DEVELOPMENT AND BENCHMARKING OF MULTIVARIATE STATISTICAL PROCESS CONTROL TOOLS FOR A SEMICONDUCTOR ETCH PROCESS: IMPROVING ROBUSTNESS THROUGH MODEL UPDATING

Neal B. Gallagher and Barry M. Wise

Eigenvector Research, Inc.
830 Wapato Lake Road
Manson, WA 98831
USA
nealg@eigenvector.com

Stephanie Watts Butler, Daniel D. White, Jr. and Gabriel G. Barna

Texas Instruments
13536 North Central Expressway, MS 944
Dallas, TX 75265
USA

Abstract: Multivariate Statistical Process Control tools have been developed for monitoring and fault detection on a Lam 9600 Metal Etcher. Application of these methods is complicated because the process data exhibits large amounts of normal variation that is continuous on some time scales and discontinuous on others. Variations due to faults can be minor in comparison. Several models based on principal components analysis and variants which incorporate methods for model updating have been tested for long term robustness and sensitivity to known faults. Model performance was assessed with about six month's worth of process data and a set of benchmark fault detection problems.

Keywords: Fault detection, Multivariate quality control, Non-stationary, Principal Components Analysis, Adapting Models

1. INTRODUCTION

Semiconductor processes, like many chemical processes, are becoming increasingly measurement rich. Large volumes of data are recorded and are often not used until the process has undergone a significant upset. This data can be very useful for process monitoring if the appropriate tools are applied. Successful applications can result in reduced costs and/or improve the final product quality through improved process control or fault detection. However, there are significant obstacles to using the data for process monitoring and fault detection, including the sheer volume of the data, large numbers of variables, and the non-stationarity of the process data due to process and monitoring sensor drift. A wide variety of data treatment methods are available, however, it is often not apparent what methods will be useful in meeting monitoring and fault detection goals.

Several chemometrics techniques are available for application to process data (Wise, *et. al.*, 1996). These applications can be roughly divided between those directed at maintenance of process instruments, e.g. calibration, and those concerned with maintenance of the process itself, e.g. statistical process control and fault detection. The focus of this paper is on the latter application in which we describe a study performed on a Lam 9600 metal etch tool at Texas Instruments. For this study principal components analysis (PCA) modeling methods, which are commonly used for multivariate statistical process control (MSPC), were modified to be robust over long time periods in the presence of process drift while remaining sensitive to faults. PCA will be briefly reviewed, along with more recent modifications which allow the PCA models to adapt with the process. The issue of sensitivity of different process sensors and methods for detecting process faults is discussed in a companion article (Wise, *et. al.*, 1997).

2. THE METAL ETCH PROCESS

There are several steps in the manufacture of semiconductors. This project focused on an Al-stack etch process performed on the commercially available Lam 9600 plasma etch tool. The goal of this process is to etch the TiN/Al - 0.5% Cu/TiN/oxide stack with an inductively coupled BCl_3/Cl_2 plasma. The key parameters of interest are the etch Al line width reduction relative to the incoming resist line width, etch uniformity across the wafer, and loss of the underlying oxide due to over etch.

The standard recipe for the process consists of a series of six steps. The first two are for gas flow and pressure stabilization. Step 3 is a brief plasma ignition step. Step 4 is the main etch of the Al layer terminating at the Al endpoint, with Step 5 acting as the over-etch for the underlying TiN and oxide layers. Note that this is a single chemistry etch process, i.e. the process chemistry is identical during steps 3 through 5. Step 6 vents the chamber. The process "profile" is shown in Figure 1, which is the etch tool Endpoint A signal (the plasma emission intensity as measured by a filter spectrometer). The stabilization step is followed by the three etch regions: Al, TiN and oxide etch. Etching of an individual wafer is analogous to a single batch in a chemical process.

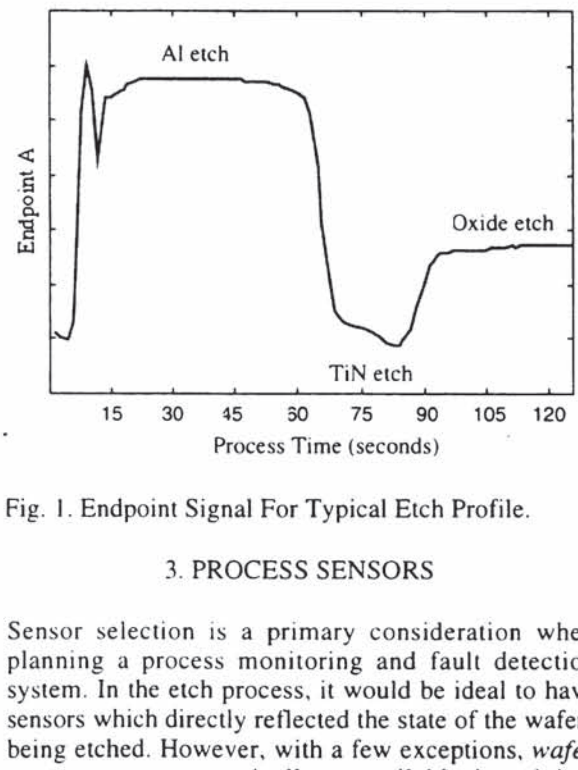

Fig. 1. Endpoint Signal For Typical Etch Profile.

3. PROCESS SENSORS

Sensor selection is a primary consideration when planning a process monitoring and fault detection system. In the etch process, it would be ideal to have sensors which directly reflected the state of the wafers being etched. However, with a few exceptions, *wafer* state sensors are typically unavailable in original equipment manufacturer etch tools. Thus, the alternative is to select more commonly available *process* state sensors, with the understanding that wafer state information will have to be inferred.

The research metal etcher used for this study was equipped with multiple sensor systems however, long term data was only available for the machine state sensors. Therefore this paper focusses on testing long term model robustness and sensitiviity using data from the machine state sensors only. These sensors, built into the processing tool, collect the available machine data during wafer processing. This data consists of 40 process setpoints, measured and controlled variables sampled at 1 second intervals during the etch. These are engineering variables, such as gas flow rates, chamber pressure and RF power. In this work, non-setpoint process variables that exhibit some normal variation were used for monitoring, as shown in Table 1. The physics of the problem also suggests that these variables should be relevant to process and final product state.

Table 1. Machine Variables for Process Monitoring

1	BCl_3 Flow	11	RF Power
2	Cl_2 Flow	12	RF Impedance
3	RF Bottom Power	13	TCP Tuner
4	RFB Reflected Power	14	TCP Phase Error
5	Endpoint A Detector	15	TCP Impedance
6	Helium Pressure	16	TCP Top Power
7	Chamber Pressure	17	TCP Reflected Power
8	RF Tuner	18	TCP Load
9	RF Load	19	Vat Valve
10	Phase Error		

4. PROCESS SHIFTS AND DRIFT

A major objective of this work was to determine if process monitoring models could be constructed using the machine state sensors. A key requirement for these models was that they remain robust over the long term i.e. did not require frequent recalibration due to excessive false alarms. These models must also remain sensitive to faults. Under ideal conditions a process would be stationary, i.e. retain the same mean and covariance structure over time. However, measurements from the etch process are clearly non-stationary. The etch process data exhibits large amounts of normal systematic variation on several time scales. This normal process drift is continuous on some time scales and discontinuous on others while variations due to faults can be relatively minor in comparison.

Normal process variation is primarily due to three sources. The first is a result of periodic cleaning and maintenance about every 1 to 2 months. After cleaning the goal is to restore the machine to its original state but this is rarely achieved, so consecutive clean cycles have different initial states i.e. large discontinuous shifts in the process mean. The second source of large variation is due to a continuous drift in the process data over a clean cycle as residue accumulates on the inside of the chamber and as the machine state sensors drift. The third major source of variation in the process data is a result of discontinuous shifts in the process mean on a lot–to–lot basis. Dozens of lots (there are approximately 24 wafers per lot processed one at a time) are processed during a clean cycle and variation in the lots is primarily due to differences in incoming wafers resulting from changes in upstream processing. In addition, process maintenance can result in sudden shifts in the process mean. Variation is also observed, although to a lessor extent, as a lot is processed. The result is that it is normal for the process data to show considerable variation over time, as illustrated in Figure 2. This variation is often much larger than changes due to process faults. It has also been observed that the process mean shows more erratic behavior than the process covariance, i.e. how the process variables co-vary.

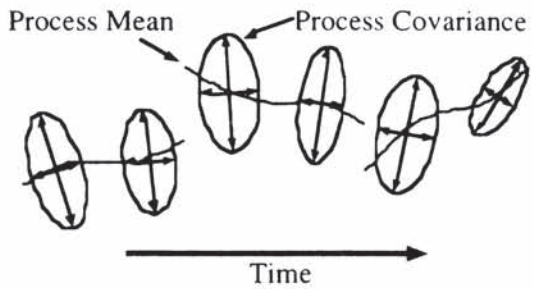

Fig. 2. Schematic Diagram Showing Drifting Process Covariance and Shifting Mean.

5. DATA TREATMENT

Processes in the semi-conductor industry are being monitored by a large number of sensors producing enormous volumes of data. Often essential information lies not in any individual process variable but in how the variables change with respect to one another, i.e. how they co-vary. However, many modeling techniques applied to data processing do not take advantage of this fact, and as a result, a great deal of data is wasted i.e. little useful information is obtained from it. The question is how to extract information and compress the data down to a few useful metrics. Also, in the presence of noise, it would be desirable to take advantage of some signal averaging between the redundant measurements. Principal components analysis (PCA) has many of these desired properties and is a common tool for multivariate statistical process control (MSPC) (Jackson, 1991; Jackson, *et. al.*, 1979; Kourti, *et. al.*, 1995; Kresta, *et. al.*, 1991; MacGregor, 1994; Wise, *et. al.*, 1988, 1990 and 1995a). This work used PCA and variants of PCA for developing MSPC tools for the semi-conductor etch process.

5.1 Principal Components Analysis

Only a brief overview of PCA is given here. For more detail the reader is referred to Jackson (1991), Wise, *et. al.* (1995b and 1996), and Wold, *et. al.* (1987). The goal of PCA is to split a data matrix $\mathbf{X}$ into two portions: one that describes the systematic variation (the process model) and the other that captures measurement noise (residual variance). Of course the split is never quite perfect but this typically does not present a problem for process monitoring.

For a data matrix $\mathbf{X}$ that is m rows by n columns (samples by variables) the PCA model is given by Equation 1.

$$\mathbf{X} = \mathbf{t}_1\mathbf{p}^T_1+\mathbf{t}_2\mathbf{p}^T_2+...+\mathbf{t}_k\mathbf{p}^T_k + \mathbf{E} = \mathbf{T}_k\ \mathbf{P}^T_k + \mathbf{E} \quad (1)$$

Here $\mathbf{X}$ has been mean centered by a 1 by n vector of means $\mathbf{a}$ i.e. adjusted to have a zero mean by subtracting the mean from each column. It may also have been autoscaled by a 1 by n vector of standard deviations $\mathbf{d}$ i.e. adjusted to zero mean and unit variance by dividing each column by its standard deviation. In PCA the data matrix $\mathbf{X}$ is decomposed into the sum of k outer products of vectors $\mathbf{t}_i$ and $\mathbf{p}_i$ (the process model) plus a residual matrix $\mathbf{E}$. The $\mathbf{p}_i$ (loadings vectors) are the eigenvectors of the covariance matrix of $\mathbf{X}$ (correlation matrix if autoscaled) defined in Equation 2. The loadings are an orthonormal set of vectors ($\mathbf{p}_i^T\mathbf{p}_j = 0$ for $i \neq j$, $\mathbf{p}_i^T\mathbf{p}_j = 1$ for $i = j$) that describe directions of systematic variation in $\mathbf{X}$, and the eigenvalues λ_i associated with each $\mathbf{p}_i$ are proportional to the variance captured by i^{th} loading.

$$\mathbf{C} = \mathrm{cov}(\mathbf{X}) = \frac{\mathbf{X}^T\mathbf{X}}{m-1} \quad (2)$$

The scores, $\mathbf{t}_i$, form an orthogonal set of vectors ($\mathbf{t}_i^T\mathbf{t}_j = 0$ for $i \neq j$) and are coordinates of the samples in the new coordinate system defined by the $\mathbf{p}_i$. The number of principal components (PCs) retained to model the systematic variation k must be less than or equal to the smaller dimension of $\mathbf{X}$, i.e. $k \leq \min\{m, n\}$. Generally it is found that the data can be adequately described using far fewer PCs than original variables. Note that for $\mathbf{X}$ and any $\mathbf{t}_i$, $\mathbf{p}_i$ pair

$$\mathbf{X}\mathbf{p}_i = \mathbf{t}_i \quad (3)$$

This provides the basis for projecting new data into a PCA model, calculating new scores, and comparing to existing control limits.

Two statistics that are commonly employed in MSPC are a *lack of model fit* statistic, Q, and a measure of the variation *within* the PCA model given by Hotelling's T^2 statistic. For each sample Q is the sum of squares of each row (sample) of $\mathbf{E}$, for example, for the i^{th} sample in $\mathbf{X}$, $\mathbf{x}_i$:

$$Q_i = \mathbf{e}_i\mathbf{e}_i^T = \mathbf{x}_i(\mathbf{I} - \mathbf{P}_k\mathbf{P}_k^T)\mathbf{x}_i^T \quad (4)$$

where $\mathbf{e}_i$ is the i^{th} row of $\mathbf{E}$. The columns of $\mathbf{P}_k$ are the first k loadings vectors retained in the PCA model and $\mathbf{I}$ is the identity matrix of appropriate size (n by n). The Q statistic indicates how well each sample conforms to the PCA model and is a measure of the amount of variation *not* captured by the model. T^2 is the sum of normalized squared scores defined as

$$T_i^2 = \mathbf{t}_i\lambda^{-1}\mathbf{t}_i^T = \mathbf{x}_i\mathbf{P}\lambda^{-1}\mathbf{P}^T\mathbf{x}_i^T \qquad (5)$$

in this case $\mathbf{t}_i$ refers to the i^{th} row of $\mathbf{T}_k$, the matrix of k scores vectors. The matrix λ^{-1} is a diagonal matrix containing the inverse eigenvalues associated with the k eigenvectors (principal components) retained in the model. Statistical limits can be developed for sample scores, Q and T^2, and individual residuals.

5.2 *Applying an Existing PCA Model:MSPC*

A PCA model is developed on a calibration data set and consists of a mean vector, standard deviation vector (or other scaling vector if applied), eigenvalues, loadings, and statistical limits on the scores, Q and T^2. The model can be used with new process data $\mathbf{X}_{new}$ to detect changes in the system. The first step is to scale the new process data to the mean $\mathbf{a}$ and standard deviation $\mathbf{d}$ of the calibration data set. New scores $\mathbf{t}_{i,new}$ can be obtained using Equation 2 with original loadings vectors, $\mathbf{p}_i$, and Q and T^2 for the new data can be obtained with Equations 3 and 4 by substituting $\mathbf{x}_{i,new}$ for $\mathbf{x}_i$ and $\mathbf{t}_{i,new}$ for $\mathbf{t}_i$. When one monitors these values as the process proceeds, the result is multivariate statistical process control (MSPC). For PCA based monitoring models of the etch process it was found that the Q and T^2 statistics were adequate for detecting system faults.

5.3 *Application of PCA with a Moving Mean*

PCA as described above works well with stationary processes. However, for cases where the mean drifts but the covariance structure does not change other strategies can be applied. For infrequent discontinuous shifts in the mean (such as that observed between clean cycles for the etch process) the strategy may be as simple as resetting the mean used to scale new process data to that observed at the beginning of a new cycle. This estimate of the process cycle mean can also be continuously updated as new data are acquired. For frequent shifts in the process mean data from each batch (each wafer processed) can be centered to the process mean for that batch. This latter strategy puts all the monitoring responsibility on the covariance structure alone and removes information related to shifts in the process mean. This could result in a significant loss of sensitivity of the monitoring model. Of course a model of the process means, either alone or in conjunction with a model of the batch covariance, can prove very useful.

For process data with a continuously drifting mean (such as that observed over a clean cycle for the etch process) new process data could be centered to the mean of the past J batches or moving window average. Selecting a good value for J is difficult but analysis suggests that the number of batches corresponding to the time scale of the process drift is reasonable. Another strategy borrowed from time series analysis uses an exponentially weighted moving average (EWMA) (Box, *et. al.*, 1994). Equation 5 shows how the moving average $\mathbf{a}'$ is updated:

$$\mathbf{a}'_{i+1} = \alpha\mathbf{a}_i + (1-\alpha)\mathbf{a}'_i \qquad (5)$$

where α ($0 \le \alpha \le 1$) is the weighting. Data from batch i+1 is then centered to the moving average $\mathbf{a}'_{i+1}$ which only depends on past values of $\mathbf{a}'$. The problem here is determining a good value for α. In all cases discussed above the centering strategy should be applied to the calibration data set as well as new data.

5.4 *Application of PCA with a Moving Covariance*

If process drift includes a varying covariance structure the PCA models can be allowed to adapt (Wold, 1994). A moving window PCA can be used when the process data provides a single sample per batch. This is the case when only the batch mean is used, data is pre-processed using speech recognition technology (White, *et. al.*, 1997), or multi-way PCA is used (Wise, *et. al.*, 1996 and 1997). Moving window PCA is analogous to centering to a moving window mean and just uses a PCA model of the past J batches. The problem again is identifying a good value for J.

When all the data from a batch is available an exponentially weighted moving covariance (EWMC) can be used. If $\mathbf{C}_i$ is the covariance for the i^{th} batch Equation 6 shows how the moving covariance $\mathbf{C}'$ is updated:

$$\mathbf{C}'_{i+1} = \beta\mathbf{C}_i + (1-\beta)\mathbf{C}'_i \qquad (6)$$

where β ($0 \le \beta \le 1$) is the weighting. Data from batch i+1 is then compared to a PCA model based on the moving covariance $\mathbf{C}'_{i+1}$ which only depends on past values of $\mathbf{C}'$. The problem again is determining a good value for β.

6. TESTS WITH LONG TERM DATA

A significant accomplishment of this project was the acquisition of long term data. Analysis of this data showed that testing models on just a few lots worth of data is highly unrealistic and will likely give optimistic results for methods not expected to remain robust over the long term. The long term data includes significant variation due to a variety of causes and provides a realistic test of each method considered.

Each monitoring model tested was calibrated on data acquired on the Lam 9600 metal etcher from 11/27/95–1/12/96 including data from about 250 wafers. During this period the etch tool was subjected to preventative maintenance (PM) and an additional cleaning (MC). Once calibrated the models were not recalibrated during testing. Test data for about 700 wafers was acquired from 1/22–4/26/96 which included three PMs, a MC, and two new equipment installations (EQ). The test data also included data from 5 experiments (EXP-29 through 33) that spanned a PM, a MC and an EQ. Figure 3 shows the calibration and test periods. Analysis and modeling of the long term data was performed using the PLS_Toolbox for use with MATLAB (Wise, *et. al.*, 1995b; Mathworks, 1992).

Sensitivity of the process monitoring models was tested using etch data from three experiments (EXP-29, 31 and 33) that included induced faults. A series of specific faults were intentionally induced by changing the TCP power, RF power, pressure, Cl_2 or BCl_3 flow rate, and He pressure. These three experiments consisted of a total of 129 wafers with 21 faults. To make the test more representative of an actual sensor failure, the analysis was done with "reset" values: values for the controlled variable which was intentionally moved off its setpoint was reset to have the same mean as its normal baseline value, i.e. the controlled variable which was changed was reset to look normal in the data file. Designed experiments (EXP-30 and 32) were also performed and since these experiments used setpoints far from normal operation they can be considered multi-variable faults. Sensitivity is discussed more fully in the companion article (Wise, *et. al.*, 1997).

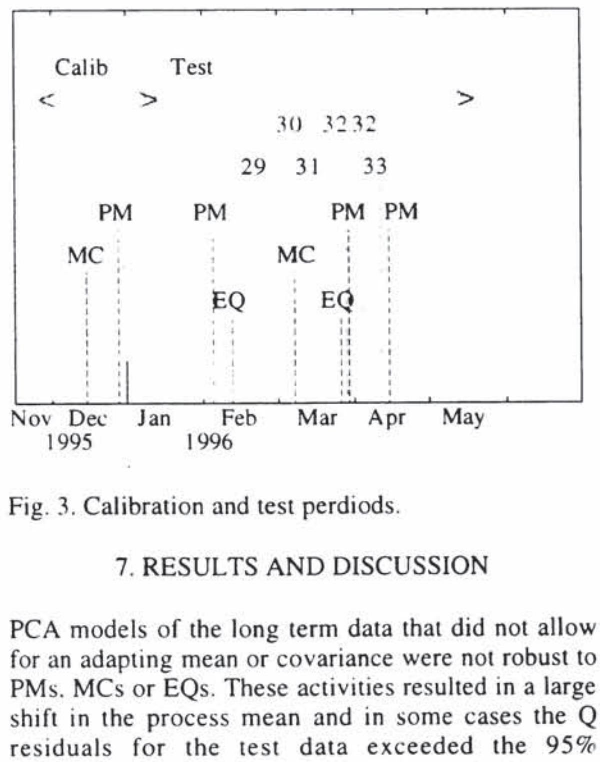

Fig. 3. Calibration and test perdiods.

7. RESULTS AND DISCUSSION

PCA models of the long term data that did not allow for an adapting mean or covariance were not robust to PMs, MCs or EQs. These activities resulted in a large shift in the process mean and in some cases the Q residuals for the test data exceeded the 95% confidence limit by 1–8 orders of magnitude. As a result all subsequent models reset the process mean after a PM, MC or EQ.

A model was also tested for wafer data centered to the process mean for that wafer. This centering strategy showed great improvement with Q residuals for the test data of the same order of magnitude as the 95% sample confidence limit. However, the model indicated faults for nearly all wafers after about 100 wafers were processed (after a PM and the first EQ). The result is that this model was too sensitive to changes in covariance structure and does not have a high longevity.

Three PCA models with an EWMA and weights α of 0.1, 0.5 and 0.9 were also tested. This strategy again showed a vast improvement in robustness compared to PCA with the assumption of a stationary process. However, all three models consistently showed false alarms after about 250 wafers were processed. The models remained robust through a PM, an EQ, EXP-29, and EXP-30, and subsequently failed after a MC. All three models caught the induced faults in EXP-29 but also included false alarms: 2 for $\alpha = 0.1$, and 4 each for $\alpha = 0.5$ and 0.9.

Three PCA models which combined an EWMA and EWMC were tested. These models used EWMA weights α of 0.1, 0.5 and 0.9 and an EWMC weight β of 0.01 which allowed the covariance to adapt slowly. These models remained robust through the entire test period and caught 8 (mostly power and pressuure faults) of the 21 induced faults (all 21 faults were identified when fault measurements were not replaced by their nominal value). As an example Figure 4 shows results for all three models in a portion of the test period. The Q residuals remain below the 95% limit line (horizontal dashed line) for nearly all wafers with normal set points, but indicates faults for the designed experiments (Exp-30 and 32), which had non-normal set points and should be indicated as a fault, and induced faults in Exp-31.

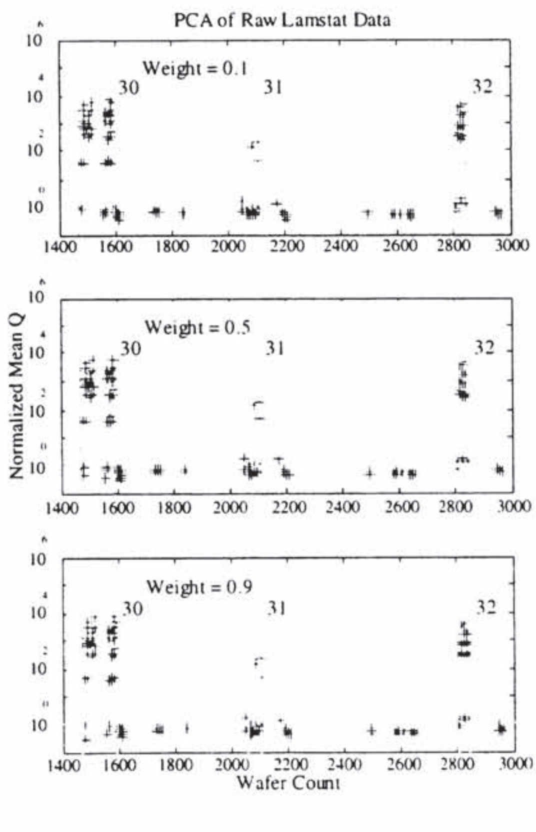

Figure 4. Q residuals normalized to the 95% confidence limit for test period 3/5–28/96 which includes Exp-30 to 32.

PCA models are typically very easy to interpret and when interogated can provide fault diagnosis. This study started with a simple PCA modeling strategy and added complexity only as it was deemed necessary. The most robust model tested used PCA combined with both an EWMA and EWMC. This strategy also proved sensitive, but at least three issues need to be addressed due to the added complexity. The first issue is the development of rules for when the moving mean is reset (in this study it was reset after each PM, MC and EQ). Secondly, strategies for identifying optimal weighting parameters for the EWMA and EWMC need to be developed. No attempt was made to optimize these parameters in this study but there did not appear to be a strong sensitivity in model performance with respect to the EWMA weight. A third issue that must be addressed is the development of rules for model updating. In this study the models were updated with data from a wafer if it was not considered an excessive fault i.e. its Q and T^2 statistics were less than 1.1 times their respective 95% limit.

8. CONCLUSIONS

The success of developing the monitoring strategy depended on the availability of long term data. The etch process is non-stationary and contains large amounts of normal variance. Testing monitoring models with only a few lots of data can lead to optimistic and erroneous conclusions about model performance over the long term.

This study showed how one can systematically step through options for developing a robust process monitoring model. In this case complexity was added only as deemed necessary. The most robust models tested used a principal components analysis based model combined with an exponentially weighted moving average and covariance.

Remaining issues include the development of rules for resetting a moving mean. This study reset the moving when when large shifts occurred as a result of maintenence, cleaning and equipment changes. Methods also need to be developed for selecting optimal weightings for the moving avarage and covariance. Rules for model updating also need to be developed to avoid updating a moving avarage and covariance with data from a faulty process.

9. ACKNOWLEDGMENTS

The authors would like to acknowlege SEMATECH for providing partial funding for this work through the J-88E project.

10. REFERENCES

Box, G.E.P., G.M. Jenkins and G.C. Reinsel, (1994). *Time Series Analysis*, Prentice Hall, Englewood Cliffs, NJ.

Jackson, J.E. and G.S. Mudholkar (1979). "Control Procedures for Residuals Associated With Principal Component Analysis", *Technometrics*, **21**(3), 341-9.

Jackson, J.E. (1980). "Principal Components and Factor Analysis: Part 1-Principal Components," *J. Qual. Tech.*, **12**(4), 201-13.

Jackson, J.E. (1991). *A User's Guide to Principal Components*, Wiley, New York, NY.

Kourti, T., P. Nomikos and J.F. MacGregor (1995). "Process analysis, monitoring and diagnosis, using multivariate projection methods," *Chemo. and Intell. Lab. Sys.*, **28**, 3-21.

Kresta, J.V., J.F. MacGregor and T.E. Marlin, (1991). "Multivariate Statistical Monitoring of Process Operating Performance," *Can. J. of Chem. Eng.* **69**, 35-47.

MacGregor, J.F. (1994). "Statistical Process Control of Multivariate Processes," *Adchem '94*. Kyoto, Japan.

Mathworks, Inc. (1992), *MATLAB, Version 4.2c*, Natick, MA.

White, D., G. Barna, S. Butler, B. Wise and N. Gallagher, (1997). "Methodology for Robust and Sensitive Fault Detection", 191st Meeting of the Electrochemical Society, May 4–9.

Wise, B.M., D.J. Veltkamp, B. Davis, N.L. Ricker and B.R. Kowalski (1988). "Principal Components Analysis for Monitoring the West Valley Liquid Fed Ceramic Melter," *Waste Management '88*, Tucson AZ.

Wise, B.M., N.L. Ricker, D.J. Veltkamp and B.R. Kowalski (1990). "A Theoretical Basis for the Use of Principal Components Models for Monitoring Multivariate Processes", *Process Control and Quality*, **1**(1), 41-51.

Wise, B.M. and B.R. Kowalski (1995a). "Process Chemometrics" in *Process Analytical Chemistry*, Blackie Acadamie Publishing, in press.

Wise, B.M. and N.B. Gallagher (1995b). *PLS_Toolbox for Use with MATLAB, Version 1.5*, Eigenvector Research, Inc., Manson, WA.

Wise, B.M. and N.B. Gallagher (1996), "The Process Chemometrics Approach to Chemical Process Fault Detection and Supervision," *J. of Process Control*, **6**(6), 329-348.

Wise, B.M., N.B. Gallagher, N.B., S.W. Butler, D. D. White Jr., G.G. Barna, (1997). "Development and Benchmarking of Multivariate Statistical Control Tools for a Semiconductor Etch Process: Impact of Measurement Selection and Data Treatment on Sensitivity", Safeprocess '97, Hull, England August 26–27.

Wold, S., K. Esbensen and P. Geladi (1987). "Principal Components Analysis," *Chemo. and Intell. Lab. Sys.*, **2**, 37-52.

Wold, S. (1994). "Exponentially Weighted Principal Components Analysis." *Chemo. and Intell. Lab. Sys.*, **23**, 149-61.