UPSET AND SENSOR FAILURE DETECTION IN MULTIVARIATE PROCESSES

Barry M. Wise, N. Lawrence Ricker and David J. Veltkamp
Center for Process Analytical Chemistry and
Department of Chemical Engineering
University of Washington
Seattle, WA 98195

ABSTRACT

Many chemical processes have a very large number of measured variables that are recorded frequently. Often, many of these variables are highly correlated and thus provide some redundant information concerning the state of the process and its sensors. When this is the case, multivariate techniques such as Principal Components Analysis and Partial Least Squares calibration can be used in conjunction with Statistical Process Control methods to identify process upsets and sensor failures. We refer to this combination of technologies as Multivariate Statistical Process Control (MSPC). Examples are shown for two types of sensor failure. The first class is where sensors develop a bias. The second class is where sensors become corrupted by noise. We show how confidence limits can be put on PCA residuals for the purpose of detecting failed sensors of both types. This method is compared to a PLS based method where individual variables are calibrated against other system variables and the prediction residual is used in a manner similar to the use of PCA residuals.

1.0 Introduction

Because of recent advances in on-line analytical instrumentation, chemical processes produce much more real-time data than in the past. Most of this data is not used to the fullest advantage, however, because of the lag time between the development of the instrumentation and the data processing algorithms. Data from processes with many sensors often is highly correlated and thus provides some redundant information, although this is not always

apparent from an initial inspection of the data. This paper discusses how two multivariate data analysis techniques, Principal Components Analysis (PCA) and Partial Least Squares (PLS) calibration, can be used in conjunction with techniques from Statistical Process Control (SPC) to detect failed process sensors and process upsets in systems that have many process sensors. We refer to this combination of techniques as Multivariate Statistical Process Control (MSPC). These techniques also have the advantage that they reduce the amount of process information that operators must review.

In this article we briefly review some of the mathematics of the techniques. We then show how these tools can be used to detect "generic" process upsets and sensor failures. We consider the cases of sensors that have developed a bias and sensors that have become corrupted by noise. The effect of serial correlation on the detection problem is also considered.

2.0 Background

This section is provided to give the reader some information concerning the current state of multivariate methods as applied to chemical process data. The ground work for this research comes from many areas, however, and it is beyond the scope of this document to trace them all. Instead, the theory behind PCA and PLS will be reviewed. The emphasis will be on the objectives of the methods rather than the algorithms, which can be found in other references.

2.1 Principal Components Analysis

In PCA an m by n data matrix $\mathbf{X}$ is decomposed into the sum of the product of n pairs of vectors [1, 2]. Each pair consists of a vector in n called the loadings, $\mathbf{p}_i$, and a vector in m referred to as the scores, $\mathbf{t}_i$. Thus $\mathbf{X}$ can be written as

$$\mathbf{X} = \mathbf{t}_1\mathbf{p}_1^T + \mathbf{t}_2\mathbf{p}_2^T + \ldots + \mathbf{t}_n\mathbf{p}_n^T \qquad (1)$$

The matrix of loadings vectors **P** forms a new orthogonal basis for the space spanned by **X** and the individual $\mathbf{p}_i$ are the eigenvectors of the covariance matrix of **X**, defined as:

$$\text{cov}(\mathbf{X}) = (1/(m-1))\, \mathbf{X}^T\mathbf{X} \qquad (2)$$

Thus

$$\text{cov}(\mathbf{X})\mathbf{p}_i = \lambda_i \mathbf{p}_i \qquad (3)$$

where λ_i is the eigenvalue associated with the eigenvector $\mathbf{p}_i$. The loadings vectors are often referred to as principal components, or as "latent variables" (particularly in PLS) because they are linear combinations of the original variables that together explain large fractions of the information in the original matrix. Each of the the $\mathbf{t}_i$ is simply the projection of **X** onto the new basis vector $\mathbf{p}_i$:

$$\mathbf{t}_i = \mathbf{X}\mathbf{p}_i \qquad (4)$$

The value of each λ_i is an indicator of the variance in the data set associated with the direction $\mathbf{p}_i$. In fact

$$\text{fraction variance in direction } \mathbf{p}_i = \lambda_i / \Sigma\lambda_i \qquad (5)$$

In a data set that has been scaled to have variables of zero mean and unit standard deviations

$$\Sigma\lambda_i = n \qquad (6)$$

where n is the number of variables in the data set. In this case, each of the scores vectors $\mathbf{t}_i$ will then have a mean of zero and a standard deviation equal to $(\lambda_i)^{1/2}$. Statistical confidence limits can be placed on the scores using the standard normal deviate and the expected variance as calculated from the eigenvalues. Thus, if the desired confidence limit is 95%, then the value of the standard normal

deviate is 1.96 and the upper and lower control limits for each vector $\mathbf{p}_i$, UCL_i and LCL_i are calculated as:

$$UCL_i = 1.96(\lambda_i)^{1/2} \quad (7)$$

$$LCL_i = 1.96(\lambda_i)^{1/2} \quad (8)$$

The scores can be adjusted to unit variance (which is convenient for other statistical tests, as will be shown) by dividing through by the associated eigenvalues

$$\mathbf{t}_{i,adj} = \mathbf{t}_i / \lambda_i \quad (9)$$

PCA is very closely related to the Singular Value Decomposition (SVD) [3] where a data matrix $\mathbf{X}$ is decomposed as

$$\mathbf{X} = \mathbf{USV}^T \quad (10)$$

where $\mathbf{V}$ contains the eigenvectors ($\mathbf{p}_i$) and $\mathbf{S}$ is a diagonal matrix containing the square roots of the eigenvalues (the singular values) of the covariance matrix of $\mathbf{X}$.

Once the eigenvectors have been determined using PCA or SVD, projections of the data onto the eigenvectors can be made. These projections are commonly referred to as "scores plots" and are often useful for showing the relationships between the data points. Plots can be done as the projections of the samples onto a single eigenvector versus sample number (or time) or onto the plane formed by two eigenvectors. A projection of the samples onto the two eigenvectors associated with the largest eigenvalues depicts the largest amount of information about the relationship between the samples that can be shown in two dimensions. It is for this reason that PCA is often used as a pattern recognition and sample classification technique.

Plots of the coefficients of the eigenvectors themselves, known as "loadings plots", show the relationships between the original

variables in the data set. Correlations between variables show up in the loadings plots.

When PCA is done on a data set it is often found (and it is generally the objective) that only the first few eigenvectors are associated with systematic variation in the data and that the remaining eigenvectors are associated with noise. Noise in this case refers to uncontrolled experimental and instrumental variations arising from random processes. PCA models are formed by retaining only the eigenvectors which are descriptive of systematic variation in the data. Determination of the proper number of eigenvectors can be done by cross-validation or other techniques. Once the PCA model is formed new data can be viewed as projections onto single eigenvectors or the plane formed by pairs of eigenvectors. The "goodness" of fit between new data and the model can be monitored by calculating the data residual. For a reduced order model $\mathbf{P}_k$ (where only the first k of the n total eigenvectors were retained) and a new sample set $\mathbf{X}$ the residual matrix $\mathbf{R}$ is given by

$$\mathbf{R} = (\mathbf{I} - \mathbf{P}_k\mathbf{P}_k^T)\mathbf{X} \qquad (11)$$

The magnitude of the residual for any sample $\mathbf{x}_i$ in $\mathbf{X}$ is

$$Q = \|\mathbf{r}_i\| = \mathbf{r}_i^T\mathbf{r}_i = \mathbf{x}_i^T(\mathbf{I} - \mathbf{P}_k\mathbf{P}_k^T)\mathbf{x}_i \qquad (12)$$

and expresses the "goodness of fit" of the new sample to the model $\mathbf{P}_k$ as a scalar. It can be calculated by taking the sum of squares of the components of $\mathbf{r}_i$. Jackson [4-7] showed that approximate confidence limits can be calculated for the model residual Q provided that all the eigenvalues of the covariance matrix are known, as shown below:

$$Q_\alpha = \frac{\Theta_1}{h_0}\left[\frac{c_\alpha\sqrt{2\Theta_2 h_0^2}}{\Theta_1} + 1 + \frac{\Theta_2 h_0(h_0 - 1)}{\Theta_1^2}\right] \qquad (13)$$

where
$$\Theta_i = \sum_{j=k+1}^{n} \lambda_j^i \text{ for } i = 1,2,3 \qquad (14)$$

and
$$h_0 = 1 - \frac{2\Theta_1\Theta_3}{3\Theta_2^2} \qquad (15)$$

In (13) above c_α is the normal deviate corresponding to the upper (1 - α) percentile. Equation (14) simply states that the Θ_i are equal to the sum of the eigenvalues for the eigenvectors not used in the model taken to the i^{th} power. Note, however, that this result was derived assuming random errors of mean zero etc. It is not clear how autocorrelated data from a real process would affect this result.

The variables responsible for large Q values can often be found through normal statistical process control methods which track single variables. However, there are instances when these methods fail to detect systematic changes in the process or its sensors because the values of the individual variables have not gone "out of bounds" but have instead just become uncorrelated (or changed their correlation) with the remaining variables.

There are several methods for determining the source of the large Q values in this case. The simplest method is to calculate the column norm (the sum of squares over the variables, instead of over the samples, as is done to calculate Q values) for the residuals matrix for the samples with large Q values. Generally, the perturbed variables will show up as having the largest residuals. This method can fail, however, because it does not account for the fact that the average size of the residual is different for different variables in a data set. What is needed is confidence limits on the residuals so that it is possible to determine if the residuals are abnormally large or noisy.

In other cases the factors responsible for large values of Q can be found by subjecting the matrix of $\mathbf{r}_i$ vectors to PCA. This determines the major source of variation in the data not accounted for by the original PCA model. Typically the variable with the largest (absolute

value) coefficient in the first eigenvector from the residuals matrix will be the variable responsible for the deviation of the PCA model.

While the Q statistics offer a way to test if the process data has shifted outside the normal operating space, there is a need for a statistic that provides an indication of unusual variability within the normal subspace. This is provided by Hotellings T^2 statistic [8]. The value of T^2 for a sample is equal to the sum of squares of the adjusted (unit variance) scores on each of the PCs in the model. The statistical confidence limits for the values of T^2 can be calculated by taking advantage of the statistical F-distribution as follows

$$T^2_{p,m,\alpha} = \frac{p(m-1)}{m-p} F_{p,m-p,\alpha} \qquad (16)$$

Here m is the number of samples in the data set used in the calculation of the PCA model, p is the number of principal component vectors retained and α corresponds to the standard normal deviate.

The residual variance for each variable can also be calculated for the PCA models. It can be shown that for a given data set $\mathbf{X}$ with a full set of PCA vectors $\mathbf{P}$, of which k are retained, and eigenvalues λ, then the variance of the residual for the j^{th} variable is

$$s_j^2 = \sum_{i=k}^{n} p_{ij}^2 \lambda_i \qquad (17)$$

where p_{ij} is the loading of the j^{th} variable in the i^{th} PC. For the data matrix from which the model was obtained this relationship will be exact. If however it is assumed that the eigenvalues of all the PCs not retained in the model are equal (which is generally the assumption when they are not used) then the variance in the residual of the j^{th} variable can be estimated using only the PCs and eigenvalues retained in the model from

$$\widehat{s_j^2} = \left(\sum_{i=1}^{n} \lambda_i - \sum_{i=1}^{k} \lambda_i \right) \left(1 - \sum_{i=1}^{k} p_{ij}^2 \right) \qquad (18)$$

The first term on the right hand side of equation (18) can be replaced with the total sum of squares which is equal to the sum over all of the eigenvalues.

Now that we have an expression for variance of the residual for each variable we can test to see if the observed variance in a residual is equal to the expected variance. For this we can use the standard F-test with the appropriate degrees of freedom. We will test to see if

$$s_j{}^2_{new} / s_j{}^2_{old} > F_{v\text{-}new,v\text{-}old,\alpha} \qquad (19)$$

where v-new is equal to the number of new samples minus the number of PCs used in the model minus one and v-old is the number of samples used to obtain the original model minus the number of PCs minus one. When the inequality in equation (19) holds then a change in the variance of the residual has occured to a confidence level of 1 - α. The F-test parameters can now be used to set upper and lower limits on the variance of the residuals.

Because of the scaling we have chosen the mean residual should be zero for all the variables. In order to detect a shift in the mean away from zero, however, we can use the t-test. In our case the hypothesis we want to test is that the means are equal. Thus the t-test reduces to

$$t_{v\,tot} = \frac{\left(\overline{x_{old}} - \overline{x_{new}}\right)\left(v_{old} + v_{new}\right)^{0.5}}{\left(1/v_{old} + 1/v_{new}\right)^{0.5}\left(v_{old}s^2_{old} + v_{new}s^2_{new}\right)^{0.5}} \qquad (20)$$

where the degrees of freedom are both one greater than for the case given above. For the purpose of setting limits, the variances can be assumed to be equal to the variance of the residuals of the calibration set. Furthermore, once the desired confidence level is chosen, it is possible to solve for the difference between the old and new means that is just significant.

2.2 Partial Least Squares

PLS regression methods are well described by Hoskuldsson [9] and the history of PLS is covered quite well by Geladi [10]. A theoretical foundation for PLS is provided in the reference by Lorber et. al. [11]. Here we present a brief description of the method and outline the computational steps.

PLS is a multivariate calibration technique where a data matrix of inputs, known as the **X** or independent block, can be calibrated to a matrix of outputs, the **Y** or dependent block [12, 13]. PLS can be thought of as a simultaneous decomposition of the **X** and **Y** blocks using PCA. In PLS, however, the eigenvectors are rotated in each of the blocks so that the samples have the same "scores". To put it another way, the projections of the independent variables onto the first "rotated eigenvector" of the **X** block will be highly correlated to the projections of the dependent variables onto the first "rotated eigenvector" of the **Y** block and so on.

Mathematically, the PLS algorithm exchanges the scores between the **X** and **Y** blocks as the matrix decomposition proceeds, resulting in highly correlated "eigenvectors" (latent variables). The PLS algorithm can be found in several of the references.

PLS can be contrasted with Multiple Linear Regression (MLR) by noting that MLR is a special case of PLS, i.e., MLR is equivalent to using all the latent variables in PLS. In MLR the vector of coefficients $\mathbf{b}_{i,mlr}$ is estimated for each of the $\mathbf{y}_i$ in **Y** as

$$\mathbf{b}_{i,mlr} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}_i \qquad (21)$$

thus the estimated value of **Y** is

$$\mathbf{Y}_{est} = \mathbf{XB} \qquad (22)$$

where **B** is composed of the column vectors calculated in (21). While MLR generally gives a better fit to the calibration data because it uses all the variation in the **X** block, PLS often gives better prediction because it uses only the predictive information.

The parameters used in PLS prediction can also be reduced to a single linear equation, similar to that of (22):

$$\mathbf{Y}_{est} = \mathbf{XC} \qquad (23)$$

where **C** is a matrix in the general case and a vector in the case of only one variable in the **Y** block. The result is equation (24) where lv is again the number of latent variables to be used in the prediction (as above) and it is assumed that the value of the term in brackets is equal to **I** for the case of i = 1.

$$\mathbf{Y} = \mathbf{X} \sum_{i=1}^{lv} b_i \left[\prod_{j=1}^{i-1} \left(\mathbf{I} - \mathbf{w}_j \mathbf{p}_j' \right) \right] \mathbf{w}_i \mathbf{q}_i' \qquad (24)$$

In (24) the b_i are the inner relation coefficients, the $\mathbf{w}_j$ and the $\mathbf{p}_j$ are the **X** block weights and loadings and the $\mathbf{q}_i$ are the **Y** block loadings.

It is proposed that PLS can also be used to determine the general state of the process in a fashion similar to the use of the Q statistics associated with PCA models. This requires that PLS models be obtained that relate each variable to the remaining variables in the system. Thus for a system with n variables, n PLS models would be required. Fortunately, using the relationship given in equation (24) the n PLS models can be formed into a single matrix, with each model being a column vector. Because each of the variables does not contribute to its own prediction, the resulting prediction matrix, $\mathbf{M}_p$, has zeros on the diagonal. Thus the PLS prediction $\mathbf{X}_{est}$, of a data matrix **X** can be obtained by simple matrix multiplication

$$\mathbf{X}_{est} = \mathbf{X}\ \mathbf{M}_p \qquad (25)$$

A residuals matrix, $\mathbf{R}_{pls}$, can be calculated from

$$\mathbf{R}_{pls} = \mathbf{X} - \mathbf{X}_{est} = \mathbf{X} - \mathbf{X}\ \mathbf{M}_p = \mathbf{X}(\mathbf{I} - \mathbf{M}_p) \qquad (26)$$

The similarity between equation (26) and the calculation of the PCA residuals in equation (11) should be readily apparent.

The residuals matrix $\mathbf{R}_{pls}$ can be used in much the same manner as the PCA residuals matrix **R** for determination of the overall state of the process (as in calculation of Q) or for determining the failure of specific sensors. In the latter case PLS predictions for each variable can then be compared with the actual values, and the off-normal variables can be identified [14]. This approach is the generally accepted method for process fault detection. A model is produced which predicts the value of a process variable from other process variables and the difference is monitored. The idea of using PLS in this fashion is related to the general idea of prediction of process output using secondary measurements, such as the example provided by Mejdell and Skogestad in [14].

2.3 Data Pretreatment

Before completely leaving the topics of PCA and PLS a word about preprocessing of data is in order. In particular, scaling of variables is very important to PCA, PLS and other eigenvalue analysis type methods. This is because eigenvectors will tend to be biased towards variables with larger numerical values since they appear to be associated with greater amounts of variance. For this reason we usually use a procedure know as "autoscaling" is with each data set prior to analysis with PCA or PLS. The mean is subtracted from each variable (mean centering) and the result is divided by the standard deviation (variance scaling) to produce a matrix with variables of zero mean and unit variance. The use of this technique implies that the variables are of equal importance which is a good starting point given no other information.

Elimination of outliers is also a good idea, since these data points have a great deal of leverage on the data models and can change them significantly. PCA can be used for data pretreatment to detect outliers.

Finally, it is important to note that the PLS and PCA techniques are linear, and as such, will not provide good models of highly non-

linear data. If the functionality of the non-linearities is known, it is best to pretreat these variables so that the resulting variables are linear.

3.0 Using MSPC to detect "Generic" Upsets and Sensor Failures

In this section we will consider the general problem of identifying failed sensors and process upsets through application of the multivariate tools and SPC techniques described previously. In this section we will assume that we have no specific upsets that we are looking for, but instead, are trying to determine the general health of the process, including its sensors. For this reason, the upset detection schemes we will apply here will all be based on process residuals, the difference between the behavior of the actual process and a model. We will consider PCA and Varimax models in the first section and PLS based models in the second section, and demonstrate them with some representative examples.

3.1 Using PCA with Non-Serially Correlated Data

In this example we will consider a synthetic process with 10 variables that has a true rank of 5. Variables will have a noise component that is equal to .5 times the standard deviation of the deterministic variation. The process for this example was generated at random, though some care was taken to assure that the process generated was not "close" to being of rank less than 5. The process is assumed to be excited by random noise. A data matrix of 1000 samples was generated for the process and a Principal Components Analysis performed. The results are shown in Table 1.

Here we will choose a 5 PC model of the process data, because we know the process that we generated is intrinsically 5 dimensional. In actual practice the order of the model may have to be determined either from cross validation of the PCA models or by comparison of the size distribution of the smaller eigenvalues to the expected distribution of "error eigenvalues". The control chart for

the first PC with 95% confidence limits is shown in Figure 1. The control charts for the other PCs are very similar so are not shown. The Q chart is shown in Figure 2.

Table 1. Eigenvalues and Variance Captured by PCA Model of Example Process Data

PC #	Eigenvalue	% Variance	Total
1.0000	4.7102	47.1020	47.1020
2.0000	1.5165	15.1646	62.2666
3.0000	1.4027	14.0273	76.2939
4.0000	0.9547	9.5475	85.8414
5.0000	0.4386	4.3858	90.2272
6.0000	0.2063	2.0635	92.2907
7.0000	0.2060	2.0597	94.3503
8.0000	0.2004	2.0038	96.3541
9.0000	0.1867	1.8672	98.2213
10.0000	0.1779	1.7787	100.0000

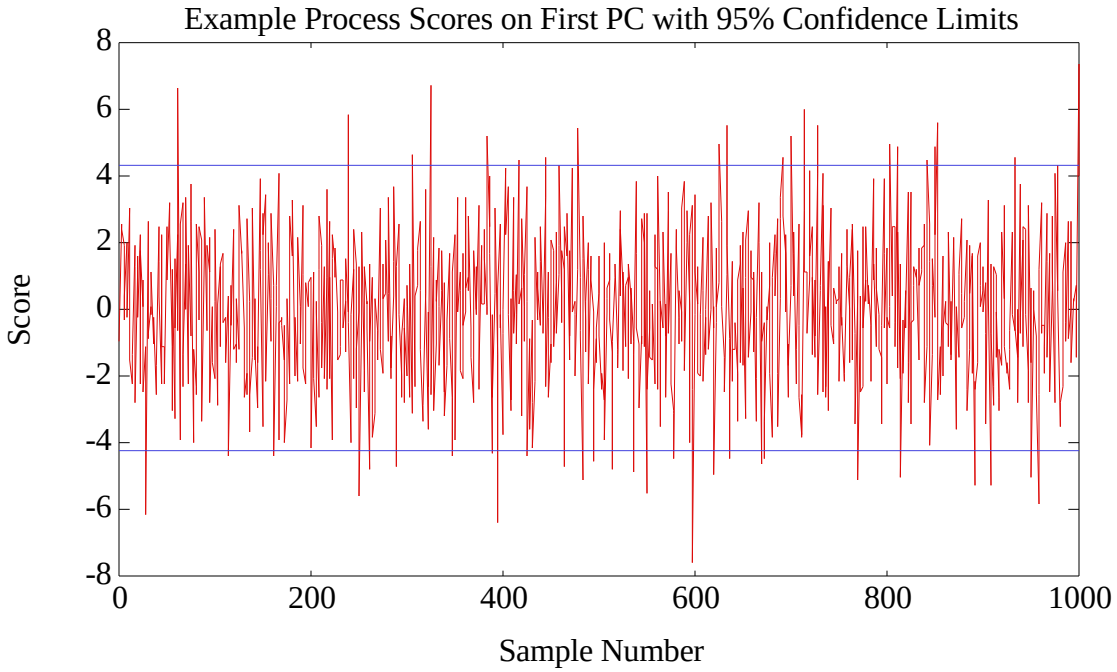

Figure 1. Example Process Scores on First Principal Component with 95% Confidence Limits

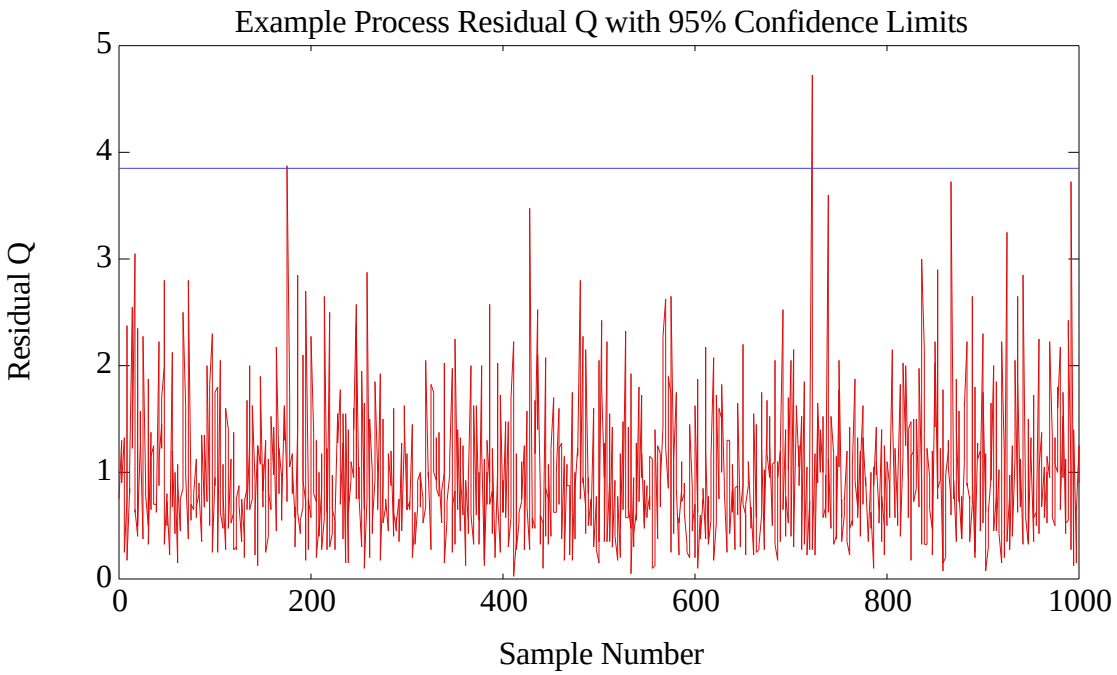

Figure 2. Data Residual Q for Example Process with 95% Confidence Limits

The variance of the residual for each variable was calculated using equation (17) which weights the residuals by the eigenvalues of the PCs not kept in the model and with equation (18) which assumes that all eigenvalues of the unused PCs are equal. The results of these calculations are compared to the actual variance of the residuals in Figure 3. The actual variance is shown as the solid line and coincides exactly with the variance calculated from equation (17) which is shown as the circles. The variance calculated from equation (18) is shown as the dashed line.

In the calculations that follow we will use the variance estimate that assumes equality of the eigenvalues since they should in principle be equal because the process is intrinsically 5 dimensional. If some small PCs that were thought to contain some intrinsic variation were not retained in the model then it may be more appropriate to use equation (17).

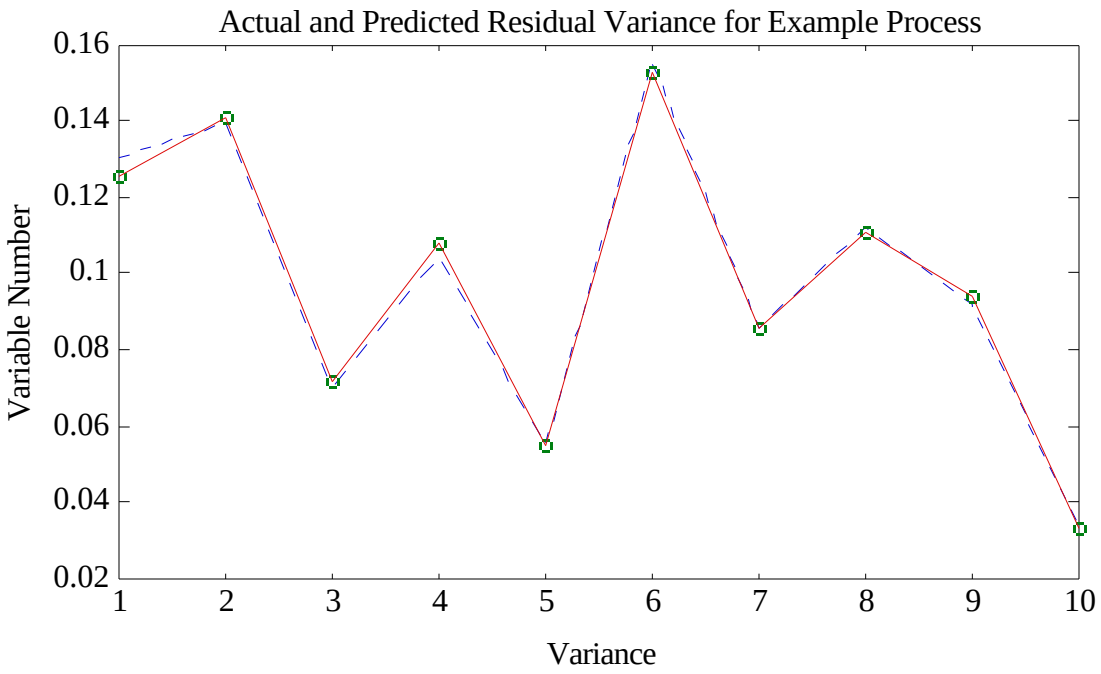

Figure 3. Actual and Predicted Variance for Example Process Using Equations (17) and (18).

Control limits on the mean and variance of the residuals can be calculated using the F and t statistics once the desired confidence level and the size of the sampling window for incoming data is known. In this example we choose a confidence of 99% and a window of 50 samples. The confidence limits calculated by this procedure are shown in Figures 4 and 5. A cross validation procedure was also performed to determine if the control limits established from theory were correct. The limits were found to be nearly identical.

At this point the reader may wonder why 50 points was chosen for the test sets. For this synthetic example the number is somewhat arbitrary. In actual practice, one would want to chose a number of samples that was relevant to the process sample time and time constants. While a large number of samples would allow us to detect smaller perturbations in the process, it will also increase the time required to detect a shift. The number of samples should therefore be the largest number possible that is within the time span of interest.

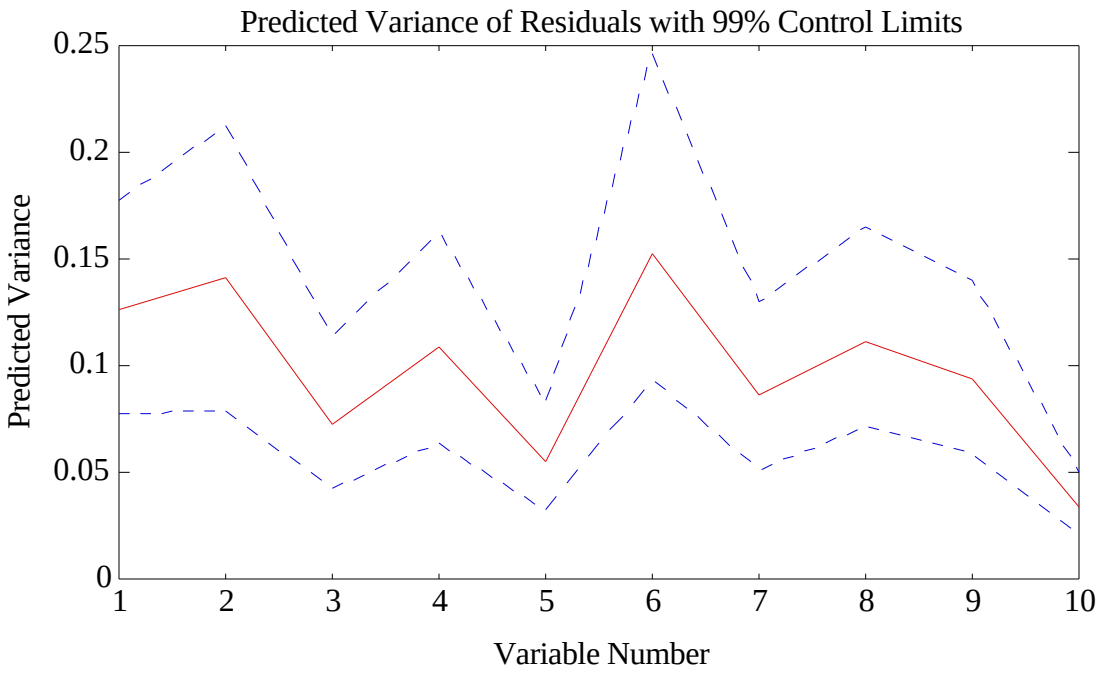

Figure 4. Predicted Variance of Residuals with 99% Control Limits

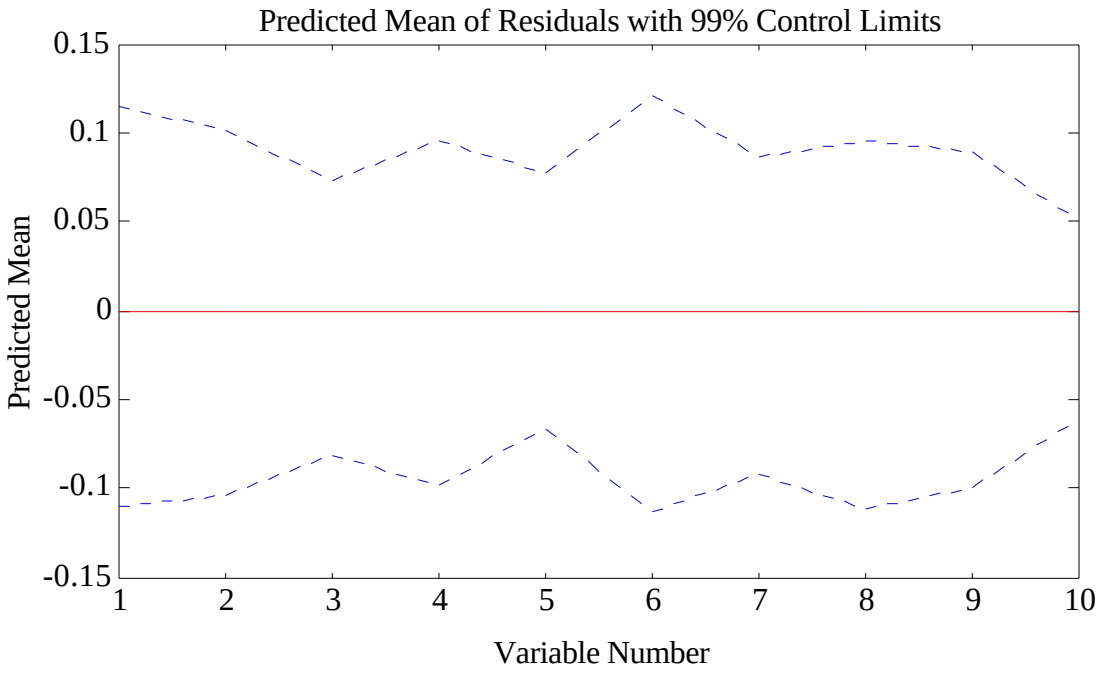

Figure 5. Predicted Mean of Residuals with 99% Control Limits.

Two types of sensor faults will be generated and used to test the method. First we will consider a bias type error, where a constant

offset is added to a variable. Figures 6 and 7 show a typical result for a bias that is equal to 0.2, 0.4, and 0.6 standard deviations of the original variable number 5. (Remember that in this example the noise component is equal to 0.5 times the intrinsic variability.) In Figure 6 we see that the mean of the residuals calculated is clearly outside the confidence limits for the high bias case. In Figure 7, however, all cases lie within the confidence limits for the variance, as would be expected. If the test was performed as an actual moving window, however, we would expect to see an increase in the amount of variance in the biased variable as the window passes over the point where the bias was initiated.

In order to make it easier to distinguish the biased variable it is convenient to ratio the mean residuals to the detection limits. When this is done all "normal" variables should have values less than one. The more biased a variable is the larger its ratio will be. Figure 9 shows that variable 5 is the most biased quite clearly. Note that other variables have also gone over the limit. This is typical for highly correlated variables.

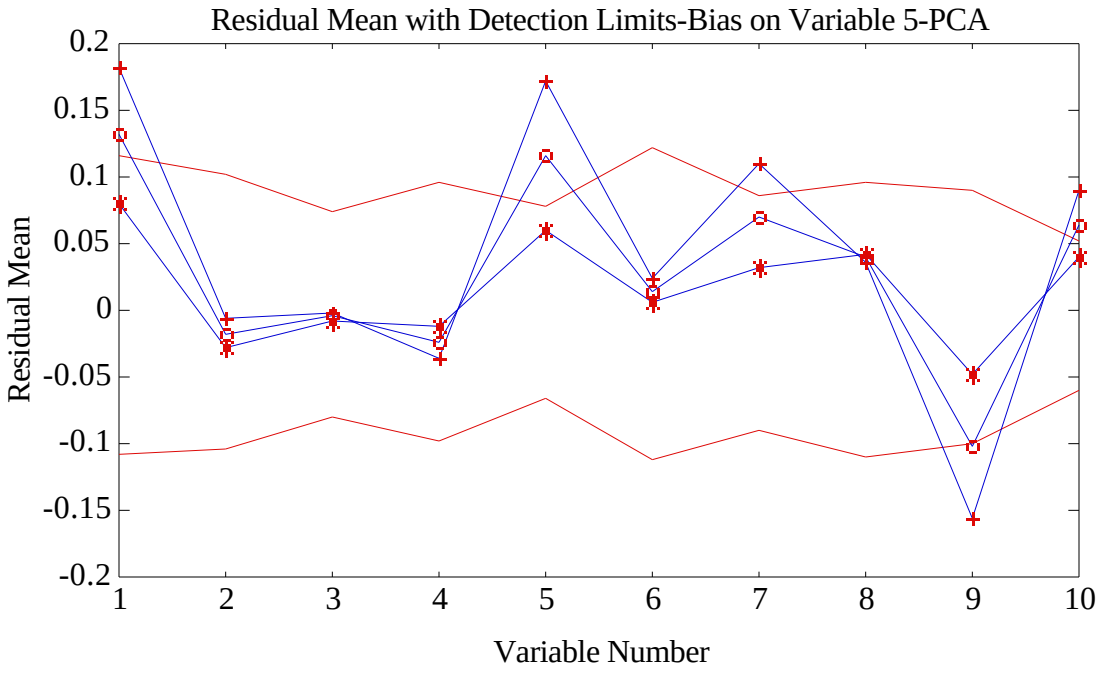

Figure 6. Residual Mean with Detection Limits for Bias of 0.2, 0.4 and 0.6 Standard Deviation on Variable 5-PCA Model

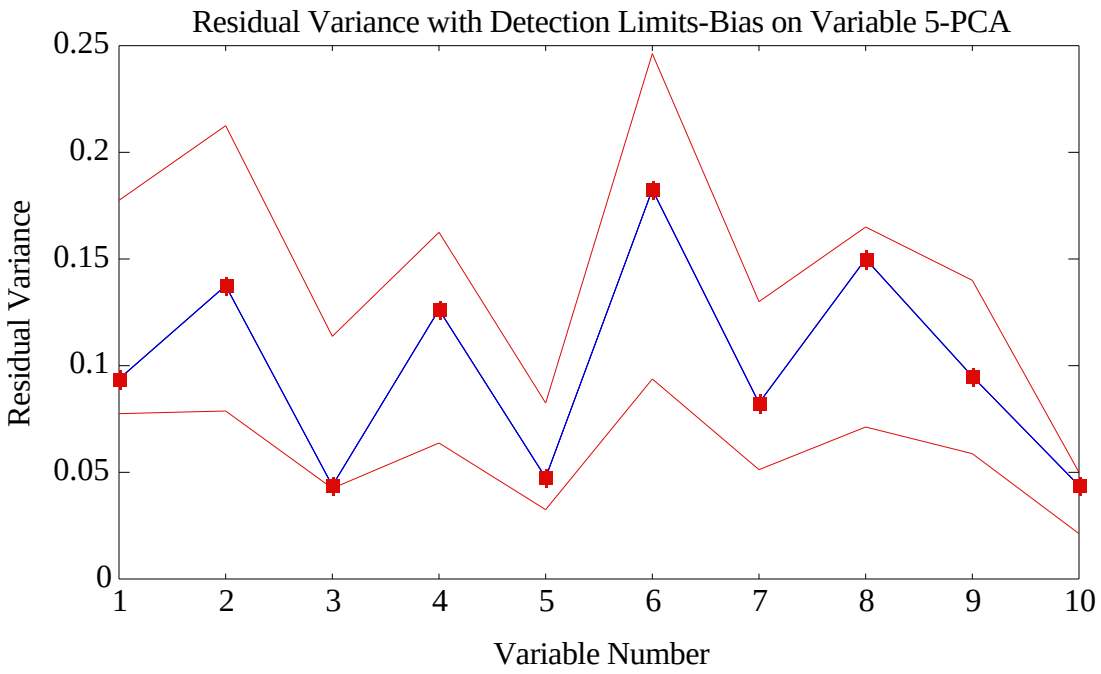

Figure 7. Residual Variance with Detection Limits for Bias of 0.2, 0.4 and 0.6 Standard Deviation on Variable 5-PCA Model

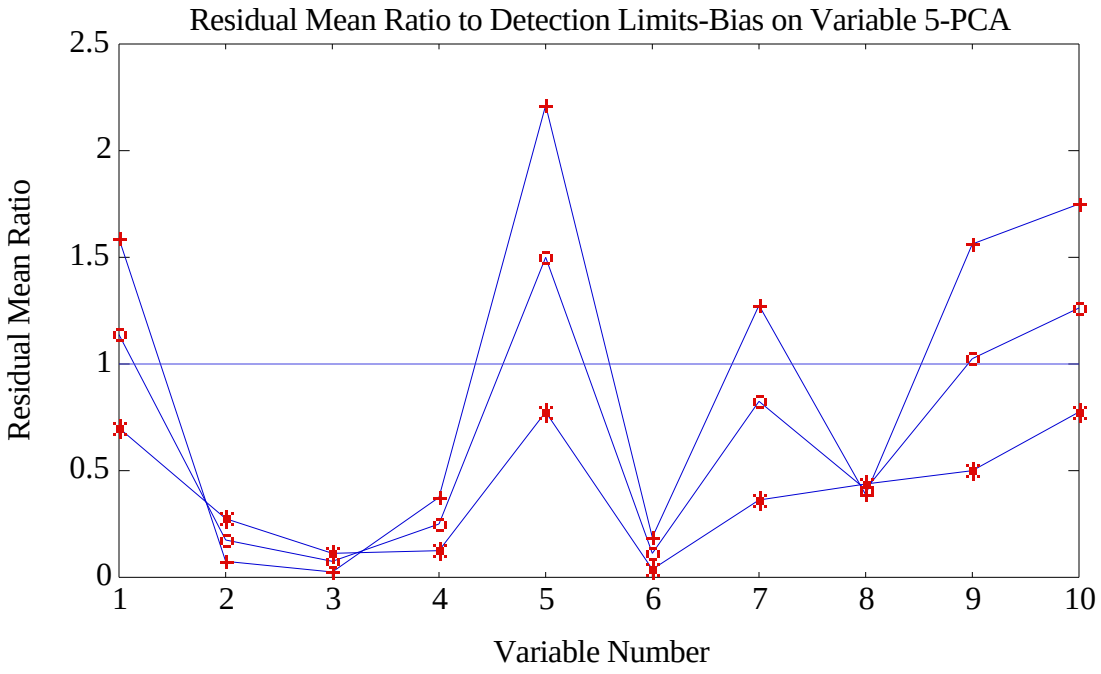

Figure 8. Mean Residual Ratio to Detection Limits for Bias of 0.2, 0.4 and 0.6 Standard Deviation on Variable 5-PCA Model

In the second test additional noise was added to selected variables. The amount of noise was equal to 0.4, 0.6 and 0.8 standard deviations of the original variables. An example of the effect of the added noise on the residuals is shown in Figures 9 and 10. In this example noise was added to variable number 6. Figure 9 shows that the mean of the residual for the 50 samples falls within the 99% confidence limits, as would be expected. There is some change in the calculated mean residual, however, due to the fact that the noise added does not have a mean of zero over the small number of samples In Figure 11 the variance of the residual for variable 7 is clearly outside the limits for the 0.6 and 0.8 cases.

The value of the confidence limits can be seen in Figures 12 and 13. For this example noise was added to variable 3. Without the limits it would not be evident that variable 3 was any different from the remaining variables, even in the high noise case. In fact, in absolute terms, the variance of variable 6 is much larger. The addition of the confidence limits, however, shows that the variance of variable 3 is outside the limits, while for variable 6 it is not.

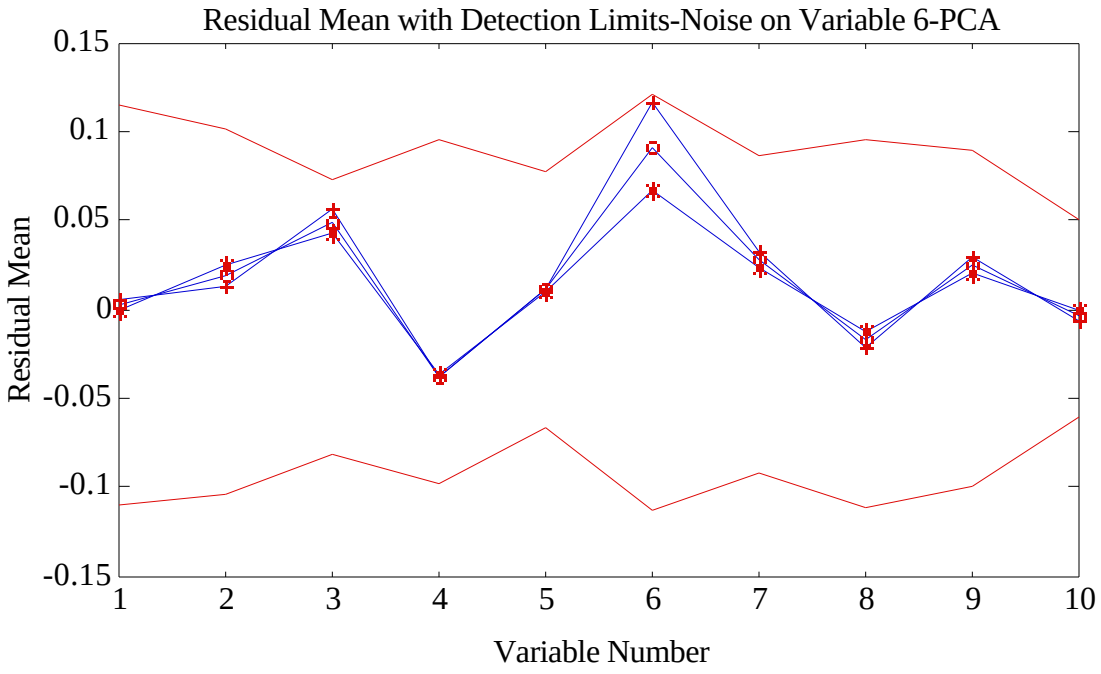

Figure 9. Residual Mean with Detection Limits for Noise Increase of 0.4, 0.6 and 0.8 Standard Deviations on Variable 6-PCA Model.

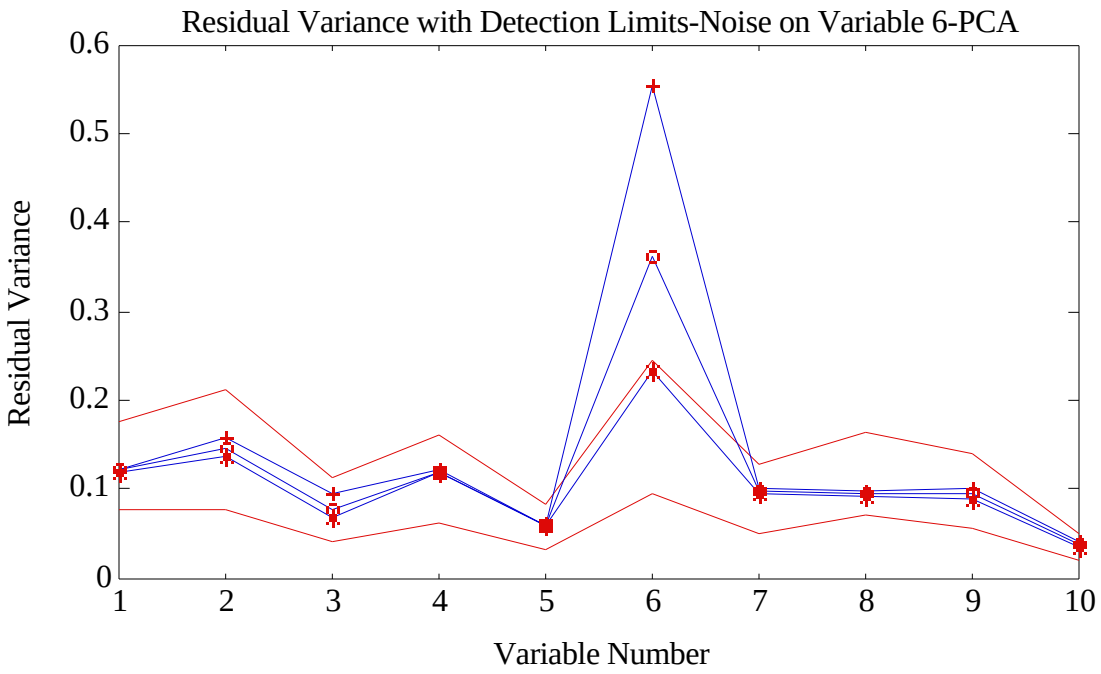

Figure 10. Residual Variance with Detection Limits for Noise Increase of 0.4, 0.6 and 0.8 Standard Deviations on Variable 6-PCA Model.

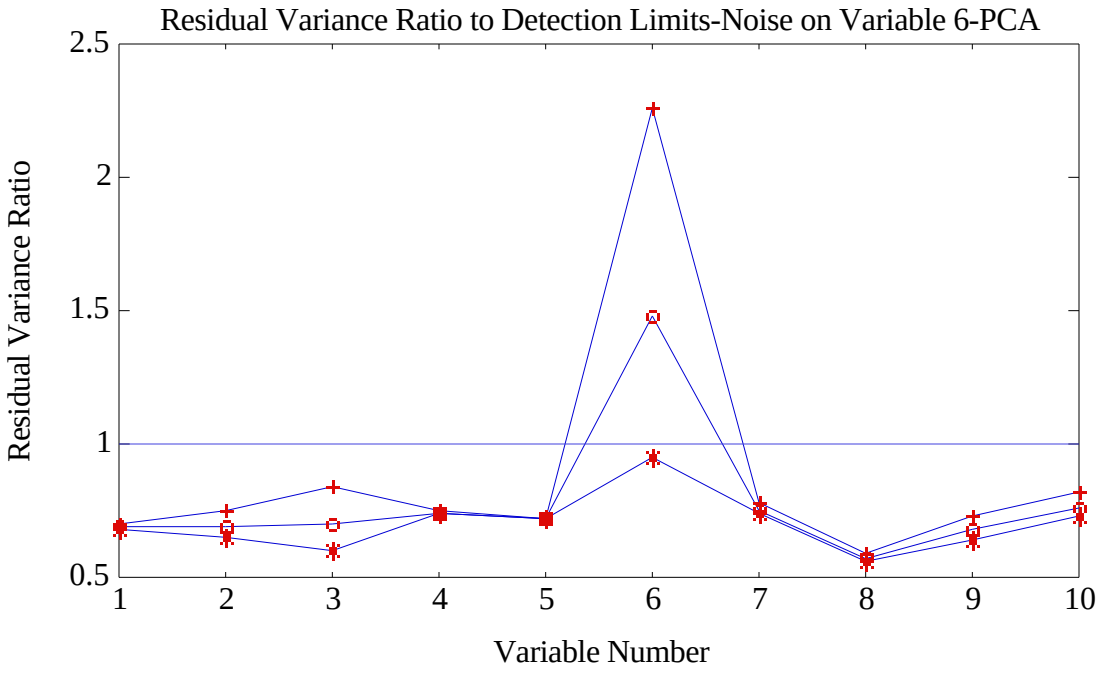

Figure 11. Residual Variance Ratio to Detection Limits for Noise Increase of 0.4, 0.6 and 0.8 Standard Deviations on Variable 6-PCA Model.

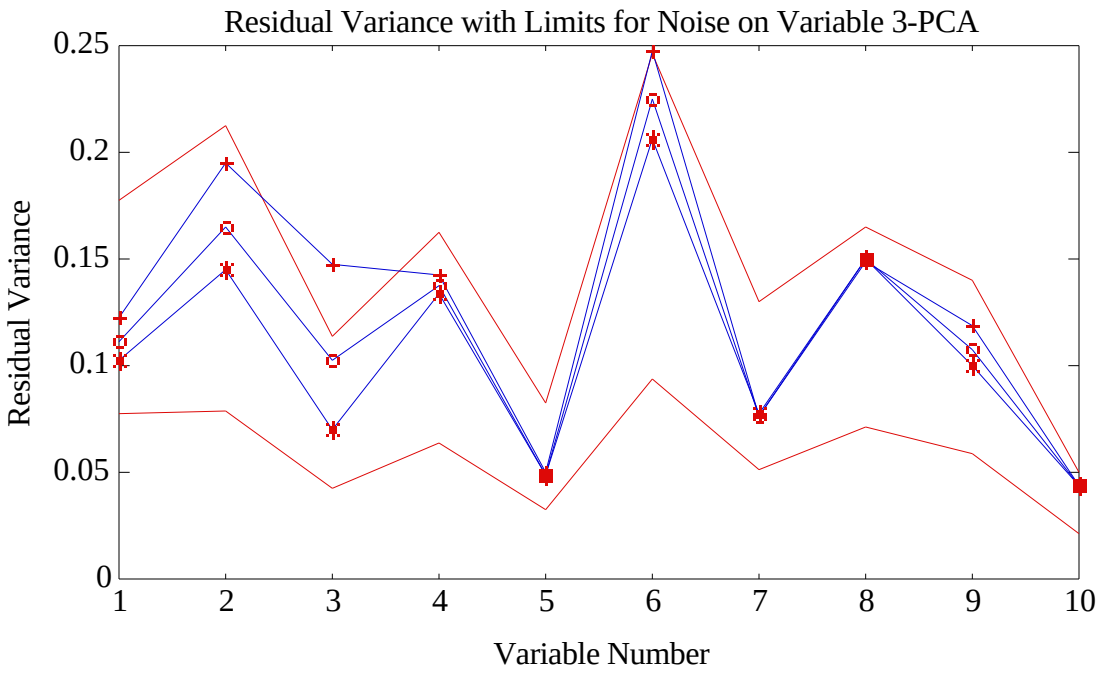

Figure 12. Residual Variance with Detection Limits for Noise Increase of 0.4, 0.6 and 0.8 Standard Deviations on Variable 3-PCA Model.

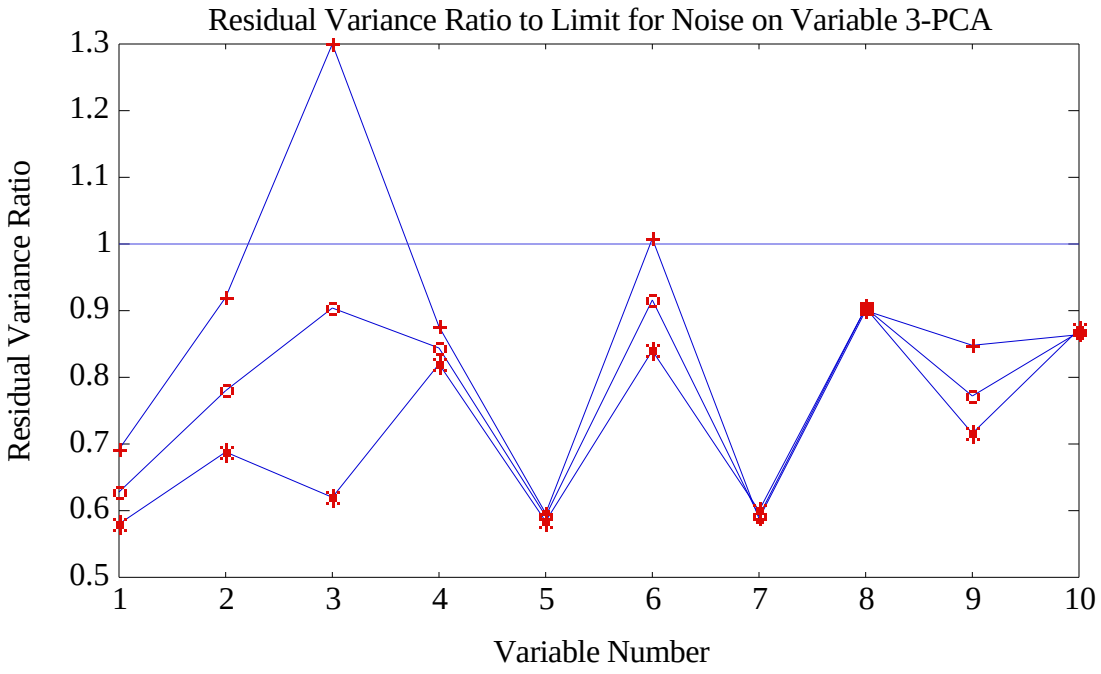

Figure 13. Residual Variance Ratio to Detection Limits for Noise Increase of 0.4, 0.6 and 0.8 Standard Deviations on Variable 6-PCA Model.

The question may arise concerning why variables other than the failed variables go outside their limits at all. This is because in the projection process any errors tend to get spread out among highly correlated variables. In fact, if one looked back at the original covariance matrix for this data set we would find that the variables that are most highly correlated with variable 7 are variables 6 and 4.

The effect of these "sensor failures" on the Q values was also investigated for these cases. It was found that the Q values did not show a statistically significant difference between "normal" data and the data from perturbed sensors in any of the cases tested. Thus it is clear that Q statistics are not very sensitive to relatively small changes in single sensors. This is not surprising since Q is an overall measure and would not be expected to have as much sensitivity for individual variables.

3.2 Using PLS with Non-Serially Correlated Data

As a comparison to the PCA based methods above we will now consider PLS based models. It was shown in the background section how PLS derived regression vectors could be inserted into a matrix and a PCA-like residual calculated. In this section we will use that method on the exact same data set used in the previous examples.

The number of latent variables to be used in the PLS model was fixed at 5. The control limits for the mean and variance of the PLS based residuals were obtained through cross validation. In this procedure a PLS model was obtained for the entire data set of 1000 points. Residuals were then calculated based on the model for 1000 test sets of 50 samples each. The 99% control limits were then selected based on the total of 1000 trials. The resulting control limits are shown in Figures 14 and 15 below, along with the worst calculated cases. Alternately, control limits could have been calculated based on the observed variance in the PLS residuals in a manner analogous to that used for the PCA model control limits.

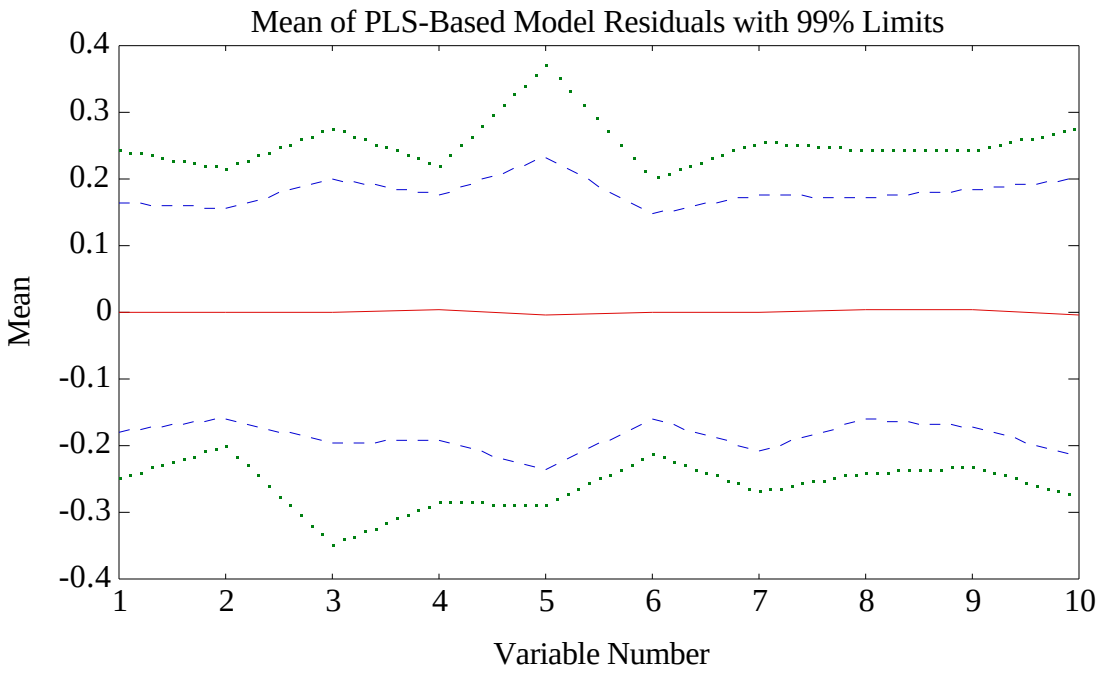

Figure l4. Mean of PLS-Based Model Residuals with 99% Confidence Limits (--) and Worst Case (...)

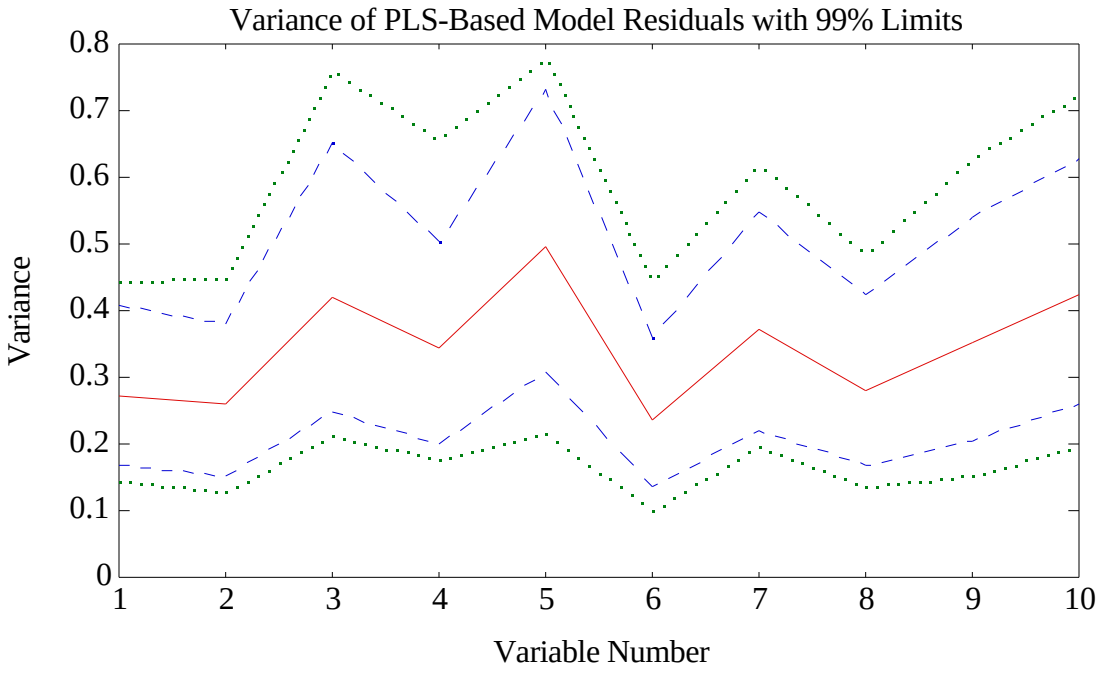

Figure l5. Variance of PLS-Based Model Residuals with 99% Confidence Limits (--) and Worst Case (...)

It is interesting to compare the variance of the PLS models residuals to those obtained for the PCA models. The PLS residuals have considerably more variance. This is not surprising, though, because of the difference in the way the residuals are actually calculated. The PLS prediction for each variable is based only on the other variables, each variable does not contribute to its own prediction. This is not the case with PCA. The projection operation in PCA includes information from every variable. In essence each variable is used in its own prediction. The result of this is that the PLS residuals are typically much larger than the PCA residuals.

The PLS model was tested with the same data used in the test of the PCA model. The same data was used in the examples below that was used for the PCA model tests shown in the previous section. The results of the bias test are shown in Figures 16 and 17 below. Figure 16 shows that in the test for a bias on sensor 5 the mean residual exceeds the limit at all three of the bias levels tested. In addition, some other variables exceed the limits for the high bias case. None of the variables exceed the limits for variance, shown in Figure 17, as would be expected. Figure 18 shows the ratio of the residual mean to the limits.

The results of the added noise test for the PLS model are shown in Figures 19 and 20. The added noise does not cause the mean residual of any of the variables to exceed the limit as expected. In the variance plot and the ratio plot shown in Figure 21, however, we see that the variance limit has been greatly exceeded on variable 7 in the high and intermediate noise case and modestly exceeded in the low noise case. Other variables exceed the limits in the high and intermediate case also, but by very small amounts relative to the PLS-based residuals.

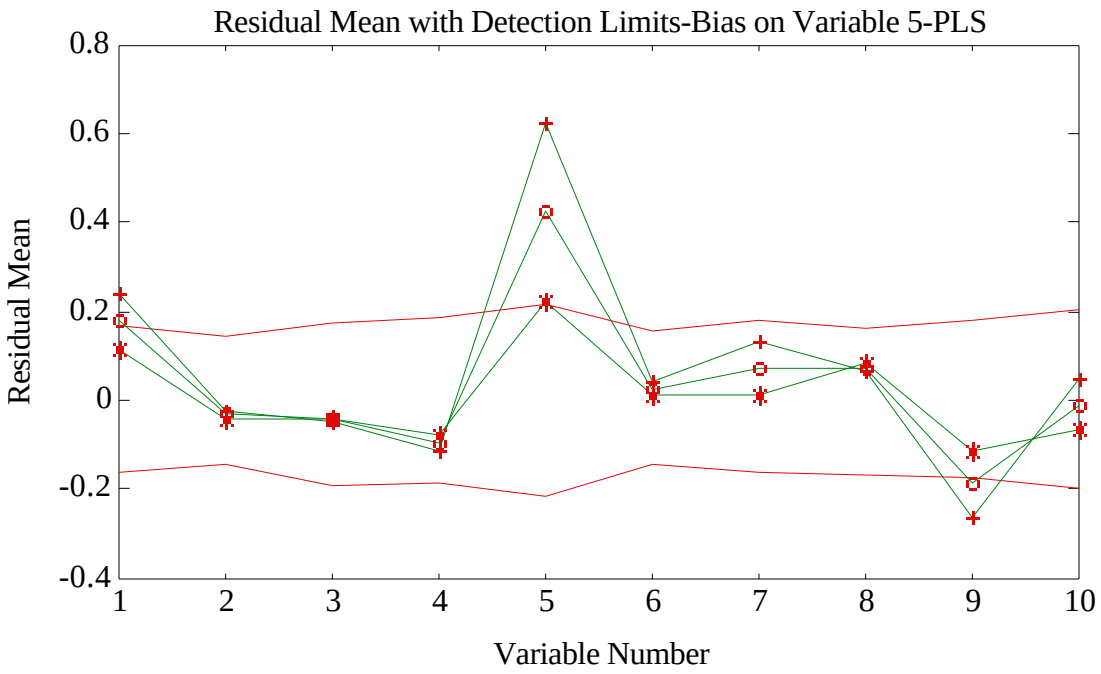

Figure 16. Residual Mean with Detection Limits for Bias of 0.2, 0.4 and 0.6 Standard Deviation on Variable 5-PLS Model

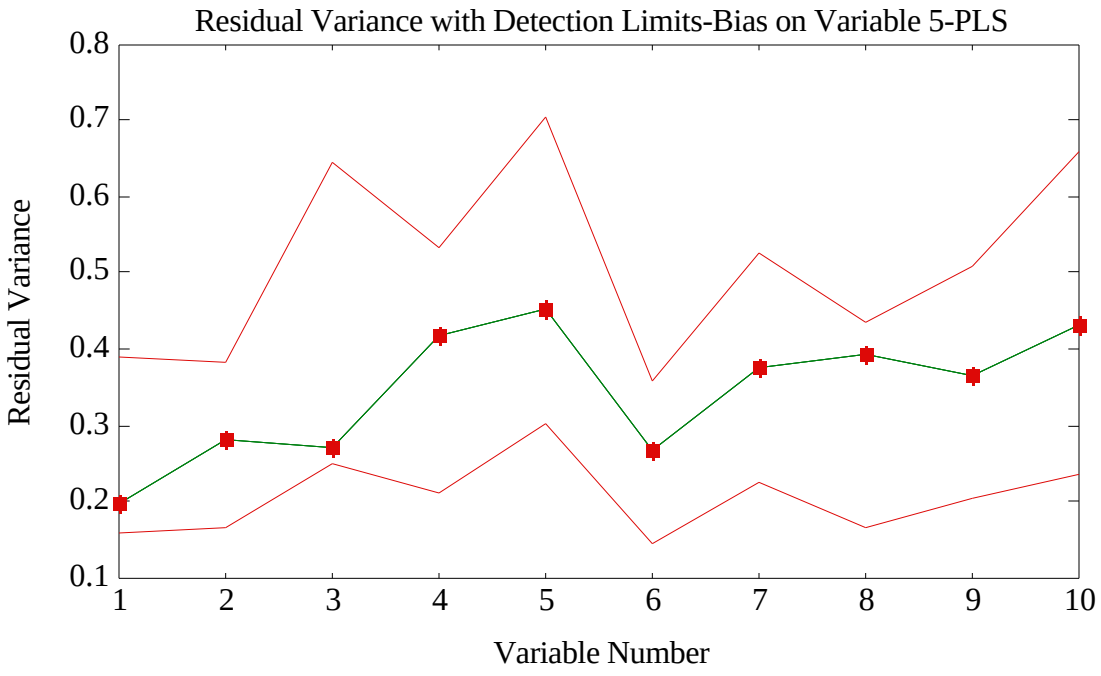

Figure 17. Residual Variance with Detection Limits for Bias of 0.2, 0.4 and 0.6 Standard Deviation on Variable 5-PLS Model

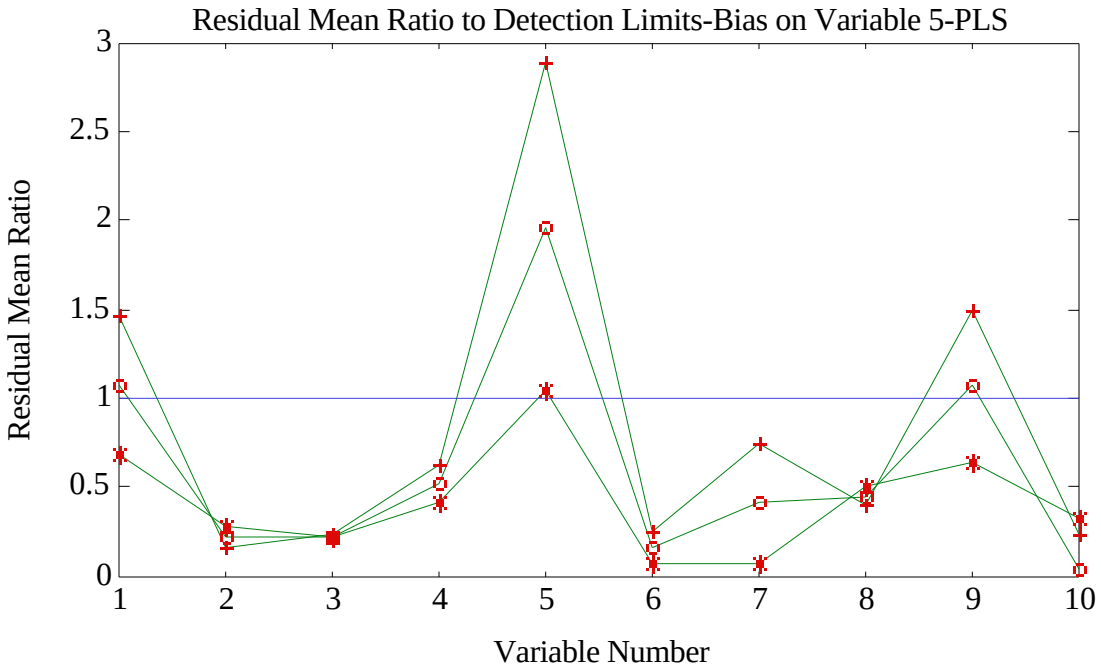

Figure 18. Residual Mean Ratio to Detection Limits for Bias of 0.2, 0.4 and 0.6 Standard Deviation on Variable 5-PLS Model

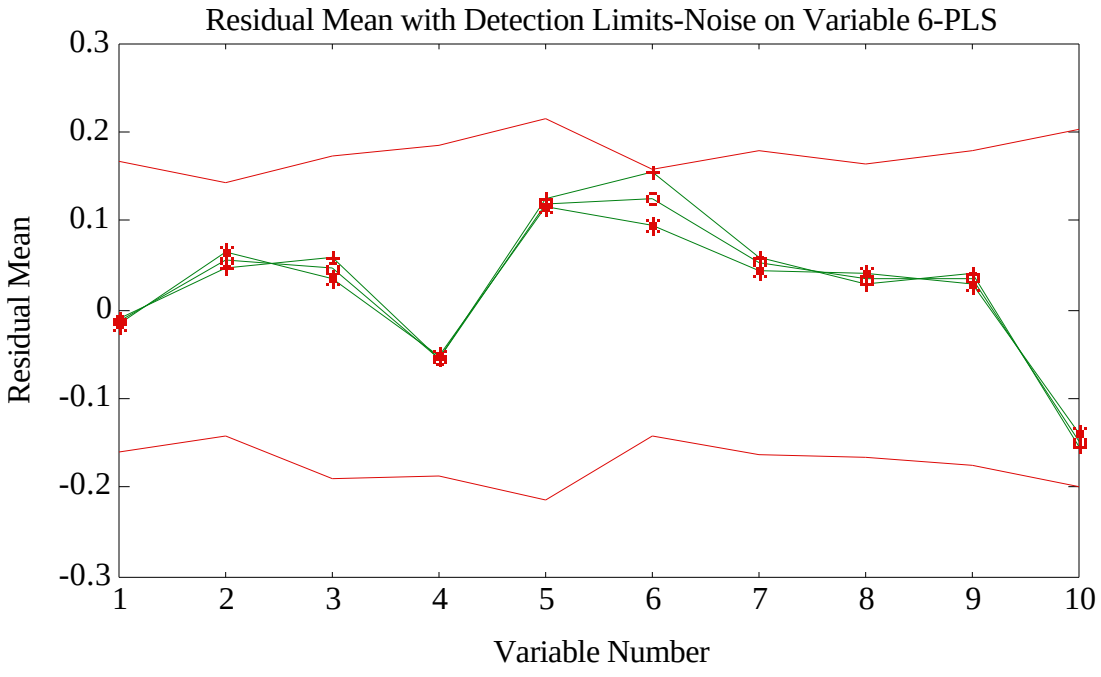

Figure 19. Residual Mean with Detection Limits for Noise of 0.4, 0.6 and 0.8 Standard Deviation on Variable 6-PLS Model

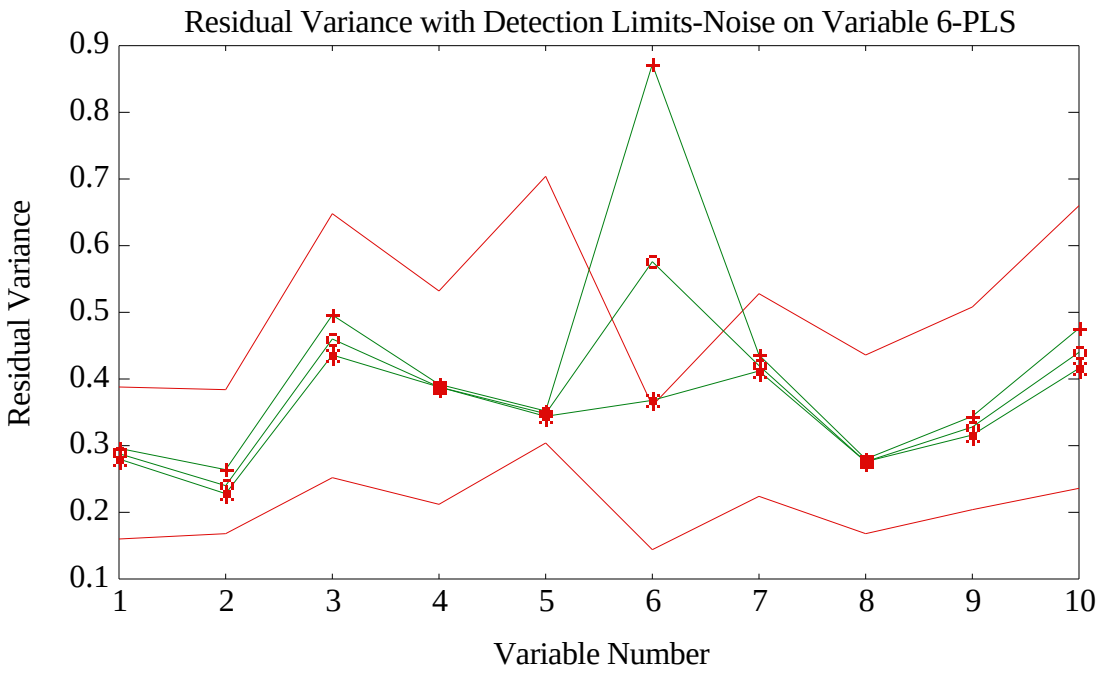

Figure 20. Residual Variance with Detection Limits for Noise of 0.4, 0.6 and 0.8 Standard Deviation on Variable 6-PLS Model

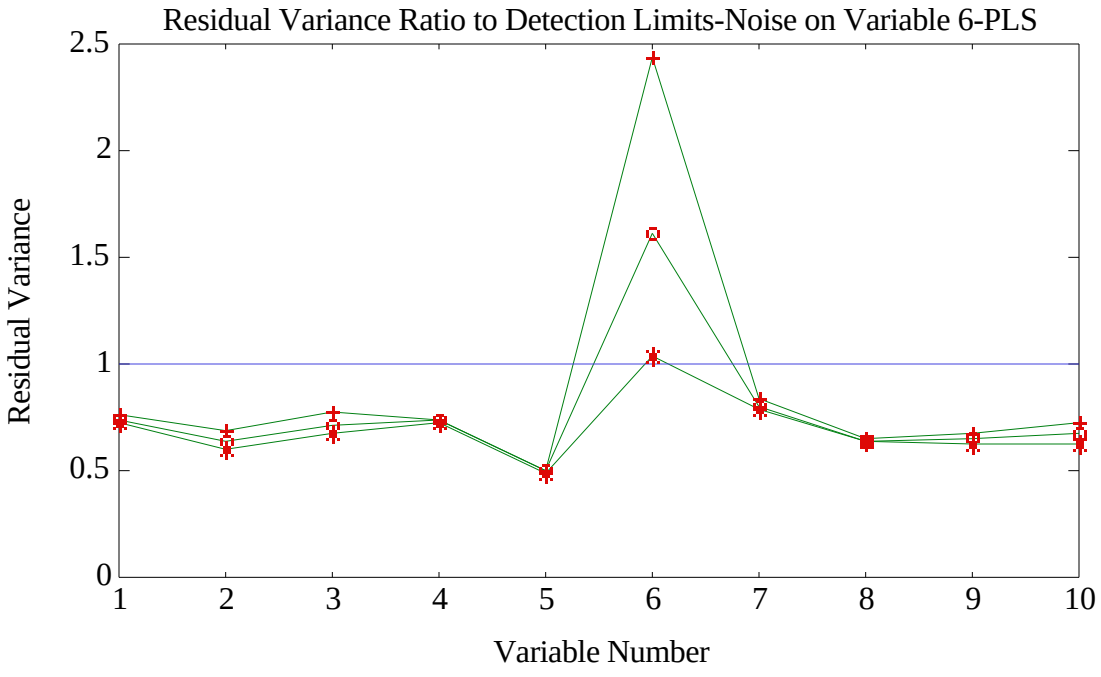

Figure 21. Residual Variance Ratio to Detection Limits for Noise of 0.4, 0.6 and 0.8 Standard Deviation on Variable 6-PLS Model

It is interesting to compare these results to those of the PCA based model. We see that the PLS based model does not tend to spread the error over other variables as much as the PCA model. Thus it appears to be more specific. At the same time, it appears to be more sensitive, exceeding the confidence limit at lower noise and bias levels. In order to compare the detection power of the models the control limits of the PCA model were converted to be on the same basis as the PLS model limits. This is done by first determining the ratio, **h**, of the change of a variable to the change in its residual, which is equal to the inverse of the diagonal elements of $(\mathbf{I} - \mathbf{P}_k\mathbf{P}_k')$

$$\mathbf{h} = (\mathrm{diag}(\mathbf{I} - \mathbf{P}_k\mathbf{P}_k'))^{-1} \qquad (27)$$

The PCA limits for mean can then be scaled by the **h** vector. The variance limits must first be converted to standard deviations, then scaled and converted back to variance limits. The results of this procedure allow us to compare the limits of the PCA and PLS models, which are shown in Figures 22 and 23.

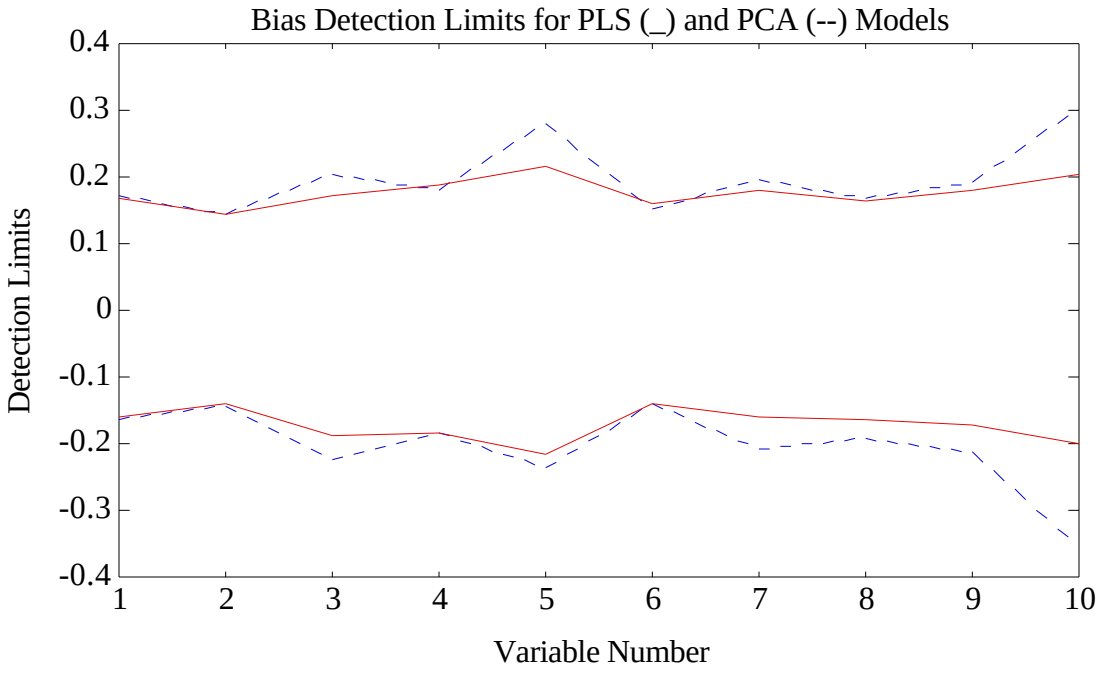

Figure 22. Detection Limits for Bias Errors For PLS(_) and PCA(--) Models

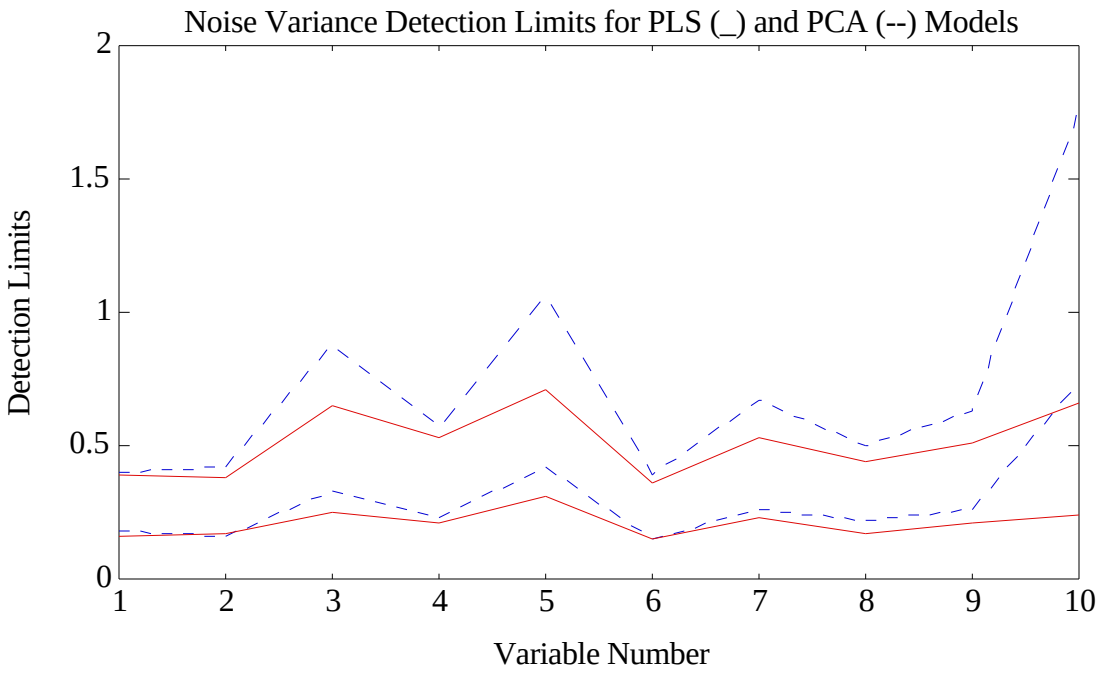

Figure 23. Detection Limits for Noise Errors for PLS(_) and PCA(--) Models

From the figures we can see that the control limits for the PLS based model are "tighter" than for the PCA model. For some variables they are about the same, but for some variables, such as 3, 5 and 10, the difference is significant.

In order to further investigate the ability of the models to detect errors a simulation was performed. In this test a set of 50 samples was selected at random from the data set 1000 times. Noise or bias errors were added to one variable in the test set. This was repeated 6 times for each variable in the test set: three amounts of bias and three noise levels were added to each variable in turn. The residuals of the perturbed data were then calculated based on the PLS or PCA models and compared to the limits using the ratio test. The number of times the perturbed variable was identified was recorded, along with the number of misidentifications and the number of cases were no variables were detected "out of bounds". The results of these trials are presented in Table 2. The number of misidentifications is listed under the column headed "Wrong" and the number of

instances where no variable was identified is under the column headed “None”.

Table 2. Results of Simulation to Test Error Detection Ability of PLS and PCA Based Models

Noise Test

	PLS Model			PCA Model		
Level	Correct	Wrong	None	Correct	Wrong	None
0.8	9893	29	78	8992	392	616
0.6	8910	293	797	7318	736	1946
0.4	4676	991	4333	3443	1137	5420
Total	23497	1313	5208	19753	2265	7982

Bias Test

	PLS Model			PCA Model		
Level	Correct	Wrong	None	Correct	Wrong	None
0.6	9998	2	0	9729	257	14
0.4	9808	157	35	8811	896	293
0.2	5536	1748	2716	4466	2262	3272
Total	25342	1907	2751	23006	3415	3579

The results show that the PLS based model provides significantly fewer “wrong” responses in any case. It is also more sensitive at the lower noise or bias levels than the PCA model.

3.3 The Effect of Serial Correlation on PCA and PLS Residuals

In this section we will consider the effect of serial correlation on the analysis of the residuals from PCA and PLS. We will use as an example data from a proprietary high temperature process that has 20 temperature measurements and 1 level measurement. In this system the temperatures are all highly correlated because the measurement locations are physically close in a process vessel. The temperatures are also correlated with the level, which varies. In addition, the data is serially correlated because of the thermal inertia and time behavior of the level.

For this example we had available a data record of 1400 samples. We will use the first 500 points in the record as the

training set and test our models using portions of the remaining 900 samples. The first step in the analysis was to scale the first 500 points to zero mean and unit variance. These scaling factors were retained so that they could be applied to the remainder of the data set. An initial survey of the data was performed with PCA so that any anomalous data points could be deleted, however, in this data set none were found.

In this process our time constant of interest is equal to ~25 sample intervals. This is based on both operator experience and calculation of the correlation time in variables. Thus we will use a "window" of 25 samples for our tests. In order to test the effect of the serial correlation we will calculate control limits for the PCA and PLS models in two ways. In the first case we will calculate the control limits just as in the previous section. In this case the order of the samples was not retained. For the PLS model 200 repetitions were used. In the second case the order of the samples was retained. A test set of 25 consecutive samples was chosen and this procedure was repeated 250 times, so that the test set indices could be incremented by 2 units each time to cover the entire data set. The control limits were set based on the 99th percentile of the calculated residual mean and variance. The control limits for both the PCA and PLS models were set in this way.

The resulting confidence limits, calculated both ways, for the mean and variance of the PCA models are shown in Figures 24 and 25. The corresponding limits for the PLS model are shown in Figures 26 and 27. The dashed line in the figures corresponds to the randomized data, while the solid line is for the data in which the order was retained. Note how the confidence limits for the mean in Figures 24 and 26 have widened considerably for the case of the serially correlated data, as might be expected. This is due to the correlation between residuals in samples that are close together in time. The confidence limits on the variance in Figures 25 and 27, however, are about the same for both the correlated and uncorrelated data. This is probably not a general result. In general we would expect to see somewhat less variance in the non-randomized data because the serial correlation. If the window width

is large relative to the correlation time, however, we would expect the variance to be the same in either case.

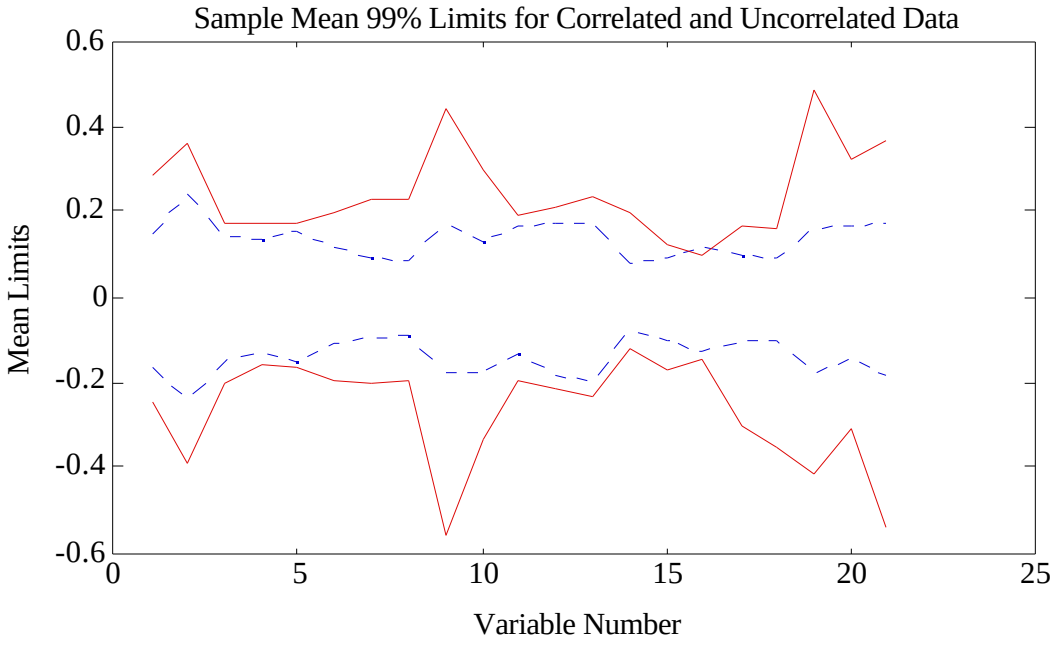

Figure 24. Residual Mean 99% Confidence Limits for Random (dashed) and Serially Correlated (solid) Data-PCA Model.

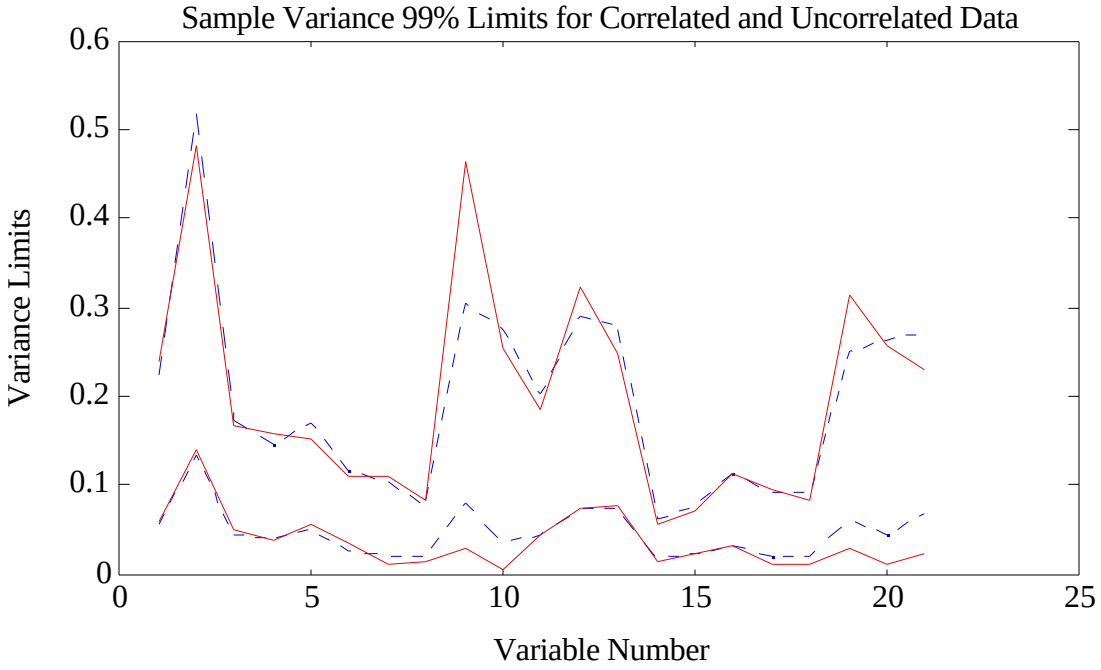

Figure 25. Residual Variance 99% Confidence Limits for Random (dashed) and Serially Correlated (solid) Data-PCA Models.

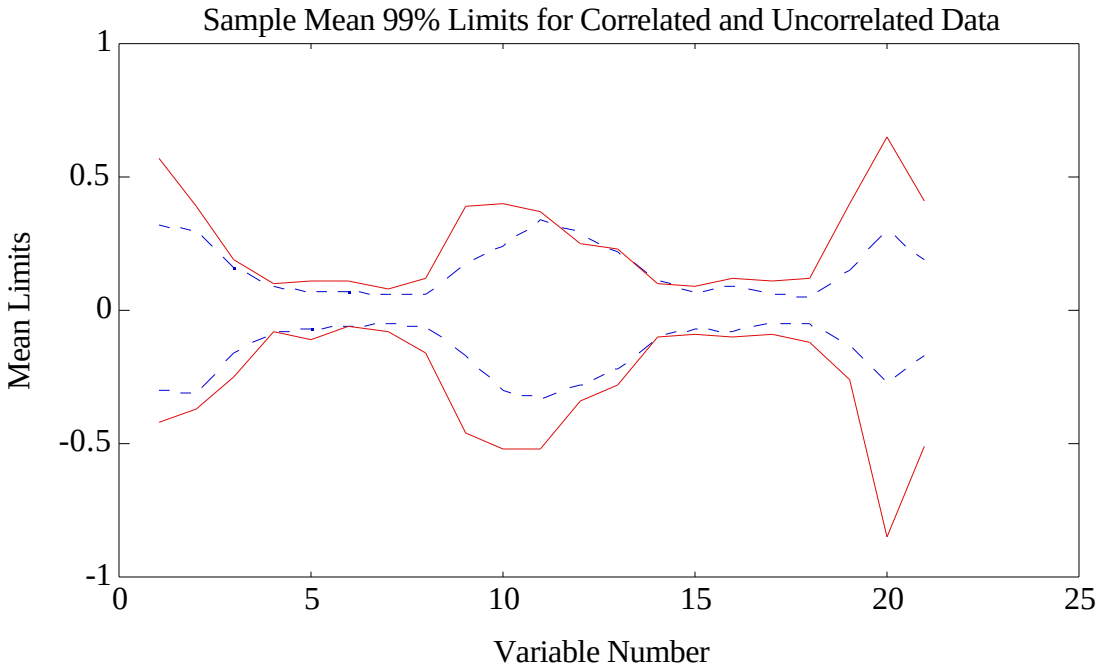

Figure 26. Residual Mean 99% Confidence Limits for Random (dashed) and Serially Correlated (solid) Data-PLS Model.

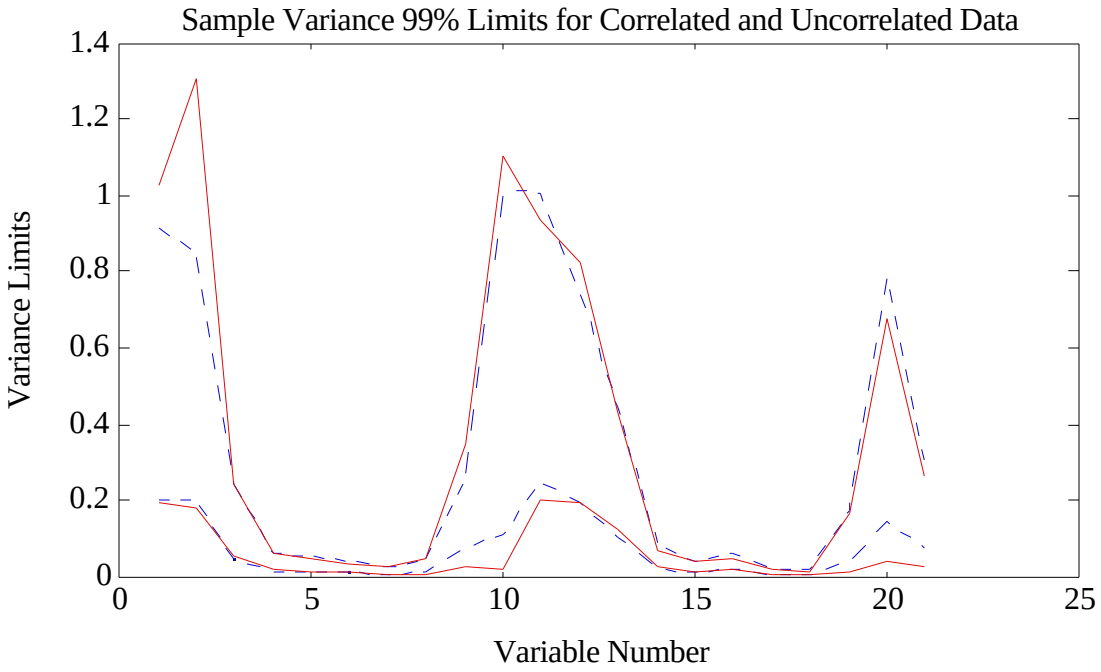

Figure 27. Residual Variance 99% Confidence Limits for Random (dashed) and Serially Correlated (solid) Data-PLS Model.

Some examples of simulated noise errors are shown in Figures 28 - 31 for the PLS and PCA models. In Figure 28 we see the effect on the PLS residuals of adding progressively more noise to variable 7. Note that variable 7 does not have the largest residual. When the residuals are ratioed to the control limit, however, the fault in variable 7 is clear, as shown in Figure 29. Similar results are shown for the PCA model in Figures 30 and 31. In the PCA case, however, the model does not detect the error for the low noise case.

The detection limits were calculated as in the previous section for the PLS and PCA models, and rescaled back to the original units of the variables. The actual detection limits in degrees C for noise and bias errors are shown for the models in Figures 32 and 33. The PLS model has considerably tighter limits for detecting bias errors than the PCA model in all but one variable. We see similar results for the detection of noise errors, with the PLS model having tighter limits in all but one variable.

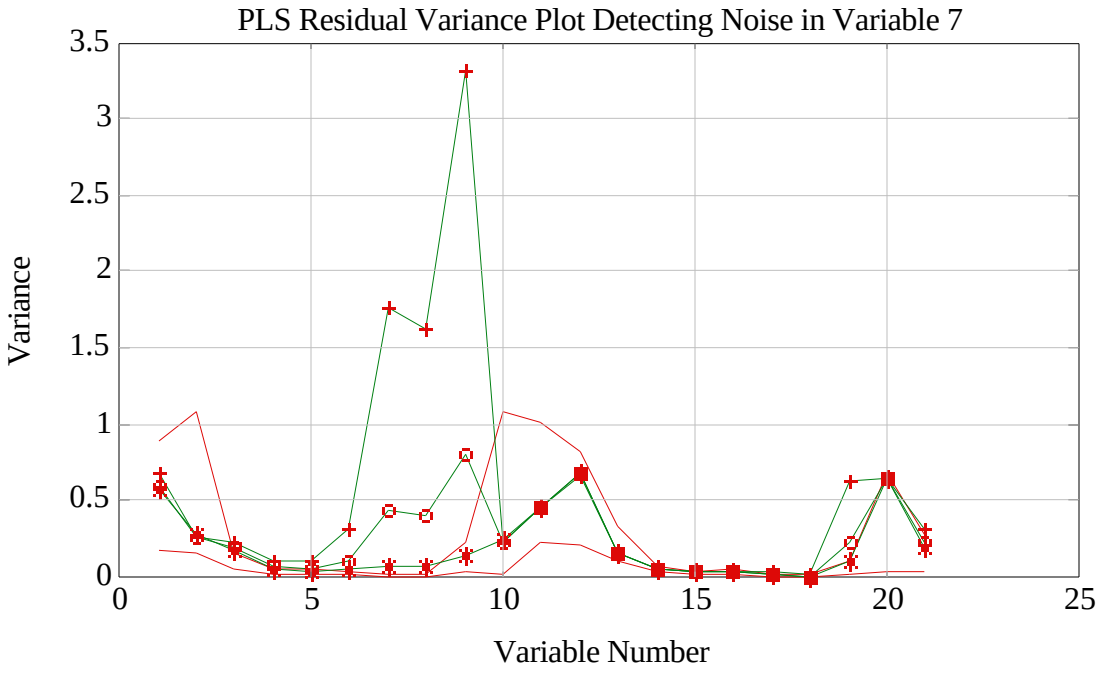

Figure 28. Variance of Residuals with Control Limits for Noise of 0.2, 0.5 and 1.0 on Variable 7-PLS Model

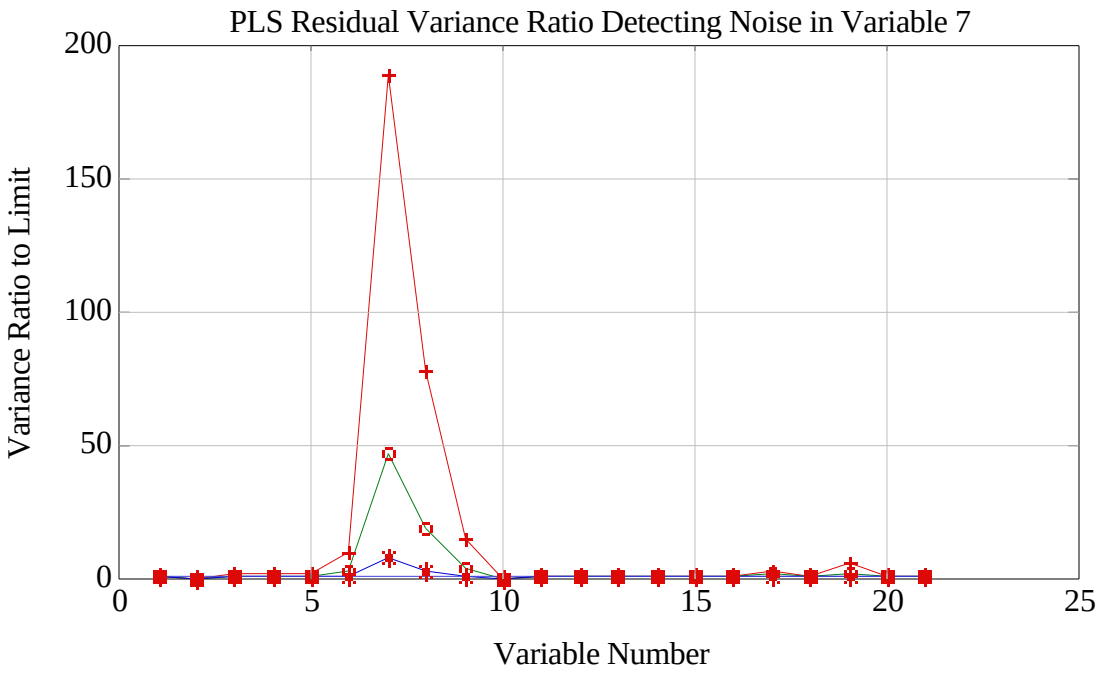

Figure 29. Variance of Residuals Ratio to Control Limit for Noise of 0.2, 0.5 and 1.0 on Variable 7-PLS Model

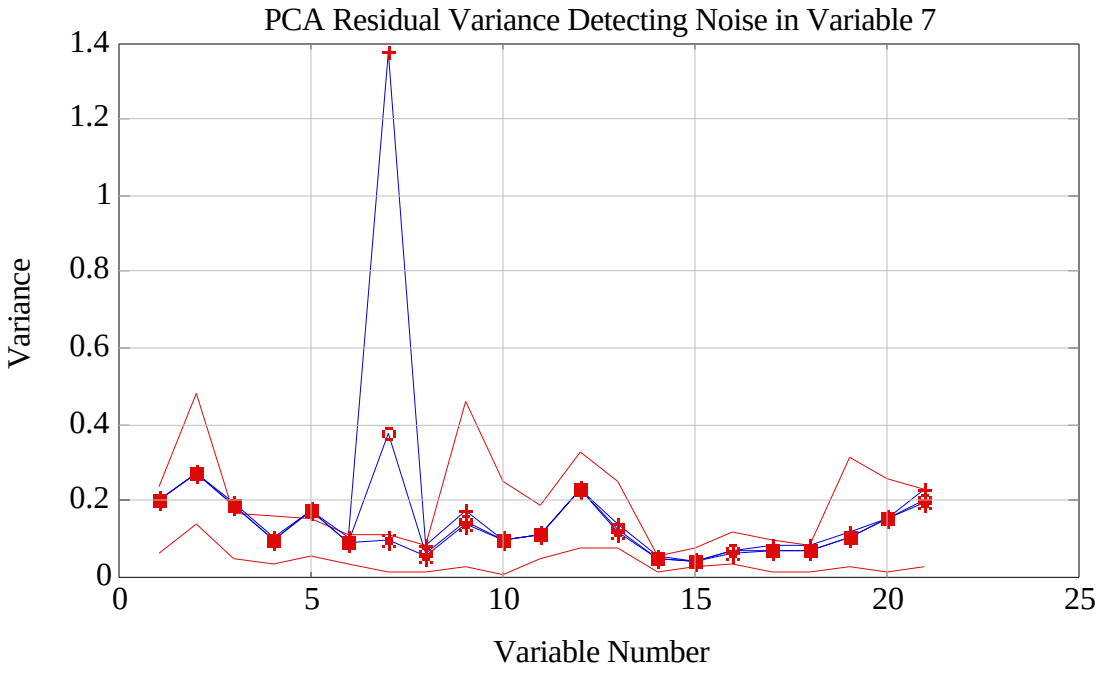

Figure 30. Variance of Residuals with Control Limits for Noise of 0.2, 0.5 and 1.0 on Variable 7-PCA Model

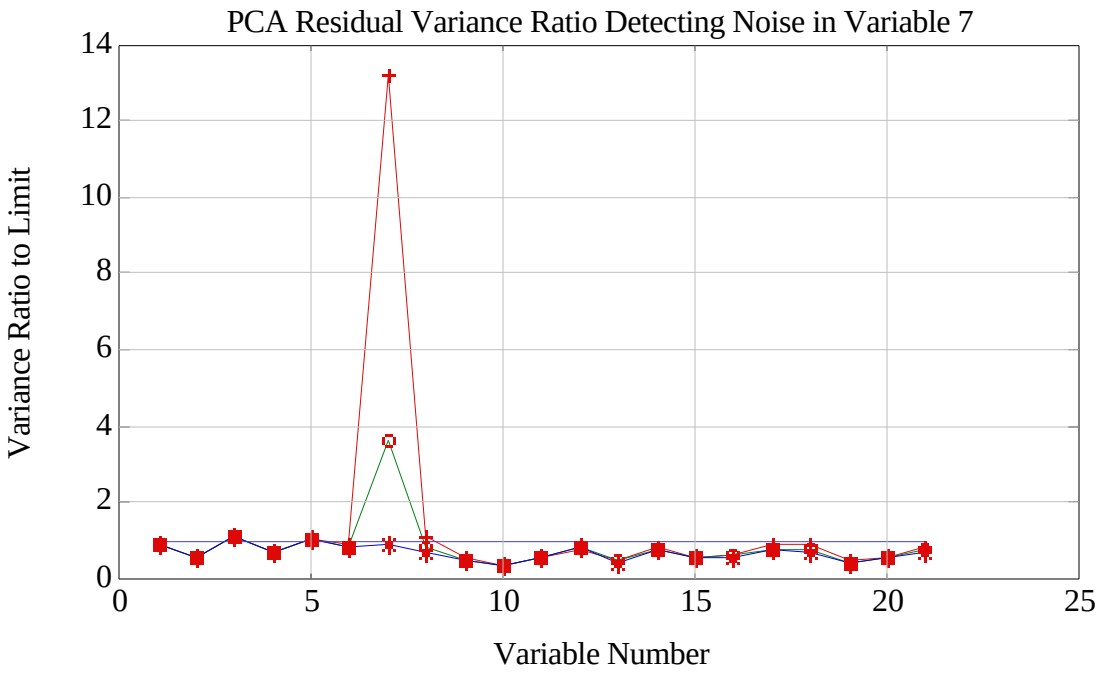

Figure 31. Variance of Residuals with Control Limits for Noise of 0.2, 0.5 and 1.0 on Variable 7-PCA Model

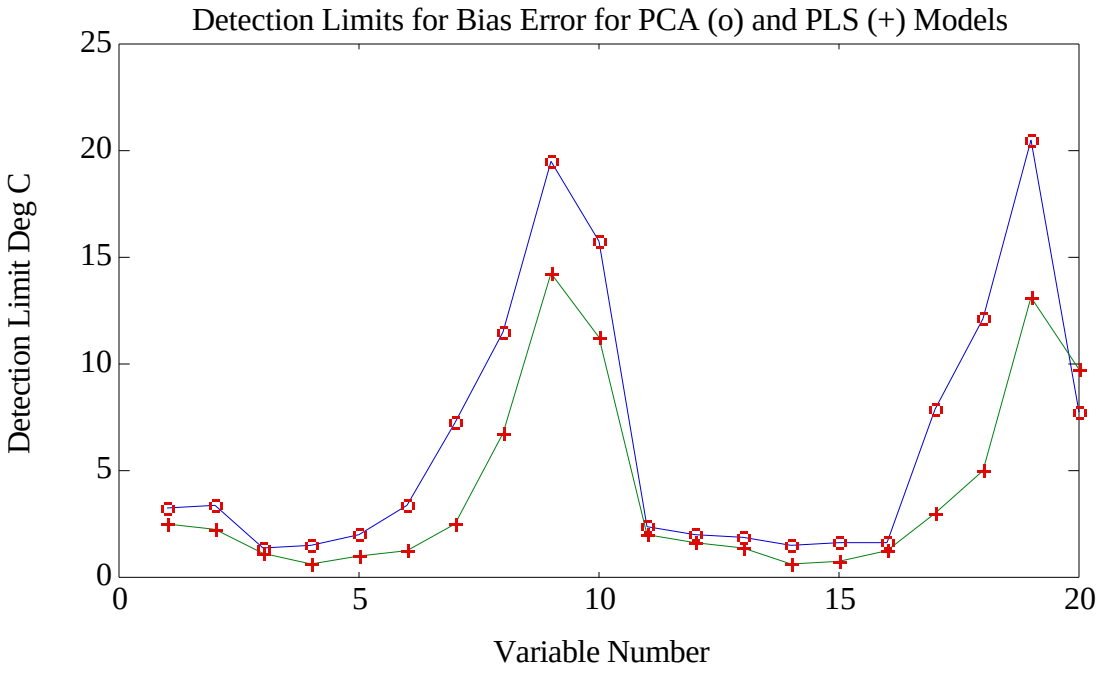

Figure 32. Detection Limits for Bias Errors For PLS and PCA Models in Original Units

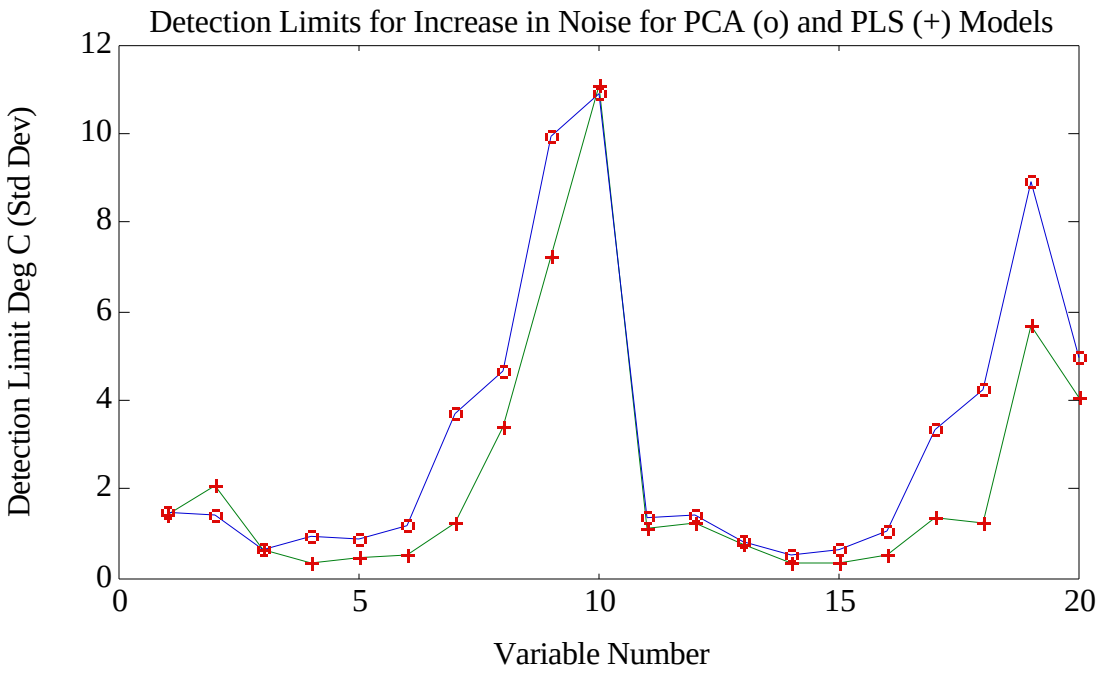

Figure 33 Detection Limits for Noise Errors For PLS and PCA Models in Original Units

In order to better determine the relative advantages of the PCA and PLS models a simulation was performed where data was taken from the test set and noise (with std. dev. of 0.2, 0.5 and 1.0 times the variable std. dev.) and bias errors (of 0.4, 0.6 and 0.8 times) were added in a manner analagous to the procedure used with the synthetic data. A 10 PC model was tested in addition to the 6 PC model used in the examples above. The results of the trials are shown in Table 3, where the number of correct responses are shown out of a possible 3675 (21 variables times 175 trials).

Table 3. Comparison of PLS and PCA Models for Detecting Errors in Process Data-Correct Responses

	Bias Errors			Noise Errors		
	PLS	PCA-6	PCA-10	PLS	PCA-6	PCA-10
High	2920	2963	2375	2603	2695	2126
Med.	2122	1929	1971	1929	1704	1791
Low	1191	362	796	1238	479	911
Total	6233	5254	5142	5770	4878	4828

For the high noise and bias cases the 6 PC model has the largest number of correct responses and was followed closely by the PLS model. At the medium and low noise and bias cases, however, the PLS model provided the most correct responses. It is interesting to note the behavior of the 6 PC model relative to the 10 PC model. The 10 PC model detects small errors much more often than the 6 PC model, but is wrong more often in the large error case.

5.0 Conclusions

In this paper we have shown how control limits can be placed on the residuals associated with PCA models. These limits enable us to detect failed sensors unambiguously in most instances. This would not be possible without the limits since the residual variance of individual variables can vary widely.

It has been shown that Partial Least Squares can be used to generate PCA-like residuals. Limits can be placed on the PLS residuals in a manner analogous to the PCA case. It appears that these PLS models are more sensitive than the PCA models for detection of sensor faults.

The effect of serial correlation on the calculated control limits has been demonstrated. This must be accounted for in setting limits.

Finally, we have some indication that for PCA models there is a trade off between sensitivity and selectivity as the model order is increased.

6.0 Acknowledgement

This work was supported in part by a supplement to National Science Foundation grant number ISI-8415075. The authors would also like to thank Bruce R. Kowalski of the Center for Process Analytical Chemistry and the Laboratory for Chemometrics for introducing all of us to the concept of MSPC.

7.0 References

[1] M. A. Sharaf, D. L. Illman and B. R. Kowalski, *Chemometrics*, John Wiley and Sons, New York (1986).

[2] D. J. Veltkamp, "User's Guide to PCA Modeling Program Version 1.0", The Center for Process Analytical Chemistry, University of Washington, (1989).

[3] G. Strang, *Linear Algebra and Its Applications*, Academic Press New York, 1980.

[4] J. E. Jackson and G. S. Mudholkar, "Control Procedures for Residuals Associated With Principal Component Analysis", *Technometrics* **21**(3), pp. 341-349, 1979.

[5] J. Edward Jackson, *J. Qual. Tech.*,**12**(4), 1980.

[6] J. Edward Jackson, *J. Qual. Tech.*, **13**(1), 1981.

[7] J. Edward Jackson, *J. Qual. Tech.*, **13**(2), 1981.

[8] H. Hotelling, "Multivariate Quality Control Illustrated by the Air Testing of Sample Bombsights," in *Techniques of Statistical Analysis*, eds. C. Eisenhart, M. W. Hastay, and W. A. Wallis, New York:McGraw, pp. 111-184 (1947).

[9] A. Hoskuldsson, "PLS Regression Methods", Journal of Chemometrics, Vol. 2, 211-228 (1988)

[10] P. Geladi, "Notes on the History and Nature of Partial Least Squates (PLS) Modelling", *Journal of Chemometrics*, Vol. 2, 231-246 (1988)

[11] A. Lorber, L. E. Wangen and B. R. Kowalski, *J. Chemometrics* **1**, 19 (1987).

[12] K. R. Beebe and B. R. Kowalski, *Anal. Chem.* **59**, 1007A (1987).

[13] N. L. Ricker, *Ind. Eng. Chem. Res.*, **27**(2), 1988.

[14] D. M. Himmelblau, *Fault Detection and Diagnosis in Chemical and Petrochemical Processes*, Elsevier Scientific Publishing Company, New York (1978).

[15] T. Mejdell and S. Skogestad, "Estimate of process outputs from multiple secondary measurements," 1989 American Control Conference Proceedings"

UPSET AND SENSOR FAILURE DETECTION IN MULTIVARIATE PROCESSES

Barry M. Wise, N. Lawrence Ricker and David J. Veltkamp

Prepared for presentation at the 1989 AIChE Annual Meeting
November 5-10, 1989.
Symposium on Statistics and Quality Control