# FAULT DECTION AND VARIATION SOURCE IDENTIFICATION BASED ON STATISTICAL MULTIVARIATE ANALYSIS

**Ming-Da Ma*. Chun-Cheng Chang**. Shi-Shang Jang**. David Shan-Hill Wong****
**Sheng-Tsaing Tseng*****

**Center for Control and Guidance Technology, Harbin Institute of Technology*
*Harbin, China, (e-mail: mamingda@hit.edu.cn).*
*** Department of Chemical Engineering, National Tsing-Hua University, Hsin-Chu, Taiwan*
*(e-mail: ssjang@mx.nthu.edu.tw; dshwong@che.nthu.edu.tw )*
**** Institute of Statistics, National Tsing-Hua University, Hsin-Chu, Taiwan,*
*(e-mail: sttseng@stat.nthu.edu.tw)*

**Abstract:** This paper aims to solve the problems of fault diagnosis and variation reduction by using multivariate statistical techniques when the quality measurements are scarce. Both single stage process and multi-stage process are considered. For the single stage process, the nonparametric statistical method, Wilcoxon rank-sum test is used to identify the key variable/step that causes the fault of the un-qualified wafers. For the multi-stage process, the most important variables are first picked out by systematic statistical analysis, and the specifications of these key variables are designated using nonparametric method to improve the product yield. Gene map which gives visual images is used to assist the analysis. Industrial examples are given to show the effectiveness of the proposed method.

*Keywords:* semiconductor manufacturing, fault detection, Wilcoxon rank-sum test, cluster analysis, stepwise regression.

## 1. INTRODUCTION

State-of-the-art semiconductor processes are often pushed to the limits of current technologies, resulting in processes that have little or no margin for error. Advanced process control (APC) and fault detection and classification (FDC) are widely applied in semiconductor industries to reduce cycle-time and improve yield. The focus of this paper is on the fault detection algorithms to find out the variation source and root-cause of scrap wafers by using statistical multivariate analysis techniques.

Detection of process and tool faults in the shortest possible time is critical for minimizing scrap wafers and improving product yields for semiconductor manufacturing. However, most of wafer-states lack in situ sensor to provide real time information and usually are measured offline and less frequently than every wafer, which can lead to a number of scrapped wafers before a fault is detected. In the meanwhile, fortunately, more and more real time measurements of manufacturing equipments like temperature, pressure, power and flow rate, etc., are available due to the advances in metrology technology. These real time measurements provide valuable information about the tool status and can be used to predict final wafer characteristics. Further, it also provides a way to improve product quality by detecting and identifying equipment malfunctions in real time without interrupting the normal operations. The difficulty is, with such an abundant amount of data available, it is usually not clear which tool-state variable is critical or closed related with the final product quality.

Principal component analysis (PCA) and partial least squares (PLS) have drawn increasing interest and have been studied extensively in semiconductor manufacturing industry. PCA and PLS are useful tools for data compression and information extraction and have the advantages of dealing with high dimension and collinearities. PCA/PLS methods find linear combinations of variables that describe major trends in a data set. Considering the batch nature of semiconductor manufacturing, multi-way PCA is usually used to unfold three dimensions data into 2-D data array (Macgregor, 1994). Yue et al. applied multi-way PCA method to optical emission spectra for plasma etchers.

Most of the methods mentioned above require a large amount of training data to build a reliable statistical model to capture the key characteristics of the process. However, in real practice, many fabs are operating with diversified products of small account (Ma, et al., 2008) which means that one has to find out the causes of un-qualified wafers with limited quality data. Compared with principal components which are combined by all process variables, engineers are more eager to know which variable exactly, or linear combinations of several variables, plays an important role on the product quality. It is also of interest to know which step is critical to the whole streamline.

In this paper, a systematic approach is proposed for fault diagnosis and variation reduction by using statistical multivariate techniques. Both single stage process and multi-

stage process are considered. For the single stage process, the nonparametric statistical method, Wilcoxon rank-sum test is used to identify the key variable/step that causes the fault of the un-qualified wafers. For the multi-stage process, homogeneous process variables are first grouped by using cluster analysis, and representative variables or linear combinations of variables of each group are picked out. Then the key clusters are selected by stepwise regression method. Further, the upper and lower limits of these selected representative variables are designated to reduce product variation. It is shown that the proposed method improves the product yield substantially.

Recently, combinatory and high throughput experiments have received widespread attention in biology. Synopsis of large amount of experiment data and subsequent information mining from such data has become a special branch of study known as bioinformatics (Baldi and Brunak, 2001). The key experimental technique that is responsible for the advancement of bioinformatics is the microarray which enables expressions of tens of thousands of genes be measured and represented on a small array of colored image dots. In this paper, we demonstrate that quick diagnosis of the key variable/step that causes the fault in final quality can be achieved by simple statistical analysis of measured values of different sensors and graphical synopsis of results of such analysis. Furthermore, specifications for the key variables, which are usually far from optimal in original settings, can be designated to improve the product yield.

## 2. FAULT DETECTION FOR SINGLE STAGE PROCESS

### *2.1 Problem statement*

Consider quality data of n wafers are collected from a tool, $n_1$ wafers are qualified, and $n_2$ wafers are un-qualified, hence $n_1+n_2=n$. Let's denote that $m$ steps with $v$ variables are implemented during the whole process. It is assumed that in each step $t_s$ seconds are carried out for some certain objective (for instance temperature ramped up, current ramped down,…,etc.), where $s$=1,…,$m$. Suppose that the total time for all steps is $t$, then $t_1 + t_2 + \cdots + t_m = t$; let $T_r = t_1 + \cdots + t_r$, where $r$=1,…,$m$. Now, let's define $X_{i,j,k,l}$ to be the kth independent variable at batch time $l$ of jth wafer, where $j$=1,…,$n_i$, $k$=1,…,$v$, $l$=1,…,$t$, and $i$=1 means the wafer is qualified, $i$=2 indicates the wafer is not qualified. Now, the problem is what is the p-value of $X_{i,j,k,l}$ to distinguish the wafer is qualified or unqualified in case $n_1$ and $n_2$ are small.

### *2.2 Statistical analysis*

It is general to apply t-test to distinguish two set of data whether or not their mean is equal to each other. However, in this case $n_1$ and $n_2$ are small, a two sample t-test is not appropriate since the above two set of data may not be in normal distribution. Therefore, a nonparametric analysis, Wilcoxon rank-sum test, is used here.

The Wilcoxon rank-sum test is a nonparametric alternative to the two-sample t-test which is based solely on the order in which the observations from the two samples fall (Higgins, 2004). It is valid for data from any distribution, whether normal or not, and is much less sensitive to outliers than the two-sample t-test. The Wilcoxon test is based upon ranking the $n_1+n_2$ observations of the combined sample. Each observation has a rank: the smallest has rank 1, the 2nd smallest rank 2, and so on. The Wilcoxon rank-sum test statistic is the sum of the ranks for observations from one of the samples.

In this work, we implement Wilcoxon rank-sum test to find the p-value of the hypothesis of

$$H_0 : \mu_{1,k,l} = \mu_{2,k,l} \quad vs. \quad H_a : \mu_{1,k,l} \neq \mu_{2,k,l} \qquad (1)$$

where $\mu_{1,k,l}$ and $\mu_{2,k,l}$ are the mean of qualified and un-qualified wafers of the kth variable at time l respectively. It is assumed that there is not much prior knowledge of the product and no evidence shows that $\mu_{1,k,l}$ is greater or smaller than $\mu_{2,k,l}$. Therefore, a two-side test is implemented here.

Let $p_{k,l}$ be the above p-value of the kth variable at time $l$, three different approaches to evaluate the above approach can be implemented

(i) Evaluate the p-value of a process variable by finding the average p-value of the process variable in the whole time horizon:

$$P_k = \sum_{l=1}^{t} p_{k,l} \Big/ t \qquad (k=1,2,\ldots,v) \qquad (2)$$

(ii) Evaluate the average of p-value of each variable at each step

$$P_{k,b} = \sum_{l=T_{b-1}+1}^{T_b} P_{k,l} \Big/ t_b \qquad (k=1,2,\ldots,v\ ;b=1,2,\ldots.m,) \qquad (3)$$

(iii) Direct observe the p-value $p_{k,l}$ of each variable at each different time.

From the statistical analysis of the above approaches, we can determine which process variable, which step, plays an important role on the quality of the wafers, and more specifically, which second is critical for the final product quality. All these information is valuable to the engineers for their further improvement of the product quality.

### *2.3 Illustrative example*

The proposed algorithm is applied to a high-density plasma chemical vapor deposition (HDP-CVD) process. HDP-CVD, which is used as the gap-filling process for the dielectric in semiconductor circuits, features a high gap-fill capability compared with conventional plasma CVD by the excitation

of a high-density plasma. The schematic diagram of the HDP-CVD reactor is shown in Fig. 1.

There are 33 process variables for this manufacturing process. 9 steps are implemented for this process and the processing time is shown in Table 1. The quality data obtained from WAT test of 25 wafers are collected, among which 21 wafers are qualified and 4 wafers are un-qualified.

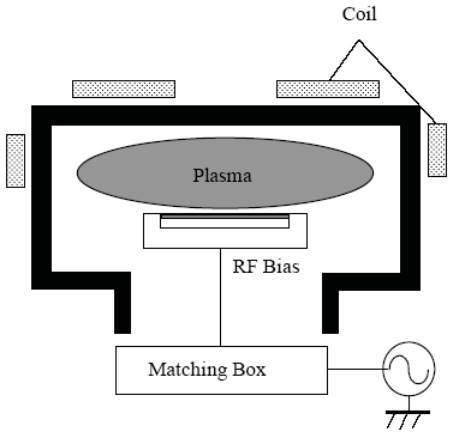


Fig. 1. Schematic diagram of the HDP-CVD reactor.

**Table 1 Processing time of each step**

| Step | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | Total |
|---|---|---|---|---|---|---|---|---|---|---|
| Seconds | 5 | 30 | 53 | 3 | 67 | 5 | 10 | 15 | 5 | 193 |

The proposed statistical method is applied to this process. The p-value of hypothesis (1) is calculated for process variables. To find out the key process variable that plays an important role on the process, equation (2) is implemented and the image plot of 1-$P_k$ is shown in Fig. 2. From Fig. 2, we can determine which process variable is more influential for the product quality. This industrial gene map can help engineers to determine which process variable is important and which one is less important at a first glance. Engineers can grasp as much as information in the shortest time with the help of industrial gene map.

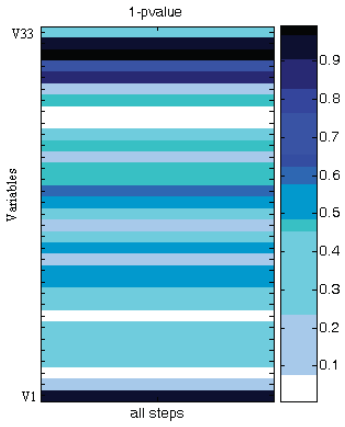


Fig. 2. Image of 1-$P_k$ of total average approach

To further know which step is critical for the process, equation (3) is evaluated for the profile variables of the process and the result is shown in Fig. 3. Similarly, Fig. 3 corresponds to a matrix of dimension 33 by 9. Obviously, the industrial gene map is more visual and straightforward. From Fig. 3, it is observed that most of critical steps are also related with the settings of temperature. The p-value of the profile variables in second are shown in Fig. 4. It can help engineers know when a fault is most likely to happen.

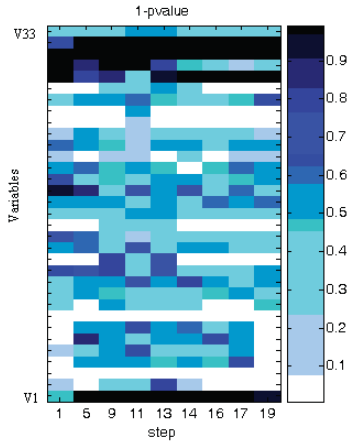


Fig. 3. Image plot of 1-$P_{k,b}$ of step average approach

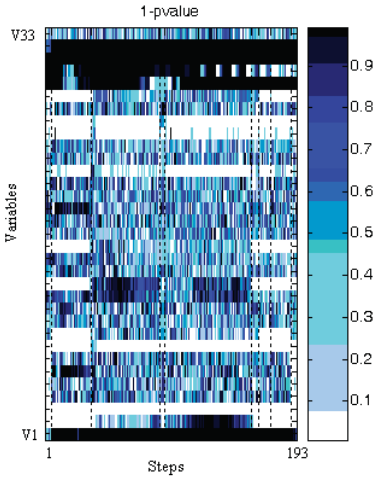


Fig. 4. Image plot of 1-$p_{k,l}$

## 3. VARIATION REDUCTION FOR MULTI-STAGE PROCESS

### *3.1 Problem statement*

In this section, a statistical method is proposed to find out the key variables that have essential effects on the product quality for the multi-stage manufacturing process. Similarly, the basic assumption is that there is relatively few quality data available compared with process variables. Then, specifications for the key variables which are usually far from optimal in original settings are designated to improve the

product yield. This framework provides a systematic method of drawing inferences from the available evidence without interrupting the normal process operation. The proposed method is directly illustrated by an industrial example. The statistical methods used in the following analysis include cluster analysis, canonical correlation analysis and stepwise regression.

*3.2 Statistical analysis and illustrative example*

Consider a CVD process. Every wafer must be processed by three chambers A, B, and C successively. Denote the process variables of chamber A, B and C as $X_A$, $X_B$ and $X_C$, respectively. The final quality variable is denoted as $Y$ which may contain wafer thickness measurements and wafer electrical measurements. In the following analysis, the method is illustrated for the wafer thickness $y$, which is one of the most important characteristics of wafers.

The numbers of steps and process variables for chamber A, B and C are list in Table 2. The data set includes measurements of 526 wafers from 22 batches. In this analysis, we want to know which variable, of which chamber, on which step, has an essential effect on the wafer thickness. Every process variable from different chambers on different is treated as an independent variable. Therefore, it is still the case that there are much more process variables than the quality data. Furthermore, process variables are usually highly correlated because of physical and chemical principles governing the process operation. To pick out the most influential variables for the quality variable $y$, the first step is to reduce the redundancy of the original data set.

**Table 2 Number of steps and variables of the three chambers**

| Chambers | Number of steps | Number of variables |
|---|---|---|
| A | 13 | 79 |
| B | 5 | 19 |
| C | 12 | 79 |

Cluster analysis is a useful technique used for combining observations into groups or clusters such that each group or cluster is homogeneous with respect to certain characteristics. Simultaneously, each group should be different from other groups with respect to the same characteristics (Sharma, 1996). The definition of similarity or homogeneity varies from analysis to analysis, and depends on the objectives of the study. In this study, it is desired to combine variables that are highly correlated into one group. Therefore, the similarity measure is defined as

$$d_{ij} = 1 - \left| r_{ij} \right| \qquad (4)$$

where $r_{ij}$ is the correlation coefficient of variables $x_i$ and $x_j$. For variables that are highly correlated, $d_{ij}$ would be small which represents similarity and vice versa. The clustering method adopted here is average-linkage method, one of the hierarchical clustering methods.

To determine the number of clusters, the rule that the correlation coefficient of the variables from the same groups should be greater than 0.9 is used. The result of cluster analysis is shown in Fig. 5-7. In these figures, variables that are filled with the same color or indicated with the same number are of the same group.

Then, the next step is to select representative variables from each group. The variables picked out should give good variance explanation which is usually evaluated by the $R^2$ statistics of the wafer thickness $y$. For example, the $R^2$ statistics of one variable selected from group 8 is 0.352 and the total $R^2$ of the whole group is 0.361. In such case, one process variable is capable of representing the group.

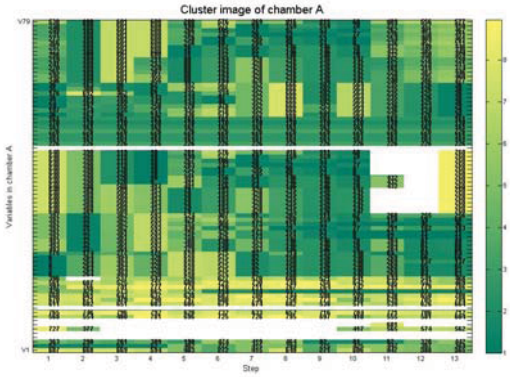


Fig. 5. Cluster image of chamber A.

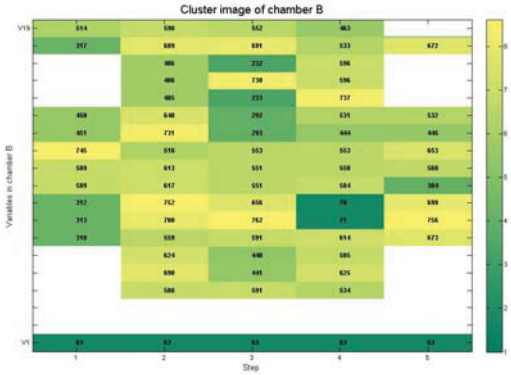


Fig. 6. Cluster image of chamber B.

However, in some circumstances, the $R^2$ of each individual variable is quite low yet the linear combination of these variables contributes a high $R^2$. In this case, it is more appropriate to use linear composites of the original variables to represent the group. This problem actually belongs to the field of canonical correlation analysis. The new variables, the linear composites, are called canonical variates. The coefficients of the canonical variates are determined to make the correlation between the linear composites maximum. For this special case, there is only one quality variable, the wafer thickness $y$. Therefore, canonical correlation analysis is essentially equal to the linear multiple regression.

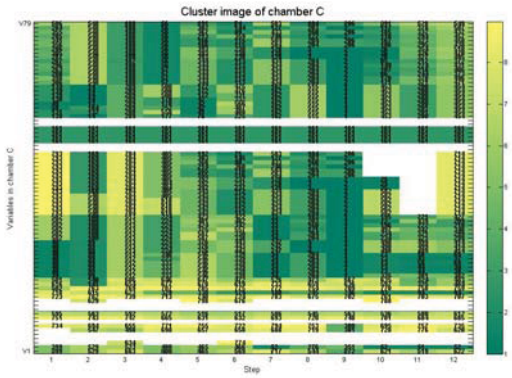


Fig. 7. Cluster image of chamber C.

Then, the question is, when a single variable should be used and when a linear composite should be used to represent a group. In this application, the following rules are adopted: if the $R^2$ of individual variable is more than eighty percent of the total $R^2$, then the single variable which has the largest $R^2$ is used to represent the whole group; otherwise, a linear composite is used. The number of variables in the canonical variate is increased till the $R^2$ of the linear composite is more than eighty percent of the total $R^2$. The coefficients of the linear composite are obtained from canonical correlation analysis.

After picking out the representative variable from each group, the next step is to select important representative variables from all the groups. The method used is stepwise regression. Stepwise regression is a statistical method used for variable selection in linear regression. The procedure iteratively constructs a sequence of regression models by adding or removing variables at each step. The criterion for adding or removing a variable at any step is usually expressed in terms of a partial *F*-test Montgomery, et al., 2001). The changes of $R^2$ and adjusted $R^2$ of stepwise regression are shown in Fig. 8. There are 68 representative variables selected by the stepwise regression.

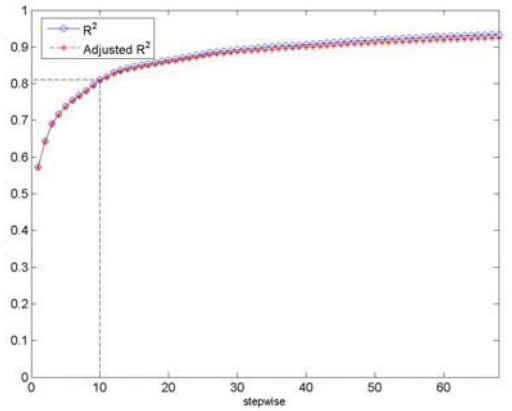


Fig. 8. Representative variables selected by stepwise regression method.

It is not an easy task to monitor 68 variables online simultaneously. Therefore, the first ten representative variables selected by stepwise regression are picked out and analyzed. The first ten representative variables listed in Table 8 give a good variance explanation because the $R^2$ and adjusted $R^2$ are higher than 0.8 which can be seen from Fig. 8.

In fact, all the 526 wafers are qualified wafers. To reduce the variance of wafer thickness further, we define $\left[\bar{y}-1.5s_y, \bar{y}+1.5s_y\right]$ as the acceptable region for the wafer thickness. Here, $\bar{y}$ is the average value of $y$ and $s_y$ is the standard deviation of $y$, respectively. The wafers fall out of this region is treated as "un-qualified" now. Among all the 526 wafers, there are 455 wafers fall into the acceptable region. Therefore, the yield is 0.865. In the following analysis, we will develop a nonparametric method to find out the new specifications for the above ten important representatives to improve the product yield.

First, the center point for all the qualified wafers in a space defined by the 10 important representative variables is determined. The Mahalanobis distance of each qualified wafer from the center point is calculated as

$$MD_i = \left(X_i - \mu\right)^T S \left(X_i - \mu\right) = c_i \quad (5)$$

where $X$ is a 10×1 vector of coordinates and $S$ is a 10×10 covariance matrix, $\mu$ is the center point. Then, the yield can be viewed as an implicit function of the Mahalanobis distance. Each value of Mahalanobis distance corresponds to a value of yield which is defined as the ratio between the number of qualified wafers and the number of all the wafers within the Mahalanobis distance. A graphical interpretation of this relationship is shown in Fig. 9. In this figure, the solid line is the relationship between the yield and the Mahalanobis distance and the dashed line is its 95% confidence interval.

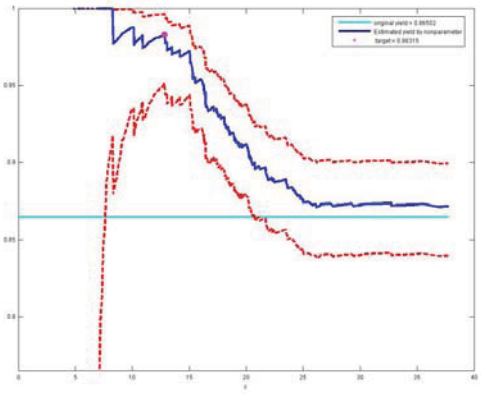

Fig. 9. Plot of Mahalanobis distance versus product yield.

It can be observed that the yield is not reliable when $c_i$ is small because the samples within the corresponding Mahalanobis distance are few. To get a balance between reliability and high yield, the point corresponds to one third of the maximum of $c_i$ which is marked as a dot in Fig. 9 is

used to derive the specifications of the ten representative variables. Once $c_i$ is determined, the joint boundary of the ten representative variables is also determined.

However, the joint boundary which is a function of ten independent variables can not be easily monitored. Therefore, the projections of the joint boundary onto the axes of coordinates are used as the new specifications of the ten representative variables. The yield increased greatly when the upper and lower bounds of the first representative variable are designated. A graphical interpretation of the increase of the yield is shown in Fig. 10. The increases of the yield are not obvious after the designation of the specification of the third representative variable.

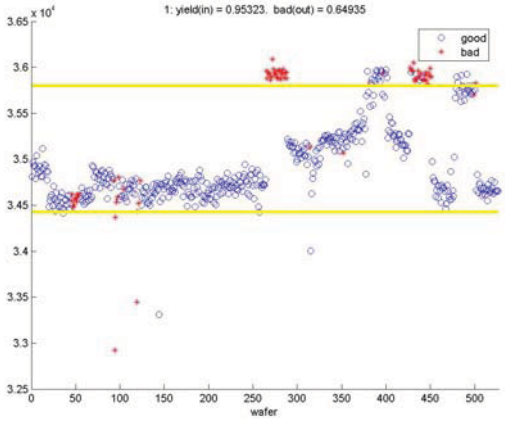


Fig. 10. Specifications of the first representative variable.

It is of interest to study the improvement of process capability ratio after the specifications of the ten representative variables are designated. The process capability ratio (PCR, or $C_p$) is defined as

$$C_p = \frac{USL - LSL}{6\sigma} \qquad (6)$$

where USL and LSL are the upper and lower specification limits, respectively. Since $\sigma$ is unknown, it is replaced by the standard deviation $s$. If the process capability ratio and standard deviation are treated as a function of Mahalanobis distance, then we can get

$$\frac{C_p(c_i)}{C_p} = \frac{s}{s(c_i)} \qquad (7)$$

The relationship between $C_p(c_i)/C_p$ and the Mahalanobis distance is shown in Fig. 11. It can be observed that there is about 40% improvement of process capability ratio for the point we used to designate the specifications of the representative variables. The changing trend of $C_p(c_i)/C_p$ is consistent in the area where the point we used also indicates that the value of $c_i$ we chose is appropriate.

## 4. CONCLUSIONS

Nowadays, many semiconductor manufacturing foundries are operating with diversified products of small account which makes the fault detection and variation reduction difficult. In this paper, systematic statistical methods are proposed to solve this difficulty. Both single stage process and multi-stage process are considered. The effectiveness of the proposed methods are illustrated by industrial examples.

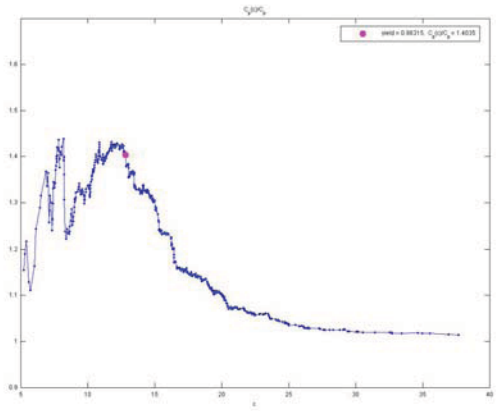

Fig. 11. Plot of Mahalanobis distance versus $C_p(c_i)/C_p$ .

## REFERENCES


Cherry, G.A. and Qin, S.J. (2006). Multiblock principal component analysis based on a combined index for semiconductor fault detection and diagnosis. *IEEE Trans. Semiconduct. Mnauf.*, 19 (2), pp 159-172.

Ma, M.D., Chang, C.C., Wong, D.S.H., and Jang, S.S. (2008). Identification of tool and product effects in a mixed product and parallel tool environment. *J. Process Contr.*, doi:10.1016/j.jprocont.2008.07.009.

Montgomery, D.C. (1993). *Introduction to Statistical Quality Control (3ed.)*. Wiley, New York.

Montgomery, D.C., Peck, E.A., and Vining, G.G. (2001). *Introduction to Linear Regression Analysis (3ed.)*. Wiley, New York.

Nomikos, P. and MacGregor, J.F. (1994). Monitoring batch processes using multiway principl analysis. *AIChE Journal*, 40 (8), pp 1361-1375.

Sharma, S. (1996). *Applied Multivariate Techniques*. Wiley, New York.

Wong, J. (2006). Batch PLS analysis and FDC process control of within lot SiON gate oxide thickness variation in sub-nanometer range. *Proc. AEC/APC Symp. XVIII*, Westminster, CO, Sep.

Yue, H.H., Qin, S.J., Markle, R.J., Nauert, C., and Gatto, M. (2000). Fault detection of plasma etchers using optical emission spectra. *IEEE Trans. Semiconduct. Mnauf.*, 13 (3), pp 374-385.

Baldi, P. and S. Brunak. (2001). *Bioinformatics: The Machine Learning Approach*. MIT Press.

Higgins, J.J. (2004). *Introduction to Modern Nonparametric Statistics*, Brooks.