Using Principal Component Analysis Modeling to Monitor Temperature Sensors in a Nuclear Research Reactor

Rosani M. L. Penha
Centro de Energia Nuclear
Instituto de Pesquisas Energéticas e Nucleares - Ipen
São Paulo, SP 05508-900 Brasil
Email: rmpenha@net.ipen.br
Phone: (11) 3817-7421

J. Wesley Hines
Nuclear Engineering Department
The University of Tennessee
Knoxville, TN 37996-0750 USA
Email: hines@utkux.utk.edu
Phone: (865) 974-6561

Abstract

Principal Component Analysis (PCA) is a data-driven modeling technique that transforms a set of correlated variables into a smaller set of new variables (principal components) that are uncorrelated and retain most of the original information. Thus, a reduced dimension PC model can be used to detect and diagnose abnormalities in the original system in a robust way. This paper presents an application of using the PCA modeling technique to detect abnormalities in temperature sensors used to monitor the primary loop of a 2MW research pool reactor. The PCA model maps the temperature variables into a lower dimensional space and tracks their behavior using T^2 and Q statistics. The Hotelling's T^2 statistic measures the variation within the PCA model while the Q statistic measures the variation outside of the PCA model. Five temperature sensors are considered in the model. Three sensors are located inside the pool and two sensors are located at the primary loop piping. The reduced dimension PCA model has well behaved T^2 and Q statistics. To test the functionality of the model, a drift was imposed sequentially to each temperature sensor, and the PCA model was able to detect and identify the faulty sensors at very low thresholds.

Introduction

Many researchers have addressed the use of Principal Component Analysis (PCA) modeling in the monitoring and fault detection of process sensors [1, 2, 3]. The objective of this work is to construct a PCA model that best maps a research reactor's primary loop temperature sensors into a lower dimensional space, in order to characterize the behavior of these variables through the use of T^2 and Q statistics.

Five sensors have been chosen to be monitored. Three sensors are located inside the pool: T1 is located near the surface of the pool, T2 is located at the midway down into the pool and T3 is located just above of the reactor core. Temperature sensor T4 is located in the outlet pipe that takes the water from the core to the decay tank. At the outlet of the decay tank we have the sensor T6. The schematic diagram of the pool research reactor is shown in the Appendix.

Principal Component Analysis

PCA is a method used to transform a set of correlated variables into a smaller set of new variables that are uncorrelated and retain most of the original information, where the variation in the signals is considered to be the information. PCA takes advantage of redundant information existent in highly correlated variables to reduce the dimensionality. After developing a model using good (training) data, the reduced dimension PC model can be use to detect and diagnose process abnormalities in a robust way [4]. For a basic reference book on PCA see Joliffe [5].

PCA decomposes the data matrix X (m samples, n variables) as the sum of the outer product of vectors $\mathbf{t_i}$ and $\mathbf{p_i}$ plus a residual matrix E [1]:

$$\mathbf{X} = \mathbf{t_1 p_1}^T + \mathbf{t_2 p_2}^T + ... + \mathbf{t_k p_k}^T + \mathbf{E} = \mathbf{T}_k \mathbf{P}_k^T + \mathbf{E} \tag{1}$$

The vectors $\mathbf{p_i}$ are orthonormal, and the vectors $\mathbf{t_i}$ are orthogonal, that is:

$$\mathbf{p_i}^T \mathbf{p_j} = 1, \quad \text{if } i = j \quad \text{and} \quad \mathbf{p_i}^T \mathbf{p_j} = 0, \quad \text{if } i \neq j; \text{ and} \tag{2}$$

$$\mathbf{t_i}^T \mathbf{t_j} = 0 \quad \text{if } i \neq j \tag{3}$$

Also we can note that $\mathbf{t}_i$ is the linear combination of the original $\mathbf{X}$ data defined by the transformation vector $\mathbf{p}_i$:

$$\mathbf{X p}_i = \mathbf{t}_i \tag{4}$$

The vectors $\mathbf{t}_i$ are known as the principal component *scores* and contain information on how the *samples* are related to each other. The $\mathbf{p}_i$ vectors are the *eigenvectors* of the covariance of matrix **X.** They are known as the principal component *loadings* and contain information on how *variables* are related to each other. In fact, the PCA splits the data matrix **X** in two parts: one describes the system variation (the process model $\mathbf{T}_k \mathbf{P}^T_k$) and the other captures the noise or unmodeled information (residual variance **E**). This division is not always perfect, but it routinely provides good results [1]. It is very important to distinguish the number of components (dimension) that are to be kept in the model.

The number of *principal components* k, to retain in the model must be less than or equal to the smaller dimension of X, i.e., $k \leq \min\{m, n\}$ [2]. The goodness of the model depends on a good choice of how many PCs to keep.

There are different criteria to choose the number of PCs [6]. The eigenvalues associated with each eigenvector or principal component: $\mathbf{p}_i$, tell us how much information (variation) each PC explains. Then we can look at the cumulative percent variance captured by the first few PCs and choose a number of PCs that accounts for most of the variability of the data (i.e. 90% to 99%).

Alternatively, we can look for a *knee* point in the residual percent variance plotted against the number of principal components. This is thought to be the natural break between the useful PCs and residual noise.

Another criterion is to accept the PCs whose eigenvalues are above the average eigenvalue. In correlated-based PCA, the average eigenvalue is 1. It is advisable to investigate more than one criterion since there is no universally accepted methodology.

PCA Model

The concept of principal components is shown graphically in Figure 1. The figure shows a PCA model constructed for a data set of three variables. We can see that the samples lie mainly on a plane, thus the data is well described by a two principal component model (2 PCs). The first PC aligns with the greatest variation in the data while the second PC aligns with the greatest remaining variance that is orthogonal to the first PC.

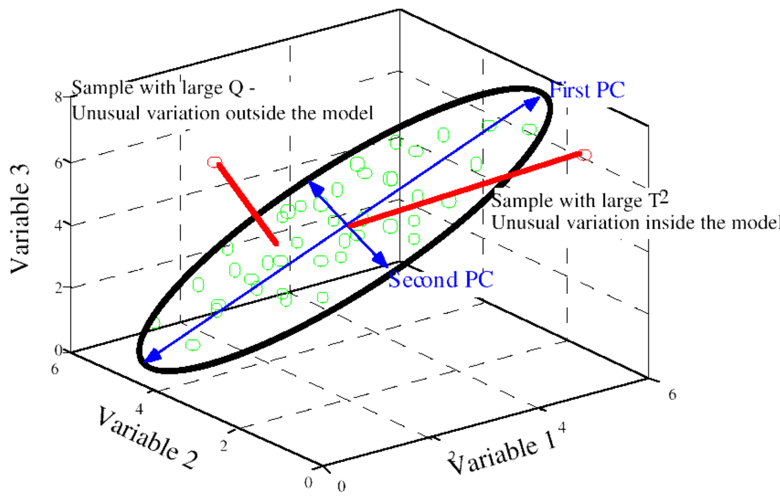

Figure 1 - Principal Component Model Staying on a Plane, Showing T^2 and Q Outliers

There are two statistics commonly employed in evaluating new data using a previously developed PCA model: Q statistic and Hotteling's T^2 statistic.

The Q statistic measures the *lack of model fit* for each sample. It indicates how well each sample conforms to the PCA model by measuring the distance a data point falls from the PC model. It is calculated as the difference between the data point and its projection on the PC model. It gives the lack of fit to the model [1].

$$\mathbf{Q}_i = \mathbf{e}_i\mathbf{e}_i^T = \mathbf{x}_i(\,\mathbf{I} - \mathbf{P}_k\,\mathbf{P}_k^T)\,\mathbf{x}_i^T \qquad (5)$$

The Hotteling's T^2 measures the variation *within* the PCA model. T^2 is the sum of the normalized squared scores defined as [1]:

$$T^2_i = \mathbf{t}_i(\mathbf{T}_k^T\mathbf{T}_k)^{-1}\mathbf{t}_i^T \qquad (6)$$

Statistical limits can be developed for sample scores, T^2 and Q, and individual residuals. If some point falls outside the limits for a specific confidence interval (95%, for example), this point may be considered to be an outlier and may be not representative of the data used to develop the PCA model.

The PCA model of a data matrix includes mean and variance scaling vectors, eigenvalues, loadings, statistics limits on the scores, Q and T^2. The model can be used with new process data to detect changes in the system that generated the original data. After detecting a probable outlier due to extreme T^2 or Q values, we can investigate the inputs that contribute to the extreme statistical value.

Methodology

A PCA model was developed to describe correlations of the primary loop temperature variables of the IPEN research reactor located in Sao Paulo, Brazil (see Appendix). The reactor data acquisition system records a snapshot of data once a minute for a complete fuel cycle that usually lasts a couple of days. A representative data set that corresponds to 25 hours of reactor operation was used to construct the model. A second data set, corresponding to 15 hours of another cycle operation, is used to validate the PCA model. Lastly, an artificial drift is imposed on each sensor to test the sensitivity of the model.

The input data used to develop the model was arranged in a matrix with 1501 rows (samples) and 5 columns (temperature variables). The time plot of these variables is shown in Figure 2.

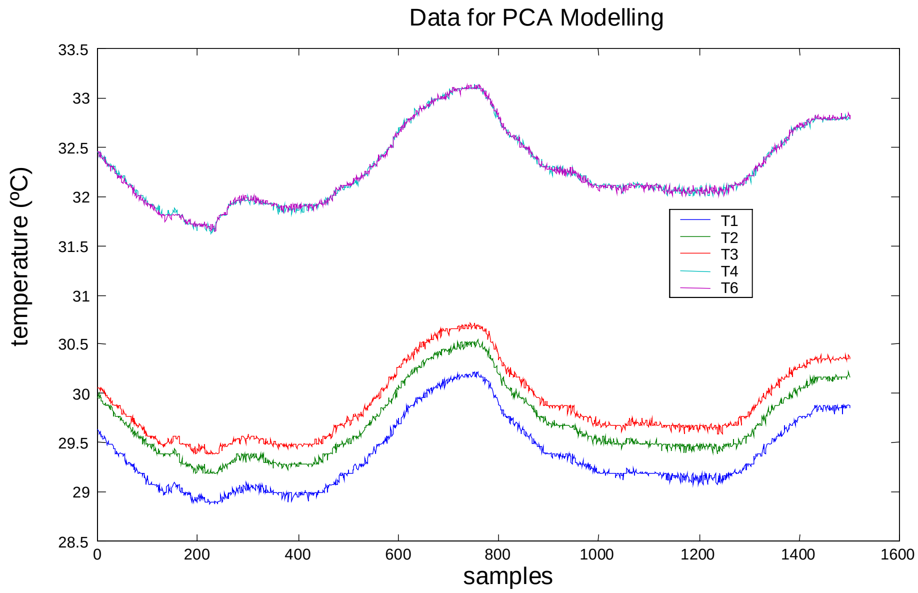

Figure 2 - Temperature Model Development Data

As shown in Table 1; T1, T2, T3, T4 and T6 are highly correlated, meaning that they vary together. This redundancy in the measurements allows us to build a PCA model that will retain the most of information in a few principal components.

corr. coef.	T1	T2	T3	T4	T6
T1	1.0000	0.9963	0.9960	0.9918	0.9903
T2	0.9963	1.0000	0.9945	0.9887	0.9862
T3	0.9960	0.9945	1.0000	0.9939	0.9927
T4	0.9918	0.9887	0.9939	1.0000	0.9975
T6	0.9903	0.9862	0.9927	0.9975	1.0000

Table 1 - Correlation Coefficients

To construct the PCA model, the input matrix was divided into a training set (to develop the model) and a test set in an odd-even manner. The input matrix was mean centered and scaled to a unit variance. This is necessary for PCA model development. PCA functions in MATLAB were used to calculate the principal components, the eigenvalues, and the amount of variance explained by each PC component.

Figure3 is a plot of the eigenvalues versus the PC number and is used to help to choose the number of PCs to keep in the model. The size of the eigenvalue equals the amount of variance explained in the corresponding PC. We use the log plot that can show the break when the eigenvalues cover several orders of magnitude, as in this case. This plot is used to identify a knee point where the PCs above the knee contain information and the PCs below the knee represent noise. The knee occurs between 1 and 2 PCs. The second PC does contribute useful information. We will keep 2 PCs that explain 99.80% of the variance.

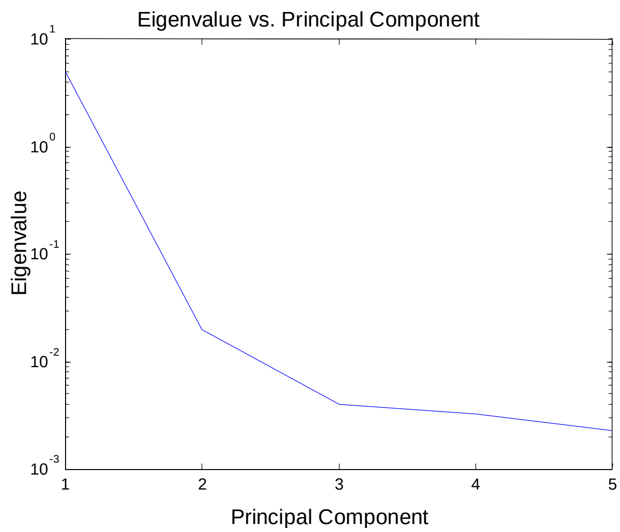

Figure 3 - Eigenvalues versus Principal Components

Next we plot and analyze the loadings on the retained principal components: PC1and PC2.

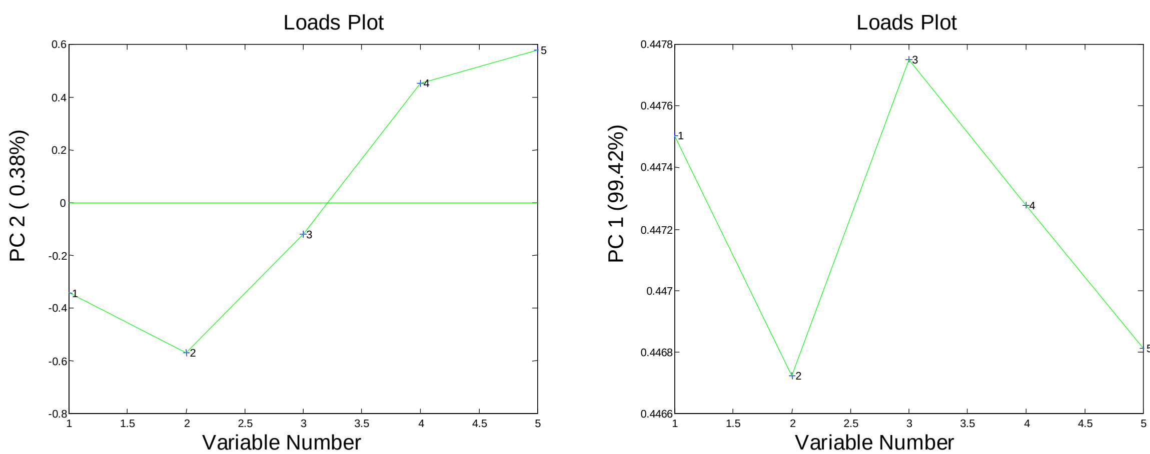

Figure 4 - Loadings on Principal Component 1 and 2

The principal component loadings are the weightings of each input to the specific PC. They show the underlying relationship among the variables. PC1 weights all the variables positively, so it is a gross measurement of the temperature in the reactor. PC2 accounts for the differences between variables {T1, T2, T3} and {T4, T6}. This makes sense, since the three first variables are located physically near each other and are related to the temperature inside the pool while the last two sensors are located outside the pool. Although PC2 accounts for a small amount of variation when compared to PC1, it is important to describe the differences between the two groups of variables.

Figure 5 is a plot of PC1 versus PC2 and shows that PC1 and PC2 are uncorrelated. If there were a noticeable relationship in this plot, it would be attributed to non-linear relationships in the data. The PC technique removes all linear correlations and results in a scatter plot when the non-linear relationships are small or nonexistent.

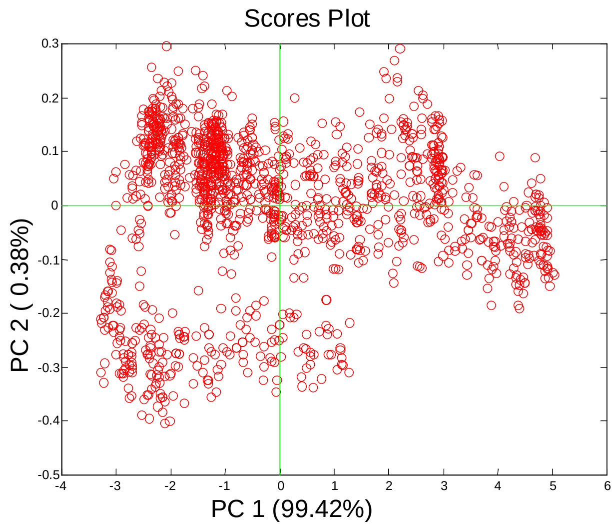

Figure 5 - Scores on PC1 versus Scores on PC2

The T² and Q statistics are shown in Figure 6. The dashed lines represent a 95% confidence interval used to identify possible outliers. The T² and Q residuals show the data fits the model well.

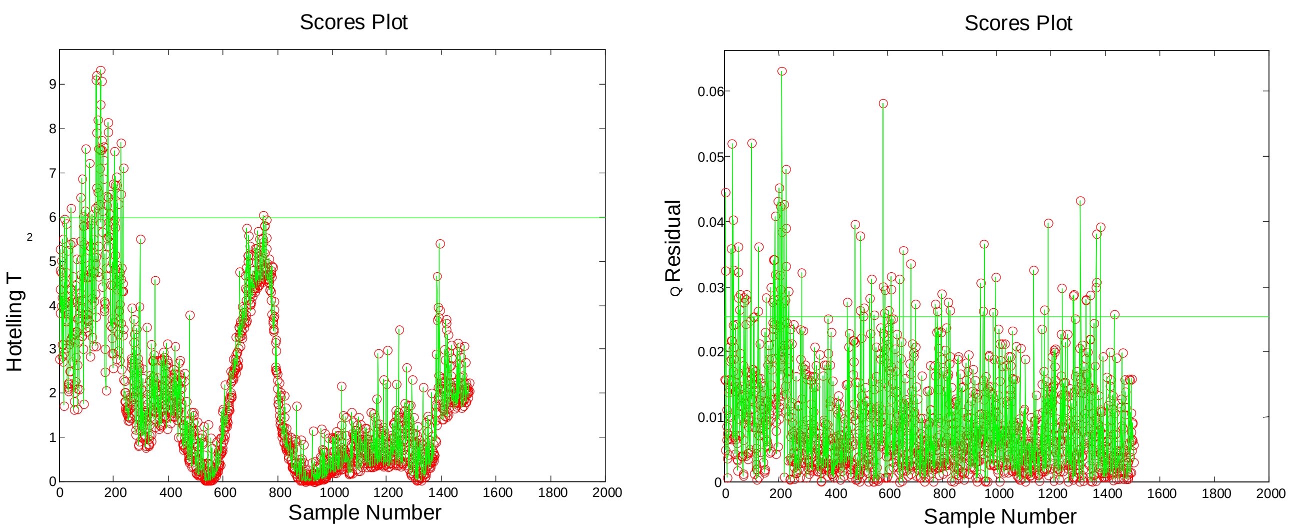

Figure 6 - T² Statistic on 2 PC Model (left) and Q Statistic on 2 PC Model (right)

Validation

To validate the PCA model, another data set corresponding to another week operation was applied to the PCA model. The resultant T^2 and Q residuals are shown in Figure 7.

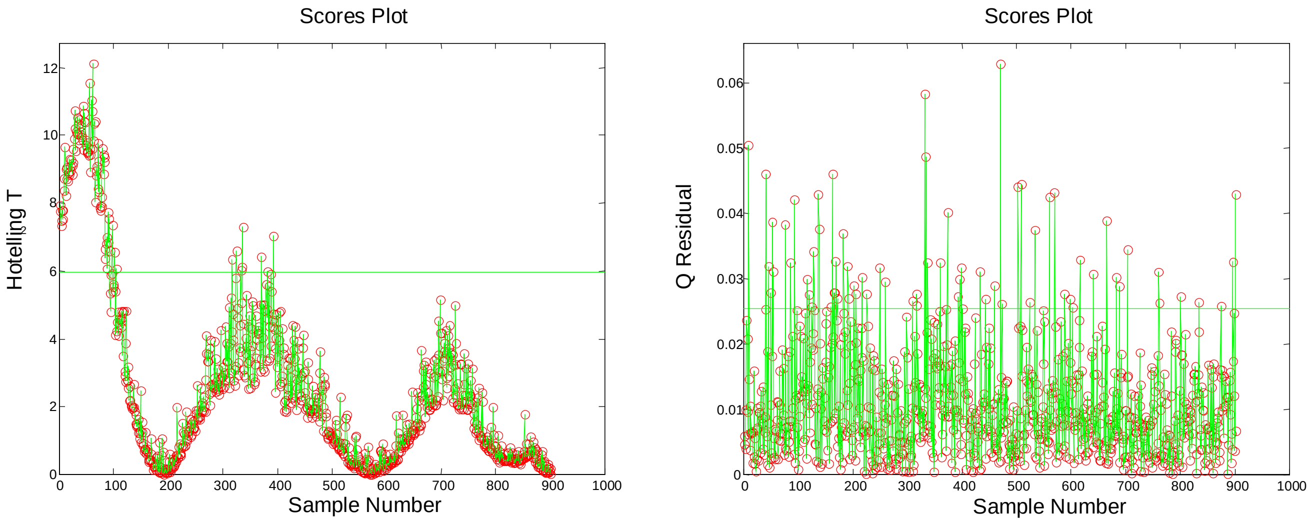

Figure 7 - T^2 Statistic on Validation Data (left) and Q Statistic on Validation Data (right)

Since the T^2 and Q statistics are within the confidence limits, the model represents the validation data set well. The PCA gives the best linear model in sense of minimum squared error.

Drifted Sensor

To verify the model ability to detect and identify the drifting sensors, an artificial drift (ramp) was applied separately to each input variable. The drift simulates a common problem that affects process sensors and may result from aging. The simulated drift is a ramp that grows to 0.05ºC (maximum value) for a temperature variable that originally varies from 28.87ºC to 30.22ºC. This small drift corresponds to a 0.17% change and is imperceptible in a time profile. Figure 8 shows the T^2 and Q statistics for this case.

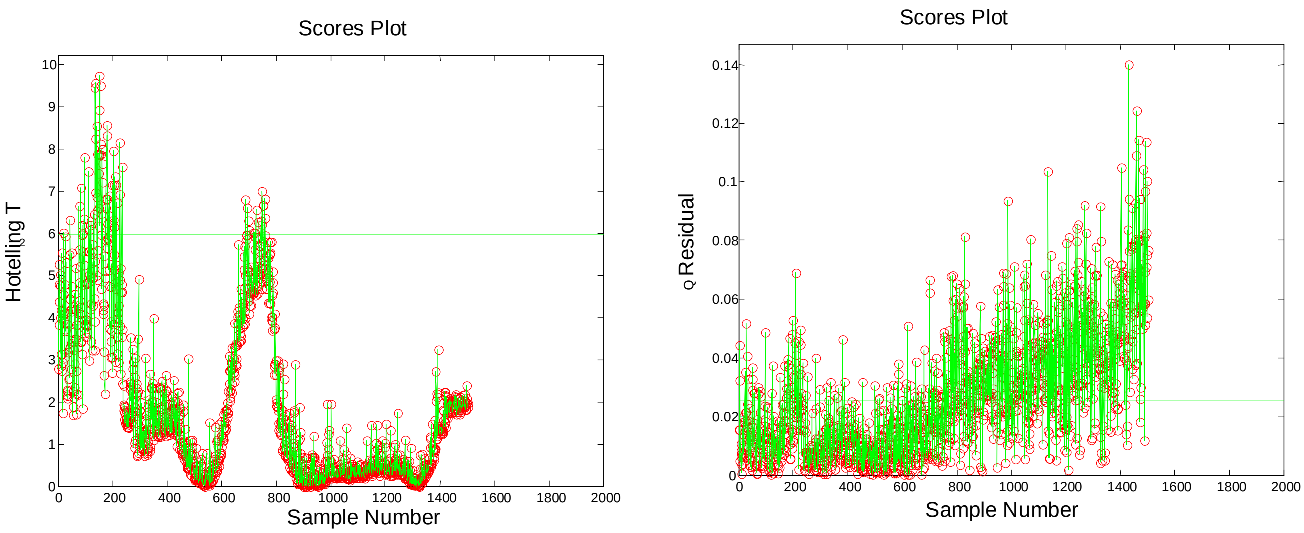

Figure 8 - T^2 Statistic on Drifted Data (left) and Q Statistic on Drifted Data (right)

The T^2 statistic doesn't seem to be out of the ordinary but the Q statistic plot shows that its values are increasing over time. This indicates that something is going on that is not in the original data. We can look at the contribution of each input to the large Q statistic. Through this analysis, it is possible to determine which variable is responsible for the unusual Q behavior. Samples 829, 1238 and 1430 were investigated. The contributions to the Q statistics are plotted in Figures 9 and 10.

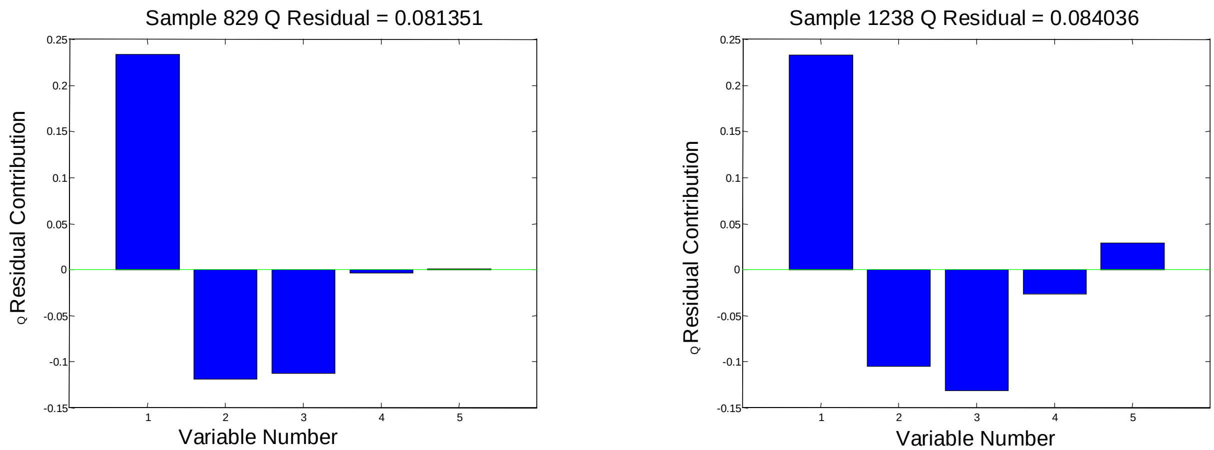

Figure 9 - Contributions (T1, T2, T3, T4, T6) to Q statistics of samples 829 (left) and 1238 (right)

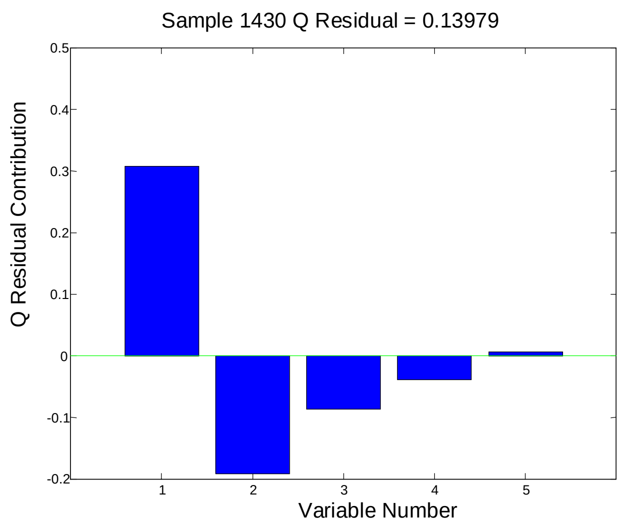

Figure 10 - Contributions (T1, T2, T3, T4, T6) to Q statistics of sample 1430

From Figures 9 and 10, it is easy to see that the variable T1 is the responsible for the unusually large Q statistic. This agrees with the fact that the drift was added to variable T1. Artificial drifts that were added to each of the other variables were detected and the drifted variable was identified using the Q statistic. When a ramp drift with 0.5ºC maximum value is added to the T1 variable, the deviation of the Q statistic is even more evident as shown in Figure 11.

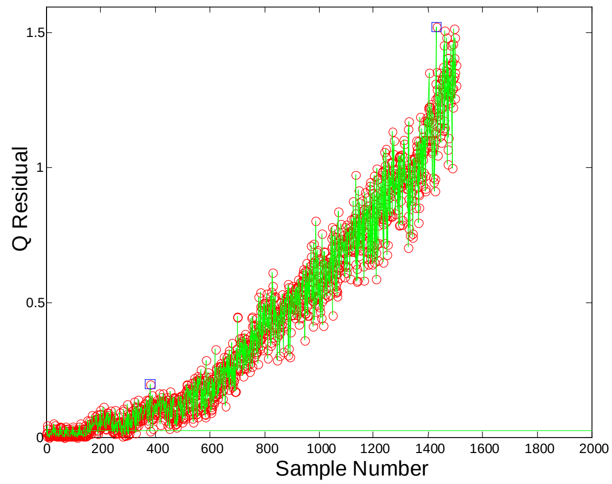

Figure 11 - Q Statistic on Drifted Data

The contributions of the variables to the Q residuals on the samples 379 and 1430 are shown in Figure 12.

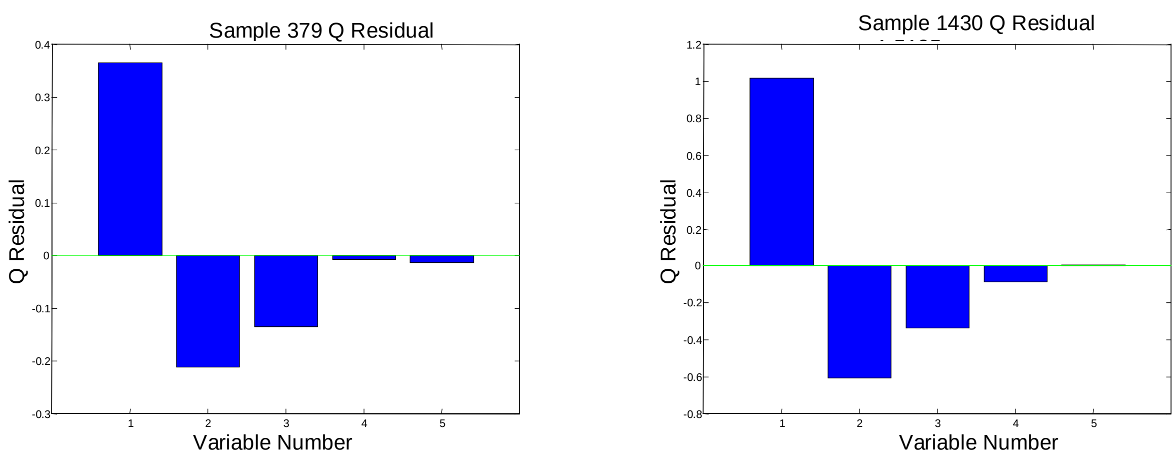

Figure 12 - Contributions of (T1, T2, T3, T4, and T6) to Q statistics of samples 379 and 1430

Again we observe the large contribution of sensor T1 to the Q statistic.

Conclusions

A PCA model with two principal components was developed to describe five temperature sensors in a nuclear research reactor. The model fitted the data well, as shown by the T^2 and Q statistics. The model was tested with a validation data set from a separate reactor fuel cycle and the model performed well. Artificial drifts were added to the variables and the model both detected and identifies the drifted variables. The PCA model was determined to be a good method to monitor the temperature sensors in this plant. This is due to the highly correlations in the data and to the insignificant non-linearities present.

If non-linearities or time delays were present in the system, other methods such as non-linear partial least squares or neural networks may be used.

Acknowledgments

The authors wish to thank the National Science Foundation (NSF) and the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) that gave support to this joint research project.

References

[1] Gallagher, N.B., B.M. Wise, S.W. Butler, D.D.White Jr. and G.G. Barna (1997), Development and Benchmarking of Multivariate of Statistical Process Control Tools for a Semiconductor Etch Process: Improving Robustness through Model Updating, ADCHEM 1997, Banff.

[2] Wise, B.M., N.B. Gallagher, S.W. Butler, D.D.White Jr. and G.G. Barna (1996), Development and Benchmarking of Multivariate of Statistical Process Control Tools for a Semiconductor Etch Process: Impact of Measurement Selection and Data Treatment on Sensitivity, Safeprocess '97, Hull, England August 26-27.

[3] Dunia R., Qin S. J. (1998), Joint Diagnosis of Process and Sensor Faults Using Principal Component Analysis. Control Engineering Practice 6 (1998) 457-469.

[4] Hines J.W., (2000), *PCA models*. NE 589 class notes, The University of Tennessee.

[5] Joliffe, L.T., (1986), *Principal Components Analysis*, Springer-Verlag.

[6] Valle S., Weihua Li and S.J. Qin. (1999), Selection of the Number of Principal Components: The Variance of the Reconstruction Error Criterion with a comparison to Other Methods, *Ind. Eng. Chem. Res*., 38, 4389-4401.

Appendix - Schematic Diagram of the IEA-R1 Pool Research Reactor