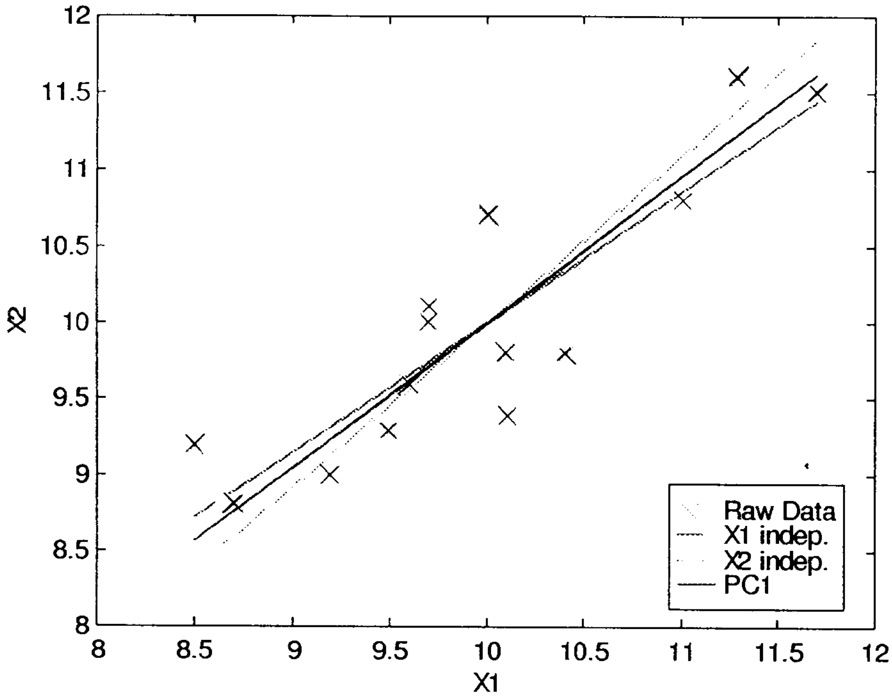


FIGURE 3: SIMPLE PCA EXAMPLE

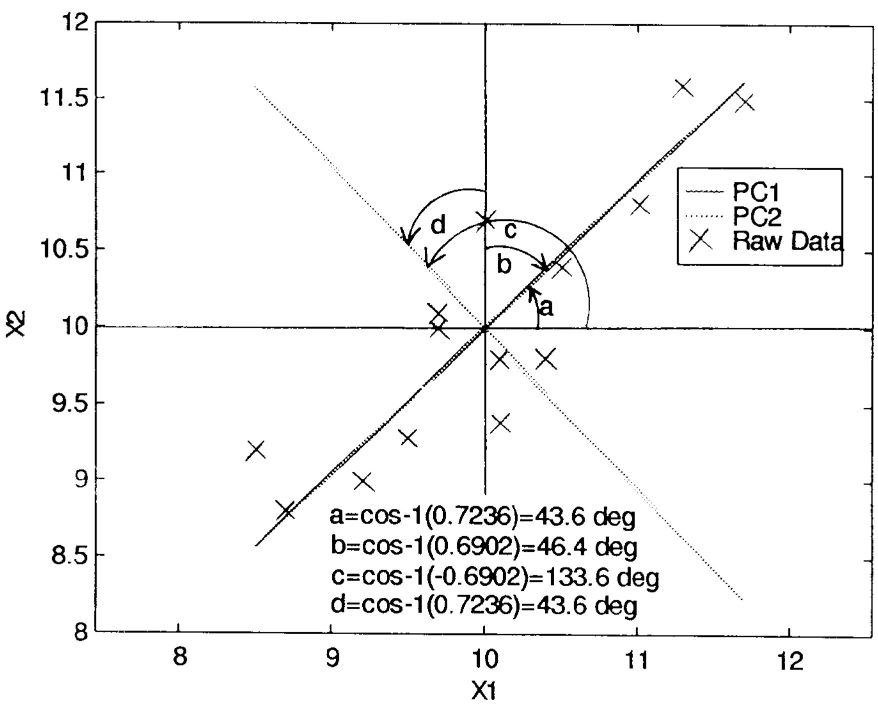
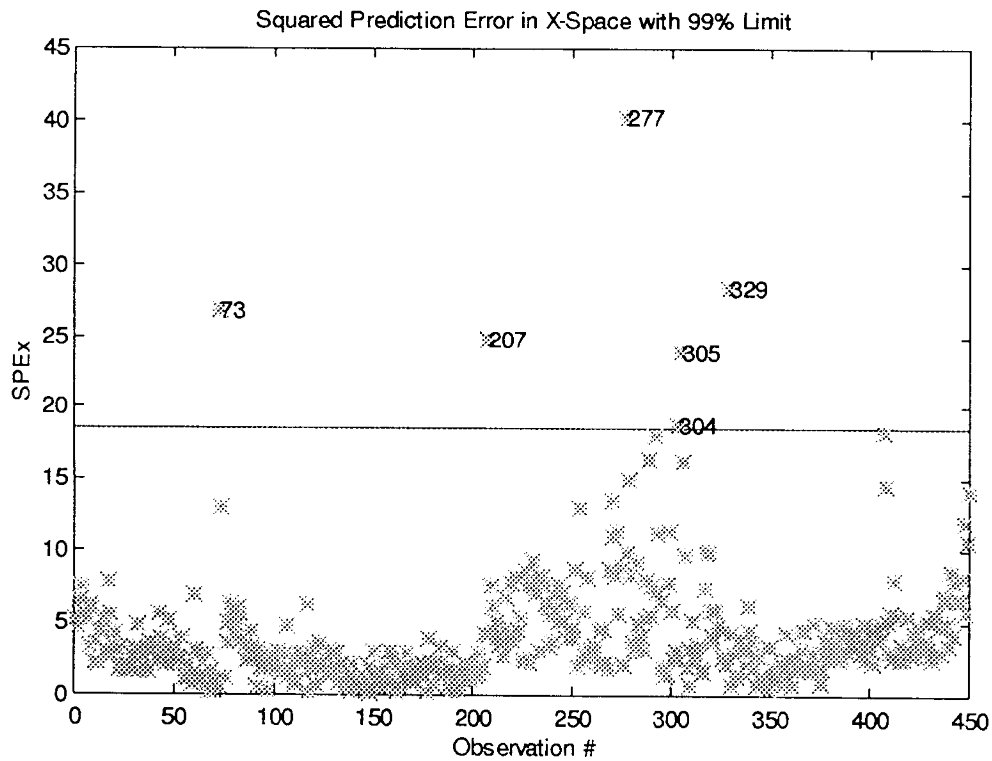


FIGURE 4: PRINCIPAL COMPONENTS FOR SIMPLE PCA EXAMPLE



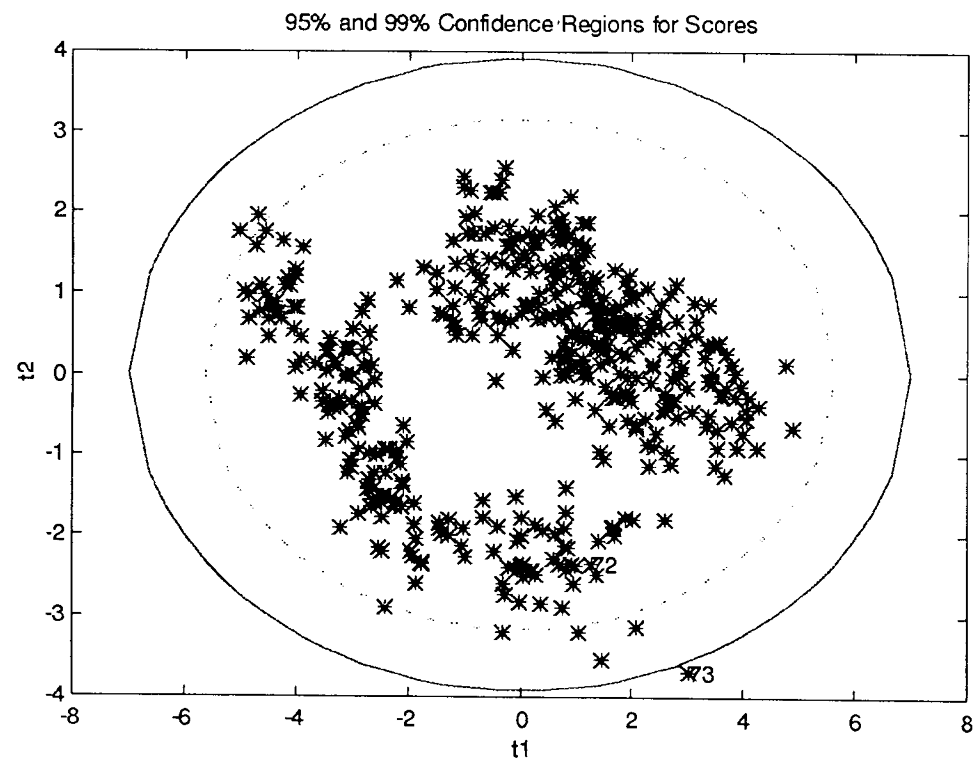


FIGURE 5: SPE AND t PLOTS FOR STEADY STATE DEVELOPMENT DATA

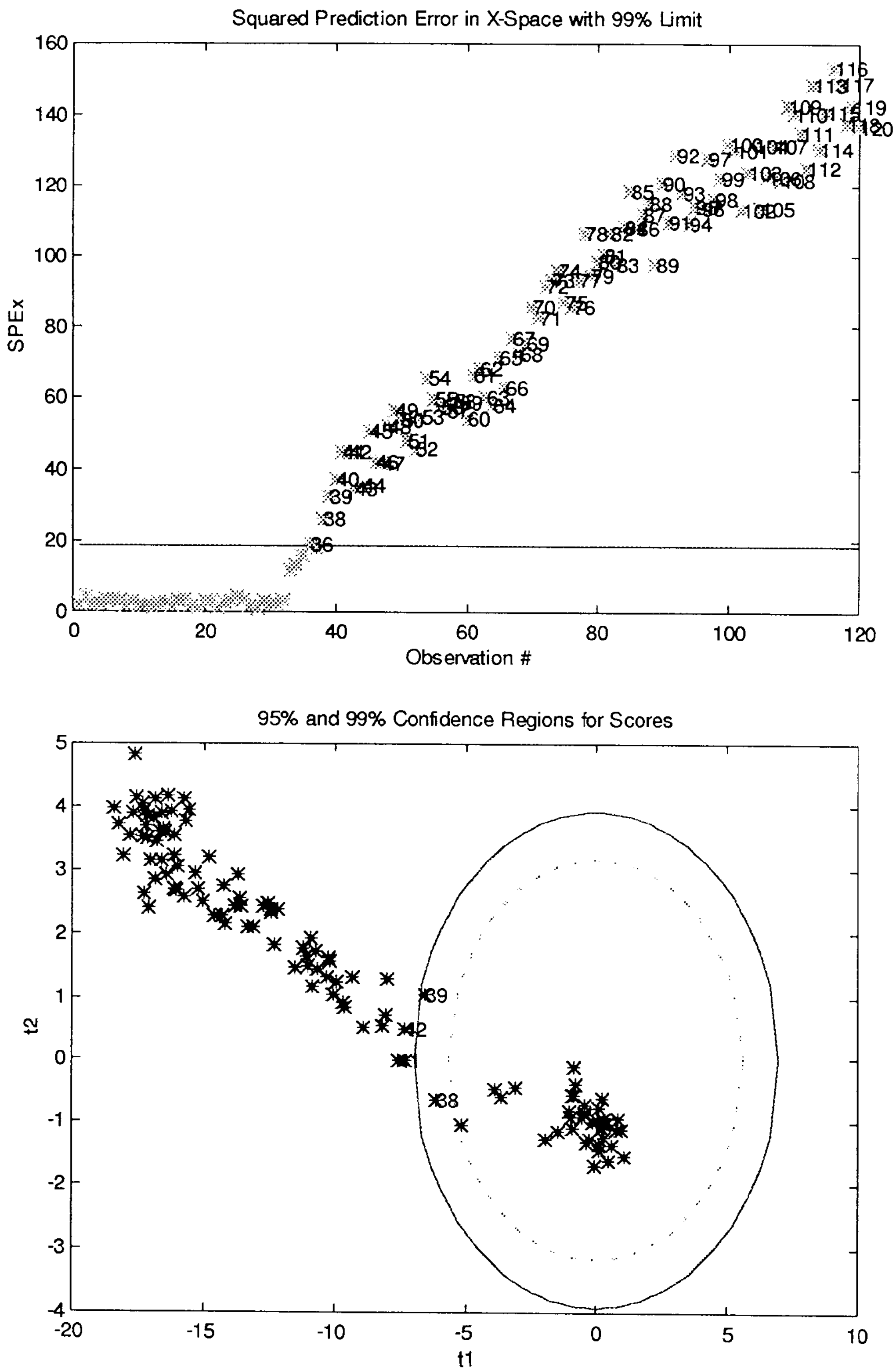


FIGURE 6: SPE AND t PLOT FOR HIGH POWER FAULT, TEST #2

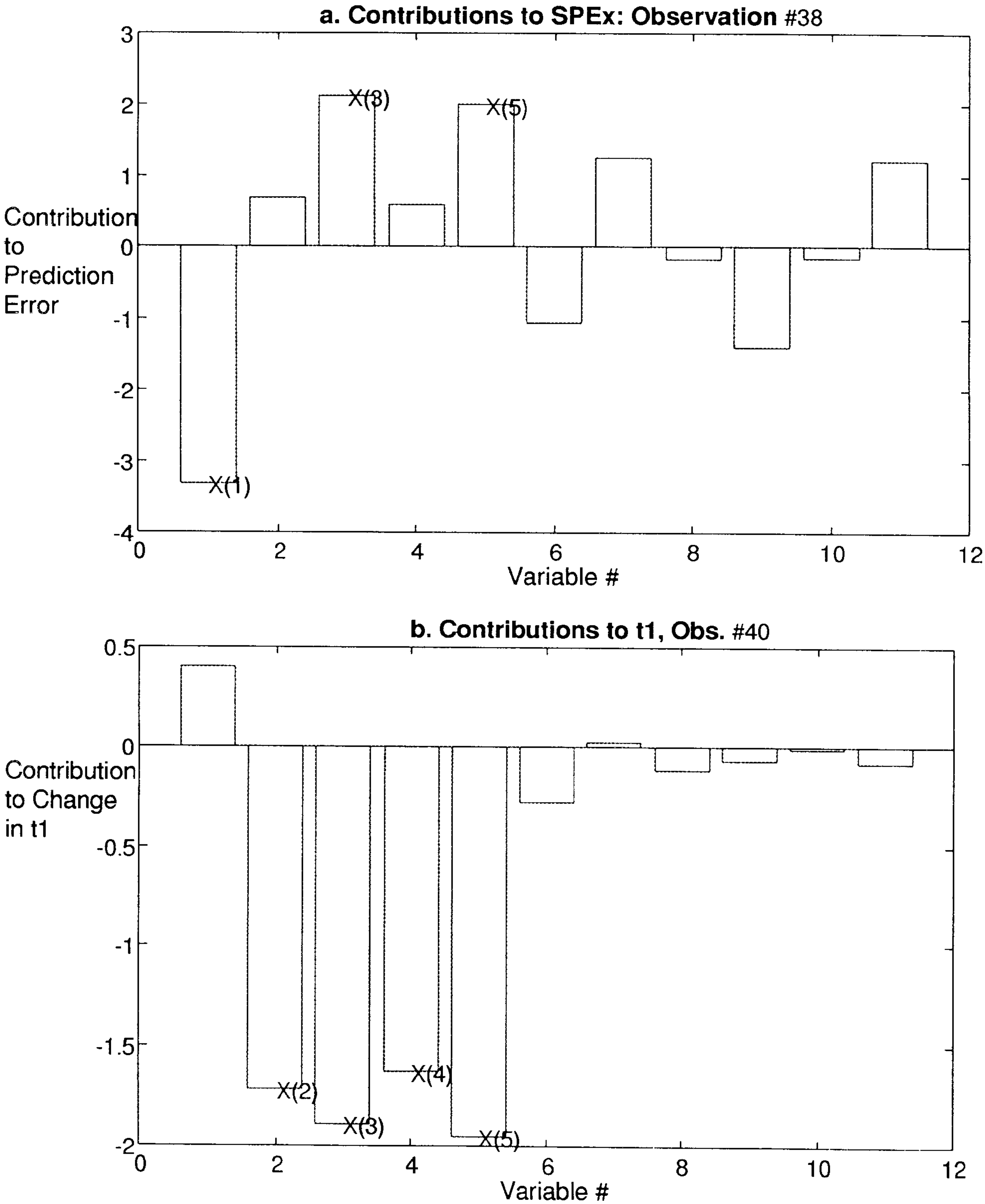


FIGURE 7: CONTRIBUTION PLOTS FOR HIGH POWER FAULT, TEST #2

7.0 REFERENCES

(1) Marquardt D.M., "PQM - Product Quality Management", E.I. du Pont de Nemours & Co. Engineering Department, Applied Statistics Group, Wilmington, Delaware, 19898. 1988 Edition.

(2) Page E.S., "Continuous Inspection Schemes", *Biometrika*, Vol. 41, pp. 100-115, 1954.

(3) Lucus J.M., "The Design and Use of V-Mask Control Schemes", *Journal of Quality Technology*, Vol. 8, No. 1, pp. 1-12, January 1976.

(4) Jackson J.E., "A User's Guide to Principal Components", Wiley-Interscience, John Wiley & Sons, Inc., New York, 1991, Chapters 1-5.

(5) MacGregor J.F., Kourti T., "Statistical Process Control of Multivariate Processes", Control Eng. Practice, Volume 3, No. 3, pp. 403-414, 1995

(6) Hodouin D., MacGregor J.F., Hou M., Franklin M., "Multivariate Statistical Analysis of Mineral Processing Plant Data", CIM Bulletin, Volume 86, No.975, pp. 23-34, November/December 1993

(7) MacGregor J.F., Jaeckle C., Kiparissides C., Koutoudi M., "Process Monitoring and Diagnosis by Multiblock PLS Methods", AIChE Journal, Volume 40, No. 5, pp. 826-838, May 1994

(8) Kresta J.V., MacGregor J.F., Marlin T.E., "Multivariate Statistical Monitoring of Process Operating Performance", The Journal of Chemical Engineering, Volume 69, pp. 35-47, February, 1991

(9) Nomikos P., MacGregor J.F., "Monitoring Batch Processes Using Multiway Principal Component Analysis", AIChE Journal, Volume 40, No. 8, pp. 1361-1375, August 1994

(10) Leger R.P., Garland Wm. J., Poehlman W.F.S., "Fault Detection and Diagnosis Using Statistical Control Charts and Artificial Neural Networks", Proceedings of the 19th CNS Simulation Symposium, Session 7, October 1995.

(11) Wold S., Esbensen K., Geladi P., "Principal Component Analysis", Chemometrics and Intelligent Laboratory Systems, 2, pp. 37-52, 1987

(12) Geladi P., Kowalski B.R., "Partial Least Squares Regression: A Tutorial", Anal. Chim. Acta, Volume 185, pp. 1-17, 1986

(13) Wold S., "Cross-Validatory Estimation of the Number of Components in Factor and Principal Component Models", Technometrics, 20, pp 397-405, 1978.

(14) Ryan T.P., "Statistical Methods for Quality Improvement", John Wiley & Sons Inc., New York, 1989, Chapter 9.

(15) Kourti T., MacGregor J.F., "Multivariate SPC Methods for Monitoring and Diagnosing Process and Product Performance", submitted to Journal of Quality Technology, September, 1995.

(16) Harris T.J., Ross W.H., "Statistical Process Control Procedures for Correlated Observations", The Canadian Journal of Chemical Engineering, Volume 69, pp. 48-57, February 1991.