DEVELOPMENT AND BENCHMARKING OF MULTIVARIATE STATISTICAL PROCESS CONTROL TOOLS FOR A SEMICONDUCTOR ETCH PROCESS: IMPACT OF MEASUREMENT SELECTION AND DATA TREATMENT ON SENSITIVITY

Barry M. Wise and Neal B. Gallagher

Eigenvector Research, Inc.
830 Wapato Lake Road
Manson, WA 98831
USA
bmw@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 (MSPC) tools have been developed for monitoring a Lam 9600 TCP Metal Etcher at Texas Instruments. These tools are used to determine if the etch process is operating normally or if a system fault has occurred. Application of these methods is complicated because the etch process data exhibits a large amount of normal systematic variation. Variations due to faults of process concern can be relatively minor in comparison. The Lam 9600 used in this study is equipped with several sensor systems including engineering variables (*e.g.* pressure, gas flow rates and power), spatially resolved Optical Emission Spectroscopy (OES) of the plasma and a Radio Frequency Monitoring (RFM) system to monitor the power and phase relationships of the plasma generator. A variety of analysis methods and data preprocessing techniques have been tested for their sensitivity to specific system faults. These methods have been applied to data from each of the sensor systems separately and in combination. The performance of the methods on a set of benchmark fault detection problems will be presented and the strengths and weaknesses of the methods will be discussed, along with the relative advantages of each of the sensor systems.

Keywords: Fault detection, Measurement Selection, Multivariate quality control, Principal Components Analysis, Multi-way Principal Components Analysis

1. INTRODUCTION

Semiconductor processes, like many chemical processes, are becoming more measurement rich all the time. A wide variety of sensors and sensor systems are available. The goal of adding sensors, of course, is to reduce costs and/or improve the final product quality through improved process control or fault detection. Often, however, it is not apparent what sensors will be useful in meeting these goals. In order for sensors to be useful, they must be sensitive to variations in the process and be stable enough to provide information over extended time periods. In addition, the method used to treat the process data must be specified, as they can also impact sensitivity and robustness performance.

Recently, "chemometric techniques" have been applied to process (as opposed to analytical chemistry) problems. These applications can be roughly divided between those directed at maintenance of process instruments, *e.g.* calibration, and those that are concerned with maintenance of the process itself, *e.g.* statistical process control and fault detection. Our focus will be on the latter area. In this paper, we describe a study performed on a Lam 9600 metal etcher to determine which of three sensor systems, alone and in combination, and what data treatment method is the most sensitive to a series of known faults. One of the most often used chemometric techniques, Principal Components Analysis (PCA) will be reviewed, along with a more recent adaptation of the method, Multi-way PCA. The

issue of robustness of the sensors and methods over long periods will be discussed in a companion article.

2. THE METAL ETCH PROCESS

There are many steps in the manufacture of semiconductors. This project was 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 line width of the etched Al line (specifically the line width reduction in relation to the incoming resist line width), uniformity across the wafer and the oxide loss.

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" as indicated the by the Endpoint A signal (the plasma emission intensity from as measured by a filter spectrometer) is shown in Figure 1 of Gallagher *et. al*. (1997). The stabilization step is followed by the three etch regions: Al, TiN and oxide etch.

3. PROCESS SENSORS

Sensor selection is a primary consideration when planning a fault detection and classification (FDC) system. In the etch process, it would be ideal to have sensors which directly reflected the state of the wafers in the process. However, with a few exceptions, wafer state sensors are typically unavailable in original equipment manufacturer (OEM) processing 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 metal etcher used for this study was equipped with 3 sensor systems: machine state, Radio Frequency Monitors (RFM), and Optical Emission Spectroscopy (OES). The machine state sensors, built into the processing tool, collect the available machine data during wafer processing. The machine data consists of 40 process setpoints and 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 with some normal variation were used for monitoring, as shown in Table 1. Also, the physics of the problem suggest that these variables should be relevant to process and final product state.

The RFM sensors measure the voltage, current and phase relationships at the fundamental frequency of 13.56 MHz and the next four harmonics at four locations in the RF control system. The resulting 70 values are sampled every 3 seconds. The OES is used to monitor the plasma in the range of 245 to 800 nm in three locations above the wafer. The original data consists of 2042 channels per location, however, in this work the data was preprocessed by integrating a much smaller number of peaks (40) in each of the three spectra which correspond to process gases and species evolving from the wafer due to the etch.

Table 1. Machine State Variables Used 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		

A major objective of this work was to determine which sensors, or combinations of sensors, are most useful for detecting process faults. Data from the three sensors systems was used to develop models of the process in a variety of ways and the ability of the models to detect faults was tested.

4. PROCESS SHIFTS AND DRIFT

Ideally, under normal conditions, a process would be stationary, *i.e.* retain the same mean and covariance structure over time. Unfortunately, measurements from the etch process are clearly non-stationary. Changes in the data are primarily due to three sources, aging of the etcher over a clean cycle as residue accumulates on the inside of the chamber, differences in the incoming materials due to changes in upstream processes, and drift in the process monitoring sensors themselves. In addition, process maintenance can result in sudden shifts in the mean. The result is that it is normal for the process data to show considerable variation over time. This variation is often much larger than changes due to actual process faults. Process means, however, normally show more erratic behavior than the process covariance, *i.e.* how the process variables covary .

5. DATA TREATMENT

Chemical and manufacturing processes are becoming more heavily instrumented and the data is recorded more frequently. This is creating a data overload, and the result is that a good deal of the data is “wasted,” *i.e.* no useful information is obtained from it. The problem is one of both compression and extraction. Generally, there is a great deal of correlated or redundant information provided by process sensors. This information must be compressed in a manner that retains the essential information and is more easily displayed than each of the process variables individually. Also, 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 covary. In this case the information must be extracted from the data. Furthermore, in the presence of large amounts of noise, it would be desirable to take advantage of some sort of signal averaging.

5.1 Principal Components Analysis

Principal Components Analysis (PCA) is a favorite tool of chemometricians for data compression and information extraction (Jackson, 1991; Wise and Kowalski, 1995a; Wise *et. al.* 1996; Wold, *et. al.* 1987a). PCA finds combinations of variables or *factors* that describe major trends in a data set. Mathematically, PCA relies on an eigenvector decomposition of the covariance or correlation matrix of the process variables. In this work we will use the convention that rows of a data matrix **X** correspond to samples while columns correspond to variables. For a given data matrix **X** with *m* rows and *n* columns the covariance matrix of **X** is defined as

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

This assumes that the columns of **X** have been "mean centered," *i.e.* adjusted to have a zero mean by subtracting off the mean of each column. If the columns of **X** have been "autoscaled," *i.e.* adjusted to zero mean and unit variance by dividing each column by its standard deviation, Equation 1 gives the correlation matrix of **X**. (Unless otherwise noted, it is assumed that data is either mean centered or autoscaled prior to analysis.) PCA decomposes the data matrix **X** as the sum of the outer product of vectors $\mathbf{t}_i$ and $\mathbf{p}_i$ plus a residual matrix **E**:

$$\mathbf{X} = \mathbf{t}_1\mathbf{p}^T_1 + \mathbf{t}_2\mathbf{p}^T_2 + ... + \mathbf{t}_k\mathbf{p}^T_k + \mathbf{E} \qquad (2)$$

Here *k* must be less than or equal to the smaller dimension of **X**, *i.e.* $k \leq \min\{m,n\}$. The $\mathbf{t}_i$ vectors are known as *scores* and contain information on how the *samples* relate to each other. The $\mathbf{p}_i$ vectors are *eigenvectors* of the covariance matrix, *i.e.* for each $\mathbf{p}_i$

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

where λ_i is the *eigenvalue* associated with the eigenvector $\mathbf{p}_i$. In PCA the $\mathbf{p}_i$ are known as *loadings* and contain information on how *variables* relate to each other. The $\mathbf{t}_i$ form an orthogonal set ($\mathbf{t}_i^T\mathbf{t}_j = 0$ for $i \neq j$), while the $\mathbf{p}_i$ are orthonormal ($\mathbf{p}_i^T\mathbf{p}_j = 0$ for $i \neq j$, $\mathbf{p}_i^T\mathbf{p}_j = 1$ for $i = j$). Note that for **X** and any $\mathbf{t}_i$, $\mathbf{p}_i$ pair

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

This is because the score vector $\mathbf{t}_i$ is the linear combination of the original **X** data defined by $\mathbf{p}_i$. The $\mathbf{t}_i$, $\mathbf{p}_i$ pairs are arranged in descending ordered according to the associated λ_i. The λ_i are a measure of the amount of *variance* described by the $\mathbf{t}_i$, $\mathbf{p}_i$ pair. In this context, we can think of variance as *information*. Because the $\mathbf{t}_i$, $\mathbf{p}_i$ pairs are in descending order of λ_i, the first pair capture the largest amount of information of any pair in the decomposition. In fact, it can be shown that the $\mathbf{t}_1$, $\mathbf{p}_1$ pair capture the greatest amount of variation in the data that it is possible to capture with a linear factor. Subsequent pairs capture the greatest possible variance remaining at that step.

The concept of principal components is shown graphically in Figure 1. The figure shows a three dimensional data set where the data lie primarily in a plane, thus the data is well described by a two principal component (PC) model. The first eigenvector or PC aligns with the greatest variation in the data while the second PC aligns with the greatest amount of variation that is orthogonal to the first PC. Generally it is found that the data can be adequately described using far fewer principal components than original variables.

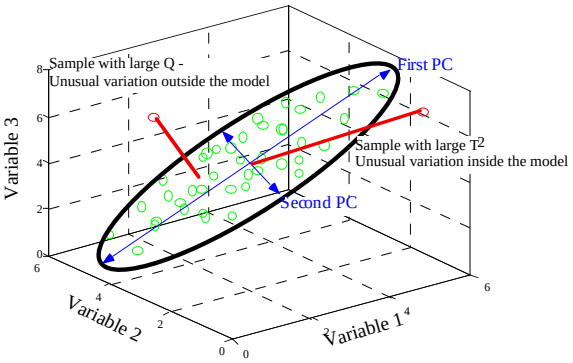


Fig. 1. Principal Component Model of Three Dimensional Data Set Lying Primarily in a Single Plane Showing Q and T^2 Outliers.

It is also possible to calculate a lack of model fit statistic, Q, for each sample. Q is simply the sum of squares of each row (sample) of **E** (from Equation 2), for example, for the i^{th} sample in **X**, $\mathbf{x}_i$:

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

where $\mathbf{e}_i$ is the i^{th} row of **E**, $\mathbf{P}_k$ is the matrix of the first *k* loadings vectors retained in the PCA model (where each vector is a column of $\mathbf{P}_k$) and **I** is the identity matrix of appropriate size (*n* by *n*). The Q statistic indicates how well each sample conforms to the PCA model. It is a measure of the amount of variation in each sample *not* captured by the *k* principal components retained in the model.

A measure of the variation *within* the PCA model is given by Hotelling's T^2 statistic. T^2 is the sum of normalized squared scores defined as

$$T_i^2 = \mathbf{t}_i\lambda^{-1}\mathbf{t}_i^T = \mathbf{x}_i\mathbf{P}\lambda^{-1}\mathbf{P}^T\mathbf{x}_i^T \qquad (6)$$

where $\mathbf{t}_i$ in this instance refers to the i^{th} row of $\mathbf{T}_k$, the matrix of *k* scores vectors from the PCA model. 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 Q and T^2, (along with limits on the scores and individual residuals).

5.2 Applying an Existing PCA Model: MSPC

Once a PCA model has been developed (including mean and variance scaling vectors, eigenvalues, loadings, statistical limits on the scores, Q and T^2) it can be used with new process data to detect changes in the system generating the data. The scores for new data $\mathbf{t}_{i,\,new}$, can be obtained for new data $\mathbf{X}_{new}$ with Equation 4 using the original loadings vectors, $\mathbf{p}_i$. In a similar fashion, new Q and T^2 can be obtained with Equations 5 and 6 by substituting $\mathbf{x}_{i,new}$ for $\mathbf{x}_i$. When one monitors these values as the process proceeds, the result is multivariate statistical process control (MSPC) (Wise and MacMakin, 1987; Wise *et. al.* 1988, 1990, 1991; Wise and Ricker, 1989; Kresta *et. al*. 1991).

In this work we will use primarily Q and T^2 for detecting system faults. Some discussion of the geometric interpretation of Q and T^2 is perhaps in order. As noted above, Q is a measure of the variation of the data outside of the PCA model. Refer again to our 3 variable where the data is restricted to lie on a plane shown in Figure 1. Such a system would be well described by a 2 PC model. Q is a measure of the distance off the plane formed by the first 2 PCs. In fact $\sqrt{Q}$ *is* the Euclidean distance of the operating point from the plane formed by the 2 PC model. A point with an unusually large Q value is depicted in Figure 1. The Q limit defines a distance off the plane that is considered unusual for normal operating conditions. T^2, on the other hand, is a measure of the distance from the multivariate mean to the projection of the operating point onto the 2 PCs. The T^2 limit defines an ellipse on the plane within which the operating point normally projects. Again, Figure 1 shows a point with a high T^2 value.

5.3 Multi-Way PCA

The PCA method outlined above take no explicit account of the ordered nature of a data set, *i.e.* the fact that the data was collected in a sequential manner. Reordering the samples in PCA would produce identical results. There are methods that explicitly consider that the data is ordered. These are referred to as multi-way methods because the data is usually organized into time ordered blocks that are each representative of a single sample or process run. The blocks are then arranged into multi-way matrices. Multi-way methods are particularly useful for the analysis of batch process data.

Consider the three dimensional data array shown in Figure 2. A data matrix of this type would be typical of a series of runs of a batch process such as our example of semiconductor processing where each "batch" is a wafer. Here there are $j = 1, 2, ..., J$ variables measured at times $k = 1, 2, ..., K$ throughout the batch. Similar data will exist on $i = 1, 2, ..., I$ runs of the batch process. The data can be summarized in the three dimensional ($I \times J \times K$) array $\underline{\mathbf{X}}$. Different batch runs (samples) are arranged along the vertical side, different process measurements (variables) along the horizontal side, and time recedes into the figure. Each horizontal slice through the array is a ($J \times K$) matrix representing the time history for all variables of a batch of a particular batch or sample. Each vertical slice made parallel to the front face of the cube is a ($I \times J$) matrix representing the values of all the variables in all the batches taken at a common time. A vertical slice made parallel to the side of the cube (the time axis) would represent a ($I \times K$) matrix of all the time histories of a single variable for all the batches.

There are several methods for decomposing the array $\underline{\mathbf{X}}$ (Geladi, 1989). These methods include the tri-linear decomposition (TLD) (Sanchez and Kowalski, 1990), parallel factor analysis (PARAFAC) (Smilde and Doornbos, 1991), and Tucker models (Smilde *et. al.*, 1994). In this work, we will consider one of the more straightforward approaches, that of multi-way PCA (MPCA) (Wold *et. al.*, 1987b). Each of the decomposition methods place different constraints on the resulting matrices and vectors.

MPCA is statistically and algorithmically consistent with PCA and has the same goals and benefits (Nomikos and MacGregor, 1994, 1995). In MPCA the array $\underline{\mathbf{X}}$ is decomposed as the summation of the product of score vectors ($\mathbf{t}$) and loading matrices ($\mathbf{P}$) plus a residual array $\underline{\mathbf{E}}$ that is minimized in a least squares sense.

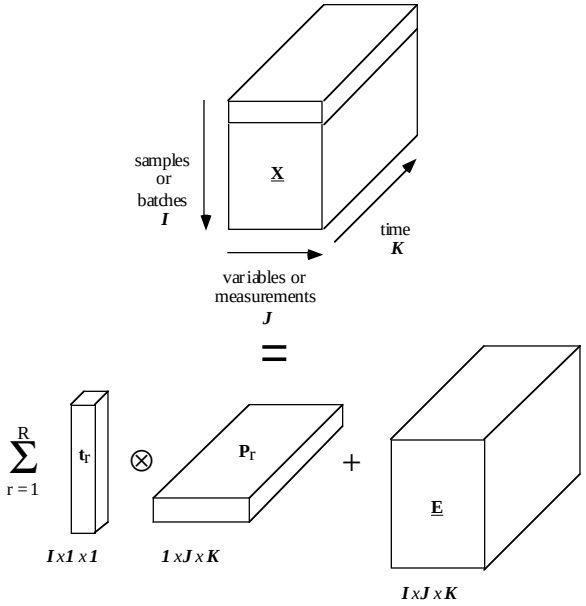


Fig. 2. Three Dimensional Data Array and Multi-way PCA Decomposition.

$$\mathbf{X} = \sum_{r=1}^{R} \mathbf{t}_r \otimes \mathbf{P}_r + \mathbf{E} \qquad (7)$$

This decomposition is shown graphically in Figure 4. This decomposition is done in accordance with the principles of PCA and separates the data in an optimal way into two parts. The noise or residual part $\underline{\mathbf{E}}$ is as small as possible and is associated with non-deterministic variation in the data. The systematic part, the sum of the $\mathbf{t}_r \otimes \mathbf{P}_r$, expresses the deterministic variation as one fraction (**t**) related only to batches and a second fraction (**P**) related to variables and their time variation.

MPCA is equivalent to performing PCA on a large two-dimensional matrix formed by unfolding the three-way array $\underline{\mathbf{X}}$ in one of six possible ways, only three of which are mathematically unique. For example, one might unfold $\underline{\mathbf{X}}$ in such a way as to put each of its vertical slices ($I \times J$) side by side to the right, starting with the slice corresponding to the first time interval. The resulting two dimensional matrix has dimensions ($I \times JK$). This particular unfolding allows one to analyze variability among the batches in $\underline{\mathbf{X}}$ by summarizing information in the data with respect to variables and their time variation. A mathematically equivalent unfolding would be to take slices off the side of $\underline{\mathbf{X}}$ and place them down the time axis, which also forms a matrix with dimensions ($I \times JK$). (The latter unfolding orders the matrix with the history of each variable kept together while the former orders the matrix with the all the measurements taken at the same time kept together.) One might also be interested in unfolding $\underline{\mathbf{X}}$ in other ways, however, the unfolding discussed above (and its mathematical equivalent) are the only ways that keep batch (sample) specific information separate from time and variable information.

The MPCA algorithm proceeds as shown in follows. First the matrix is unfolded in one of the two equivalent ways described above. Each column of the resulting matrix is then mean centered and, if appropriate, scaled to unit variance (autoscaled). An eigenvector decomposition as described in Equations 1 to 3 is then applied to the unfolded $\underline{\mathbf{X}}$. Each of the **p**, however, is really an unfolded version of the loadings matrix $\mathbf{P}_r$. After the **p** are obtained, the $\mathbf{P}_r$ can be obtained by reversing the unfolding procedure. In a similar manner, the three way array $\underline{\mathbf{E}}$ can be formed by folding the PCA residual matrix **E**. The Q and T^2 statistics can be calculated using the unfolded solution as shown in Equations 5 and 6.

This version of MPCA explains variation of measured variables about their average trajectories. Subtracting the average trajectory from each variable (accomplished by mean centering the columns of the unfolded matrix **X**) removes the major nonlinear behavior of the process. The *i*th elements of the **t**-score vectors correspond to the *i*th batch (sample) and summarize the overall variation in this batch with respect to the other batches in the database over the entire history of the batch. The **P** loading matrices summarize the time variation of the measured variables about their average trajectories. The elements of **P** are the weights, which when applied to each variable at each time interval within a batch, give the **t** scores for that batch. Additional examples of MPCA for MSPC can be found in Kosanovich *et. al.* 1994; Gallagher et. al. 1996, Wise and Gallagher, 1996).

5.4 Data Preprocessing

Before applying PCA or MPCA, several options are available for preprocessing the data. In PCA, one would often simply determine a single mean and variance for scaling the data and apply this scaling to all additional data. In our current example, however, it is known that process drift occurs, and that the process mean may shift. Thus, one might consider mean centering the data from each wafer in order to eliminate the effect of drift. It might also be possible to continually rebuild PCA models so that they are based only on recent data. (This approach is discussed at length in a companion paper).

An additional complication involves stretching of the time axis in the data record. In the etch process, timeline stretching causes blocks of data from each wafer to have different numbers of samples. This is due to differing lengths of the etch because of changes in layer thickness. One way to approach this is to simply average the data from each wafer over all available samples and work with only a mean. Another approach would be to select a specified number of samples where some point in the selected record corresponds to some particular process event. In related work, we have also used speech recognition methods such as Dynamic Time Warping to map the process response back onto a reference trace.

As will be seen in the following sections, the data pretreatment method can have a significant impact on the overall sensitivity and robustness of the method.

6. INDUCED FAULT EXPERIMENTS

A series of three experiments (EXP-29, 31 and 33) were performed where faults were intentionally induced by changing the TCP power, RF power, pressure, Cl_2 or BCl_3 flow rate, and He chuck 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. For example, if the induced fault was a change of the TCP power from 350 to 400 watts, the data file value of the TCP power was reset from a mean of 400 back to 350 by addition of a constant bias. The resulting data looks as if the

controller was seeing a biased sensor for TCP power and adjusting accordingly: TCP power would appear normal, but it would not be. The effect of a TCP power offset, however, should be evident (we hope) in the remaining process variables because the apparent relationship between the TCP power and the remaining variables should be different.

The three induced fault experiments were run at widely spaced intervals (in February, March and April 1996, respectively). Process drift is apparent in the data: each experiment has a significantly different multivariate mean. This is evident in Figure 3, which shows the scores on the first two PCs of the machine state data for all three experiments. The data clearly splits into three groups, one for each of the experiments. This suggests that models based on all of the data will define a much larger region of the multivariate space as normal variation than would a model of a single experiment. We will refer to a model of all of the data as a global model, and a model of each of the lots as a local model.

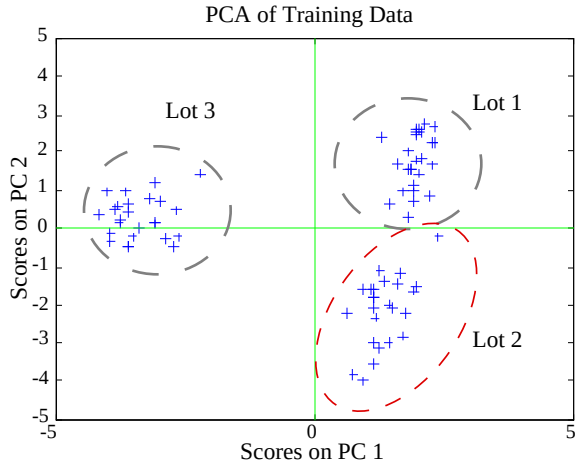


Fig. 3. Scores on First Two PCs from Analysis of Experiment 29, 31 and 33 Induced Fault Data.

7. RESULTS

Data from experiments 29, 31 and 33 was used to test the sensitivity of PCA and MPCA for detecting the induced faults. Machine state, RFM and OES data was available for each of these experiments. As described above, these experiments included some wafers where the setpoints for some variables were offset from the normal recipe. Prior to the analysis, the data from the sensors that measure each of these parameters (and is used for feedback control) were "reset" to their means from previous runs. All subsequent analysis was performed using the PLS_Toolbox software (Wise and Gallagher, 1995)

Several different approaches were used in the development of the fault detection models used in this test. In order to get a direct comparison of the sensitivity of the process sensors, only data from these experiments was used in model development (very little additional data exists where all 3 sensor systems are available). Models were developed that were intended to mimic the local and global behavior of the process. Local models were built using only data from a particular experiment, *i.e*. a model was built using data from the normal wafers from an experiment and was used to test the remaining wafers. The local models were intended to represent the upper limit of what might be achievable with models that update themselves continuously and thus are always local. Global models were developed using the normal wafers from all of the experiments simultaneously and then tested on the fault wafers. This represent the case where models span a large amount of process variation, *i.e*. include lot-to-lot and over a maintenance cycle type effects. These models included a larger amount of variation as normal than the local models.

The data was also preprocessed in a number of different ways prior to analysis. For some tests, the data from each wafer was reduced to a single vector of means of the variables over the entire wafer. In other cases, raw data was used for model development. Analysis was also performed using raw data where the data from each of the wafers was centered to its own mean. Multi-way analysis was also performed. In these instances each sample in the analysis includes the time history of the process sensors. As described above, the same number of samples were used for each wafer during model development and testing. For machine state data, 70 samples were used, including the last 25 data points from step 4 and the first 45 data points for step 5. Similarly, 25 and 28 data samples were used from the RFM and OES, respectively. RFM and OES variables that mirrored the process end point trace were found, and a consistent number of samples were selected on either side of the peak of the TiN etch.

The sensitivity results for the machine state, RFM and OES sensors used individually are shown in Table 2. The results for the sensors in combination are shown in Table 3. The faults are listed down the side of each table. Note that only faults where all data was available are considered in the table; thus, there are 19 faults listed rather than the original 21. The results for straight PCA models are shown on the left, for 5 different data pretreatment approaches. MPCA model results are shown on the right for 3 different data pretreatment approaches. Six different combinations of sensors are considered for each method/preprocessing combination: machine state, RFM, OES (Table 2), machine state + RFM + OES, machine state + RFM and machine state + OES (Table 3). A symbol in the body of the table indicates that the particular combination of data analysis method, pretreatment and sensors caught the particular fault. An open symbol indicates that the fault exceeded the 99% confidence limit, while a filled in symbol indicates that the fault exceeded the 99% limit by a factor of 5 or more. Note, however, that for analysis of the raw data, an open symbol indicates that more than 15% of the samples exceeded the 95% confidence limit, while a filled symbol indicates that over 30% of the samples exceeded the 95% confidence limit. Also, sensors were not

considered in combination using the raw data since the data acquisition times are not synchronized between the sensors.

Table 2. Results of Sensitivity Tests for Single Sensor Systems.

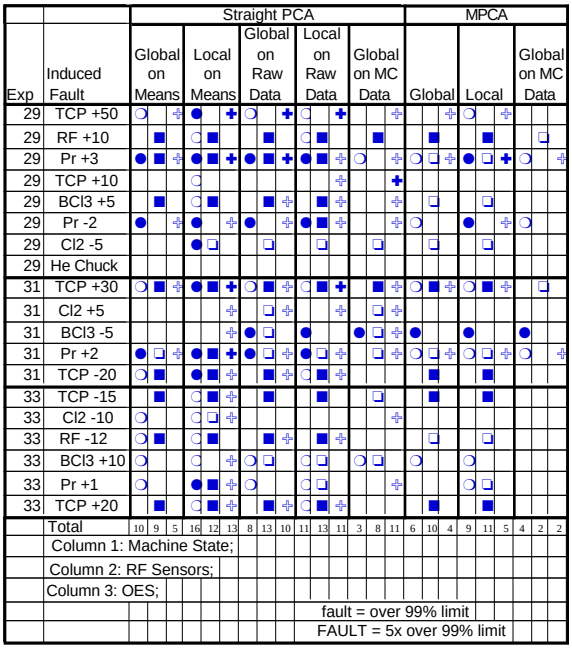
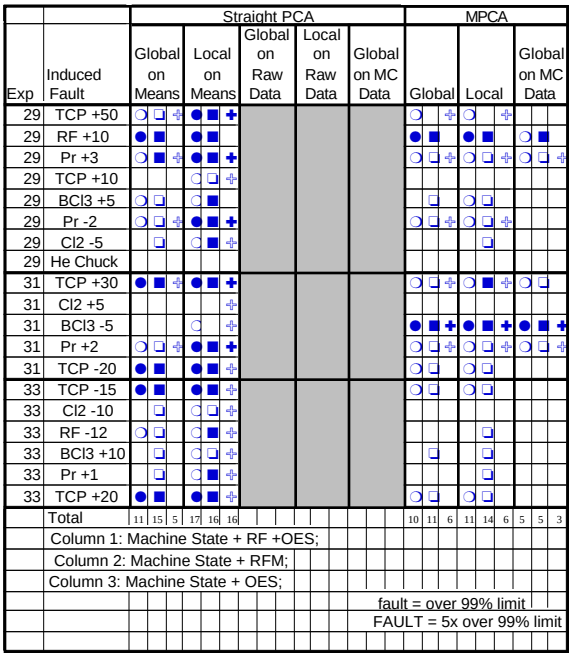


Table 3. Results of Sensitivity Tests for Combinations of Sensor Systems.



7. DISCUSSION

Several trends are evident upon examination of the results. It is clear that local models out-perform global models. In Tables 2 and 3 all local models performed better than all similarly configured global models. This is expected because global models include a larger amount of variation as normal. Thus, faults are smaller relative to the normal variation included in global models and are therefore more difficult to detect. PCA on the wafer means is somewhat more sensitive than PCA on raw data for machine state data but not for RFM and OES. The increased sensitivity with machine data, which tends to have a larger proportion of unmodeled variance, is probably signal averaging, *i.e*. it is easier to see a shift in the mean when signals are averaged over many samples. With OES and RFM data there is generally very little unmodeled variance in the raw data, and changes are more easily detected in the raw data.

In this analysis, MPCA does not perform better than PCA of the raw data. However, in previous analyses, with different arrangements of the data, MPCA did perform better on the machine state and OES sensors, but not the RFM. It is expected that MPCA will be more sensitive to some types of faults than PCA because the time-ordered nature of the data is considered explicitly. Faults which change the shape of the process trajectory, but not the overall mean and covariance, would be detectable with MPCA but not with PCA. Changes in shape can include stretching due to lengthening or shortening of some periods of the etch. In previous analysis, data was arranged for MPCA by including a specified number of samples starting from the beginning of the run, as opposed to including data centered on a particular feature near the middle of the run. MPCA models are more sensitive to stretching when the starting points are fixed in the data record, rather than a point near the center. It is not clear why MPCA does not lead to increased sensitivity when used with the RFM data. It may be that the shape of the RFM trajectories are inherently more variable, making changes to them harder to detect.

Methods based on wafers centered to their own mean are less sensitive that those based on raw data, as might be expected. However, there are several instances where the analysis on mean centered wafers detects faults where analysis of the means does not. This suggests that these techniques could be used simultaneously.

The overall performance of the different sensors is approximately equal, however, the OES sensor appears to degrade the most as the models are changed from local to global. This is no doubt due to the vary large amount of drift in the OES signals over the course of a clean cycle due to residue buildup on the chamber window. The sensitivity of RFM models, on the other hand, changed little when they were local or global. This suggest that the RFM sensors are the most stable, and/or least sensitive to lot-to-lot changes that do not affect processing.

The best combination of sensors is machine state plus RFM. This combination is more successful for detecting faults than any of the sensors alone or in combination when global models are considered. When OES data is combined with machine state and RFM data, the ability to detect faults generally decreases in the global models. This is not necessarily

true for local models where long-term variation in the OES is not important.

8. CONCLUSIONS

This study has shown how one can systematically step through the options for sensor systems and data treatment for fault detection systems in order to select the best measurements and analysis method for the particular job. For this particular application, simpler methods, such as PCA on the means, tended to work best. The major unresolved issue in this article concerns dealing with process and sensor drift. It is apparent that this had a major impact in this study. This issue is the subject of our companion article.

9. REFERENCES

Gallagher, N.B., B.M. Wise and C.W. Stewart (1996). "Application of Multi-way Principal Components Analysis to Nuclear Waste Storage Tank Monitoring," *Computers in Chem. Engng.*, Vol. 20, suppl., pps s739-44.

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

Geladi, P. (1989). "Analysis of multi-way (multi-mode) data," *Chemo. and Intell. Lab. Sys.*, Vol. 7, pps. 11-30.

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

Kosanovich, K.A., M.J. Piovoso, K.S. Dahl, J.F. MacGregor and P. Nomikos (1994). "Multi-Way PCA Applied to an Industrial Batch Process," 1994 American Control Conference.

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

Nomikos, P. and J.F. MacGregor (1994). "Monitoring Batch Processes Using Multiway Princiapl Component Analysis," *AIChE Journal*, Vol. 40, No. 8, pps. 1361-1375.

Nomikos, P. and J.F. MacGregor (1995). "Multivariate SPC charts for monitoring batch processes," *Technometrics*, Vol 37, No. 1.

Sanchez, E. and B.R. Kowalski (1990). "Tensorial Resolution: A Direct Trilinear Decomposition," *J. Chemo.*, Vol. 4, pps. 29-45.

Smilde, A.K. and D.A. Doornbos (1991). "Three way methods for the calibration of chromatographic systems: comparing PARAFAC and Three-way PLS," *J. Chemometrics*, Vol. 5, pps. 345-360.

Smilde, A.K., Y. Wang and B.R. Kowalski (1994). "Theory of Medium-Rank Second-Order Calibration with Restricted-Tucker Models," *J. Chemometrics*, Vol. 8, pps. 21-36.

Wise, B.M. and A.H. McMakin (1987). "A Statistical Technique for Analyzing Data from Liquid-Fed Ceramic Melters," PNL-SA-15267, Pacific Northwest Laboratory, Richland, WA.

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 Proceedings*, Tucson AZ 1988.

Wise, B.M. and N.L. Ricker (1989), "Feedback Strategies in Multiple Sensor Systems," *AIChE Symposium Series on Process Sensing*, No. 267 Vol. 85.

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", *Proc.Cont. and Quality*, Vol. 1, Number 1, pages 41-51.

Wise, B.M., D.J. Veltkamp, N.L. Ricker, B.R. Kowalski, S.M. Barnes and V. Arakali (1991). "Application of Multivariate Statistical Process Control (MSPC) to the West Valley Slurry-Fed Ceramic Melter Process," *Waste Management '91 Proceedings*, Tucson AZ 1991.

Wise, B.M. and B.R. Kowalski (1995). "Process Chemometrics"in *Process Analytical Chemistry*, pps 259-312, Blackie Acadamic & Professional, London.

Wise, B.M. and N.B. Gallagher (1995). *PLS_Toolbox for Use with MATLAB, Version 1.5*.

Wise, B.M. and N. B. Gallagher (1996a), "The Process Chemometrics Approach to Chemical Process Fault Detection and Supervision," *J. Proc. Control*, Vol. 6, No. 6, pps 329-348, 1996.

Wise, B.M. and N.B. Gallagher (1996b), "Multi-way Analysis in Process Monitoring and Modeling," *Computers in Chemical Engineering*, in press.

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

Wold, S., P. Geladi, K. Esbensen and J. Ohman (1987b). "Multi-way principal components and PLS analysis," *J. Chemo.* Vol. 1, pps. 41-56.