Fault Detection Based on Statistical Multivariate Analysis and Microarray Visualization

Ming-Da Ma, David Shan-Hill Wong, Shi-Shang Jang, and Sheng-Tsaing Tseng, *Member, IEEE*

***Abstract*—In this work, a statistical method is proposed to mine out key variables from a large set of variables recorded in a limited number of runs through a multistage multistep manufacturing process. The method employed well-known single variable or multivariable techniques of discrimination and regression but also presented a synopsis of analysis results in a colored map of p-values very similar to a DNA microarray. This framework provides a systematic method of drawing inferences from the available evidence without interrupting the normal process operation. The proposed concept is illustrated by two industrial examples.**

***Index Terms*—Fault detection, microarray, quality improvement, semiconductor manufacturing, Wilcoxon rank-sum test.**

I. 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) methods such as run-to-run control (RtR) 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 to mine out key variables from a large set of variables recorded in a limited number of runs for sophisticated multistage multistep semiconductor manufacturing process.

Detection of faults in the shortest possible time is critical for minimizing scrap wafers and improving product yields for semiconductor manufacturing. One can monitor wafer-state data such as critical dimension, uniformity, and thickness, etc., and compare the result with a specified target and control limits using statistical process control (SPC). 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. Meanwhile, more and more real-time measurements of process variables such as temperature, pressure, power and flow rate, etc., are available as the manufacturing tool become more sophisticated and expensive. 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.

Manuscript received December 31, 2008; revised July 06, 2009; accepted July 19, 2009. Paper no.TII-08-12-0235.R1.

M.-D. Ma is with the Center for Control and Guidance Technology, Harbin Institute of Technology, Harbin 150001, China (e-mail: mamingda@hit.edu.cn).

D. S.-H. Wong and S.-S. Jang are with the Department of Chemical Engineering, National Tsing-Hua University, Hsin-Chu, Taiwan 30013 (e-mail: dshwong@che.nthu.edu.tw; ssjang@mx.nthu.edu.tw).

S.-T. Tseng is with the Institute of Statistics, National Tsing-Hua University, Hsin-Chu, Taiwan 30013 (e-mail: sttseng@stat.nthu.edu.tw).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TII.2009.2030793

Traditionally, statistical process control using single variable control charts or multivariate statistical analysis (MVA) such as factor analysis, principal component analysis (PCA), and partial least square (PLS) could be applied to monitor tool status. For example, Wise *et al.* [4] have compared a wide assortment of factor analysis techniques for use in fault detection of plasma etcher. Spitzlsperger *et al.* [6] used Hotelling T^2 for fault detection of a via etch process by compressing multivariate data into a small number of latent variables or statistical parameters that can be monitored using control chart. Goodlin *et al.* [2] proposed a method to simultaneously detect and classify faults in a single-step using fault-specific control charts which are designed to discriminate between specific fault classes and the normal process operation. Yue *et al.* [5] applied multiway PCA method to optical emission spectra for plasma etchers.

Despite the effectiveness of the PCA/PLS method in dealing with the problems of high dimension and collinearities, each PC or latent variable is a linear combination of all variables, it is often difficult to interpret it. Selecting the variables with high loadings on the PCs as key variables is error prone as highlighted by Cadima and Jolliffe [13]. Variable selection techniques as an alternate approach to dimensionality reduction which seeks to identify a subset of measurement space that contains as much information as possible was proposed by McCabe [10]. The most informative variables are termed as principle variables. The variable selection techniques are extended to Longitudinal data and multistage processes recently [11], [12].

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 the development phase of a product, one has to find out the causes of unqualified wafers with limited amount of operating data. Furthermore, although these real-time sensors provide valuable information about the tool status but their relations to final wafer characteristics are unknown. Engineers are eager to know which variable in which step in a multistage-, multistep-, and multivariable production line is likely the key for inferior product quality. A quick diagnosis of possible cause of unqualified product is more desirable.

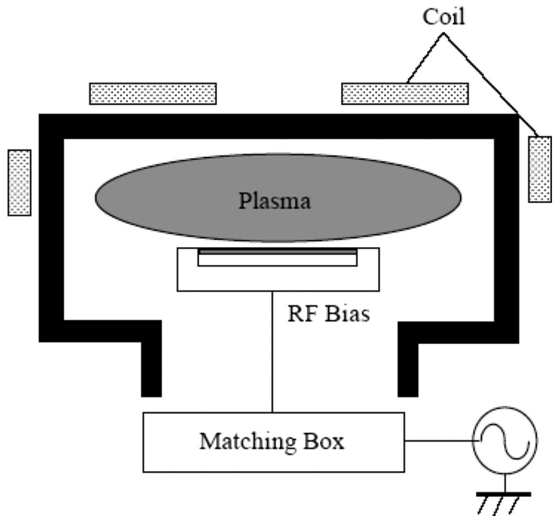


Fig. 1. Schematic diagram of the HDP-CVD reactor.

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 [7]. 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 [8]. 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.

II. Visualization of Simple Fault Detection

Consider a high-density plasma chemical vapor deposition (HDP-CVD) process, as shown in Fig. 1. There are 33 process variables for this manufacturing process and 9 steps with 193 measurement times. The quality data of 25 wafers are collected, among which 21 wafers are qualified and 4 wafers are defective.

More generally, assume that quality data of n wafers are collected from a tool, n_1 wafers are qualified, and n_2 wafers are unqualified, hence $n_1 + n_2 = n$. Let X_{ijk} be value of j^{th} ($j = 1 \ldots J$) variable at the k^{th} time points ($k = 1 \ldots K$) of the i^{th} wafer. Now, the question becomes what is the probability that the value of X_{ijk} measured for the qualified wafers is different from unqualified wafers. We can calculate the p-value of testing the mean of two groups at each batch time under the following hypothesis:

$$\begin{aligned} H_0 &: \mu^1_{jk} = \mu^2_{jk} \\ H_a &: \mu^1_{jk} \neq \mu^2_{jk} \end{aligned} \tag{1}$$

where H_0 is the null hypothesis and H_a is the alternative hypothesis, μ^1_{jk} and μ^2_{jk} are the means of qualified and unqualified wafers of the j^{th} variable at k^{th} time point, superscript 1 and 2 stand for qualified and unqualified wafers, respectively. A two-side test is implemented here since there is no prior knowledge of whether that μ^1_{jk} should be greater or smaller than μ^2_{jk}. This p-value can be used to determine whether variable j at time k is a good candidate for discriminating qualified and unqualified wafers.

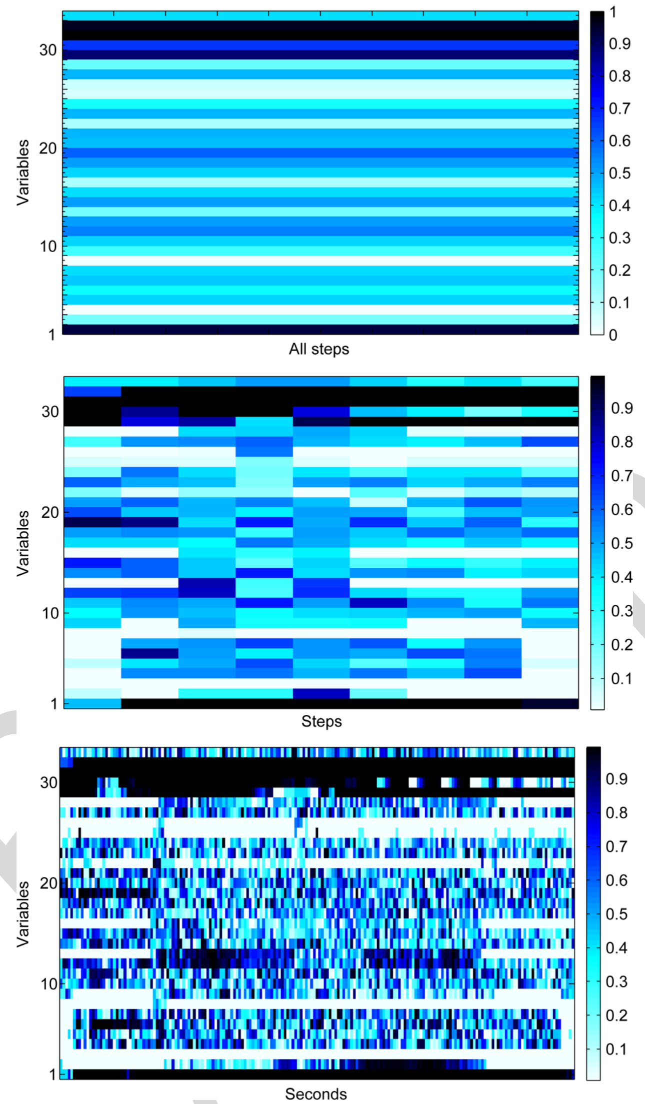


Fig. 2. Colored map synopsis images of different variables, different variables at different steps, and different variables at different time points.

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 two sets of data may not be normally distributed. The Wilcoxon rank-sum test is a nonparametric alternative which is based solely on the order in which the observations from the two samples fall. 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 rank-sum test statistic is the sum of the ranks for all observations from each set (qualified or unqualified) of samples [9].

Let p_{jk} be the p-value of the hypothesis (1) for the j^{th} variable at k^{th} time point. Note that the above analysis can be used

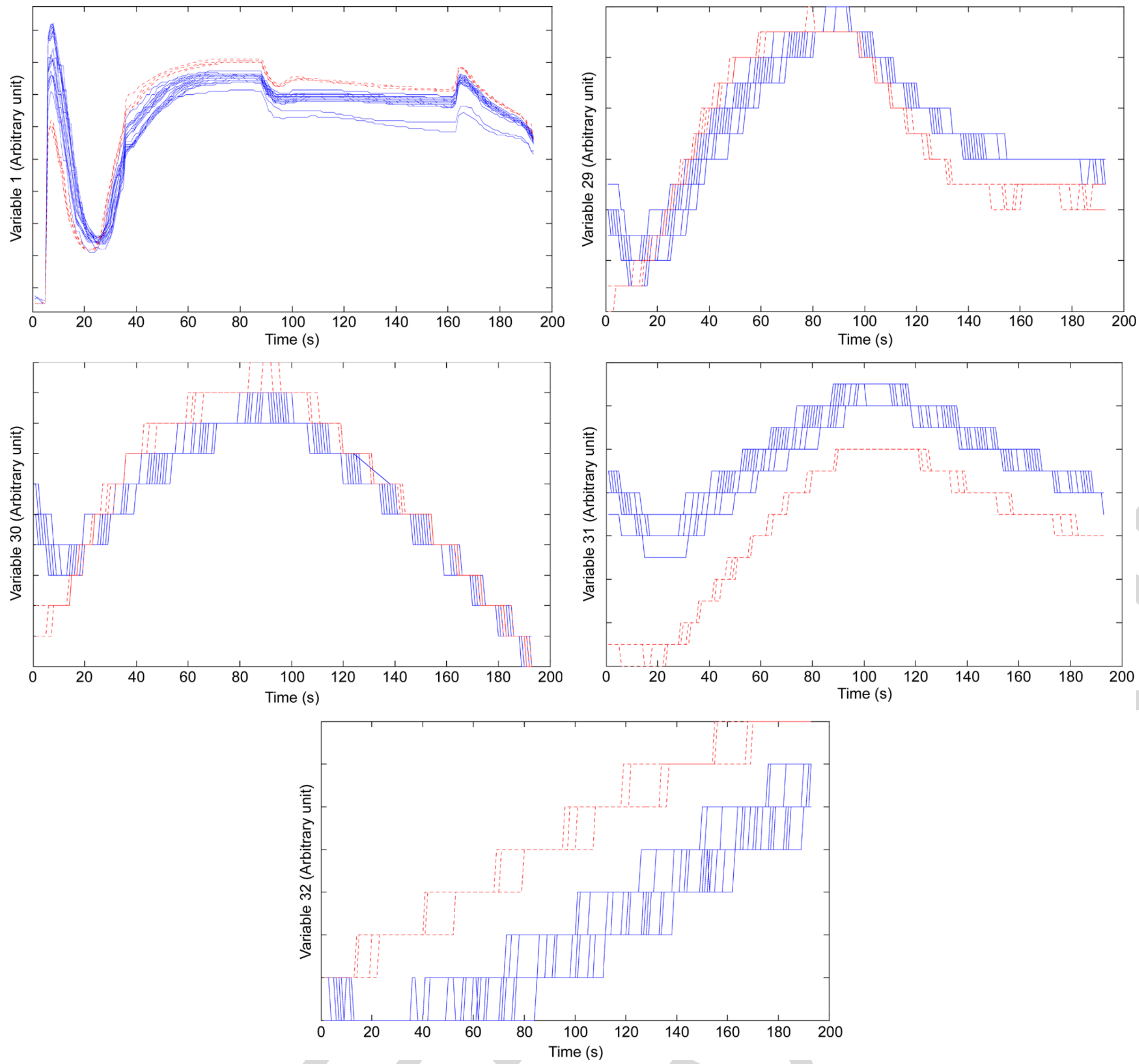


Fig. 3. Profiles of the top five influential process variables (solid line, qualified; dashed line, unqualified).

for different step s $(s = 1 \dots S)$ of the process or the entire process by taking the appropriate averages of p_{jk}

$$P_{js} = \frac{\sum_{k=1}^{K} p_{jk}\delta_{sk}}{\sum_{k=1}^{K} \delta_{sk}}$$

$$\delta_{sk} = \begin{cases} 1, & \text{time point } k \text{ belongs to step } s \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

$$P_j = \frac{1}{K}\sum_{k=1}^{K} p_{jk} \tag{3}$$

where δ_{sk} is Kronecker delta. The results of above discrimination analysis can be easily visualized in a microarray-type expression shown in Fig. 2. Engineers can easily learn that variables 1, 29, 30, 31, and 32 are crucial to the wafer quality. It also can be learned from these figures which step or which time point is important for the final product quality. These are valuable information for engineers. The profiles of these five variables for qualified and unqualified wafers are plotted in Fig. 3.

The key variables found by statistical analysis are dome temperatures and wafer temperature. Fluctuation in dome temperature may generate particles that deposit on the wafer and may result in nonuniformity and lack of repeatability of a deposited film of material. In addition, variation in temperature over regions of the dome may result in excessive mechanical stress that can ultimately result in dome fracture. The result of the above statistical analysis is tested and confirmed by operating engineers.

III. Variation Reduction for Multistage CVD Process

Consider a chemical vapor deposition CVD process in which every wafer must be processed in 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 can be either film thickness measurements or wafer electrical measurements. There are 13, 5, 12 steps and

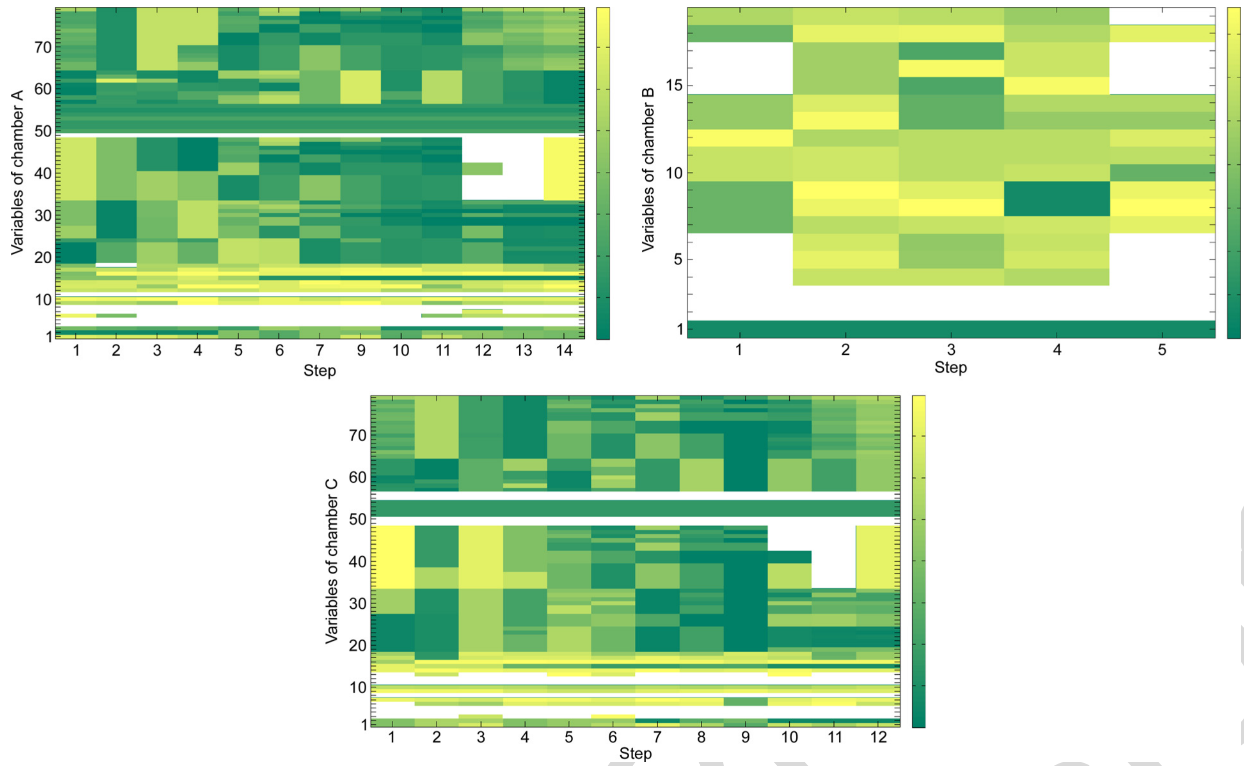


Fig. 4. Grouping of feature variables for the three chambers.

TABLE I
VARIABLES IN GROUP 8

Group 8	Total $R^2 = 0.36113$			
	Individual R^2	Correlation	Variable 1	Variable 2
	0.31203	Variable 1	1	0.97619
	0.35208	Variable 2	0.97619	1

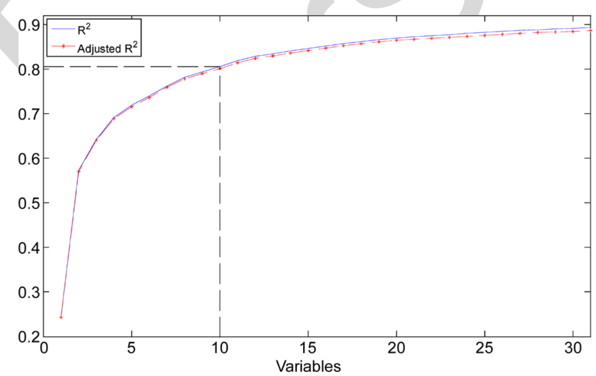


Fig. 5. Changes of R^2 and adjusted R^2 of stepwise regression.

79, 19, and 79 process variables for chamber A, B and C, respectively. The data set includes measurements of 526 wafers from 22 batches. A feature variable, e.g., average, max, min, range, is provided for each variable in each step. In this analysis, we want to know which of these feature variables will have significant effect on final quality and how the specifications of these process variables can be tightened or changed to improve final product quality.

To pick out the most influential variables, the first step is to reduce the dimension of the original data set. Process variables are usually highly correlated because of physical and chemical principles governing the process operation. The correlation is assumed to be linear in the following analysis. Cluster analysis is a useful technique used for combining variables into groups or clusters such that each group or cluster is homogeneous with respect to certain characteristics; while each group should be different from other groups with respect to the same characteristics. 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

$$d_{ij} = 1 - |r_{ij}| \tag{4}$$

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 [1].

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. 4. In these figures, variables that are filled with the same color are of the same group. The blank areas or the white color

TABLE II
VARIABLES IN GROUP 1

Group 1	Total $R^2 = 0.6425$							
	Individual R^2	Corr	Variable 1	Variable 2	Variable 3	Variable 4	Variable 5	Variable 6
	0.033207	Variable 1	1	0.97299	0.974	0.99135	0.95736	0.97312
	0.063261	Variable 2	0.97299	1	0.96825	0.98674	0.95048	0.98961
	0.017806	Variable 3	0.974	0.96825	1	0.97822	0.97716	0.95183
	0.03761	Variable 4	0.99135	0.98674	0.97822	1	0.97323	0.98846
	0.0010343	Variable 5	0.95736	0.95048	0.97716	0.97323	1	0.93974
	0.085809	Variable 6	0.97312	0.98961	0.95183	0.98846	0.93974	1

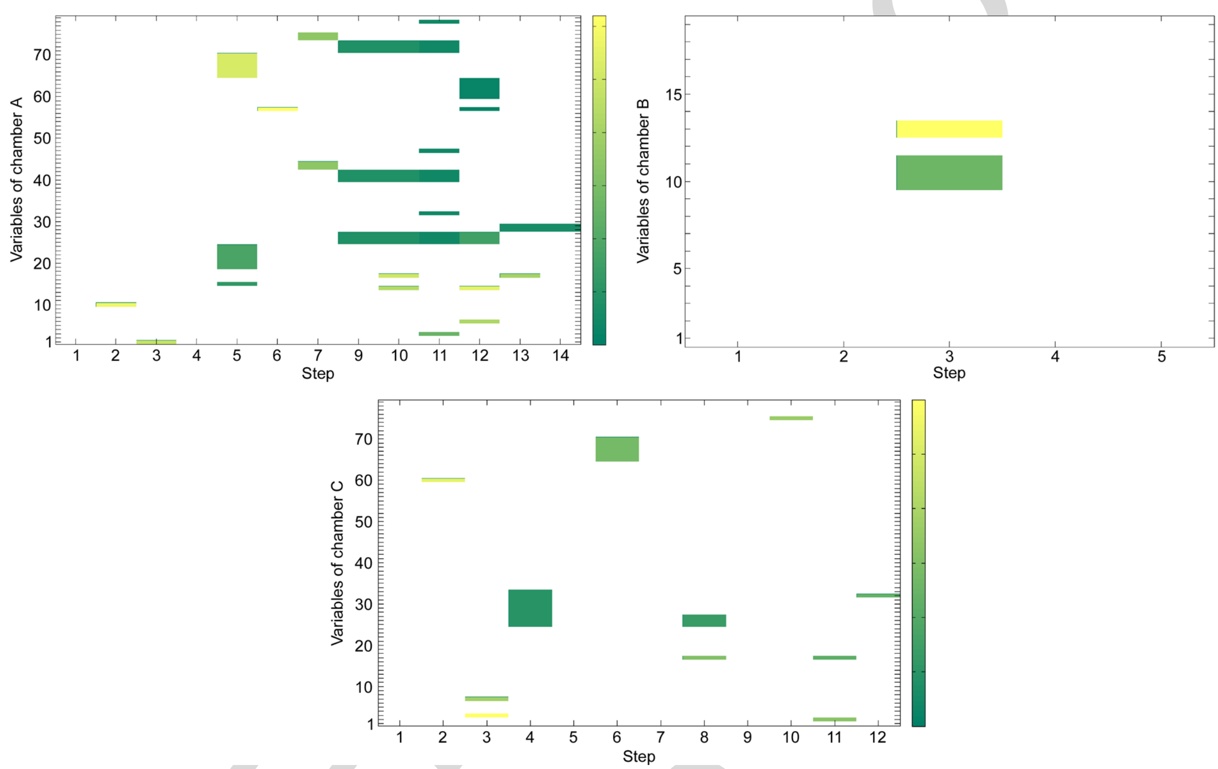


Fig. 6. The distributions of key clusters for the three chambers.

areas are caused by the result that some variables do not have observations at some steps.

The next step is to select representative variables from each group. The variables picked out should give good variance explanation of the quality index y which is usually evaluated by the R^2 statistics. For example, in Table I, the R^2 statistics of the variable 2 selected from group 8 is 0.352 and the total R^2 of the whole group is 0.361. In such a case, the variable 2 of group 8 is capable of representing the group. 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 which is the case of group 1 shown in Table II. 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 quality index and linear composites maximum. In this application, the following rules are adopted: if the R^2 of individual variable is more than 80% 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 until the R^2 of the linear composite is more than 80% of the total R^2 of all the variables in the group.

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

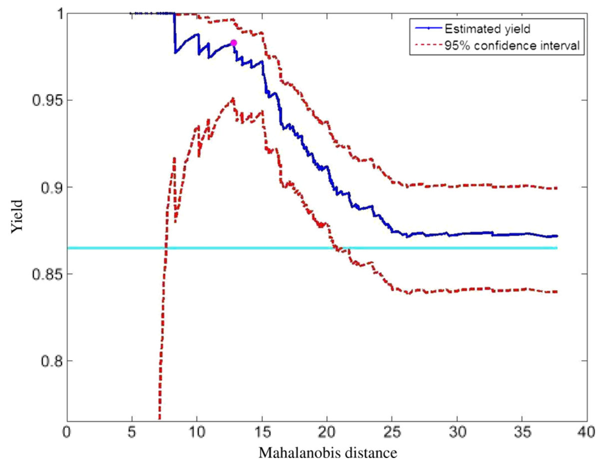


Fig. 7. Relations between the mean and 95% confidence estimates of yield and Mahalanobis distance from the center point.

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 [3]. The changes of R^2 and adjusted R^2 of stepwise regression are shown in Fig. 5. Thirty-one representative variables were selected by the stepwise regression algorithm. The distributions of the corresponding key clusters are shown in Fig. 6. In these figures, the clusters are labeled and colored by the rank of their order of selection. It is observed that most of the highly ranked variables are found in Chamber A.

In this preliminary experiment, all the 526 wafers are qualified wafers. To reduce the variance of quality characteristics, we define $[\overline{y} - 1.5s_y, \overline{y} + 1.5s_y]$ as the acceptable region for the wafer thickness. Here, $\overline{y}$ is the average value of y and s_y is the standard deviation of y, respectively. The wafers fall out of this region are treated as "unqualified" 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 selected key variables to improve the product yield. First, the center point for all the qualified wafers $\vec{\mu}$ in a space defined by the 31 important representative variables is determined. The Mahalanobis distance of each qualified wafer from the center point is calculated as

$$MD_i = (\vec{X}_i - \vec{\mu})^T \vec{S} (\vec{X}_i - \vec{\mu}) = c_i \tag{5}$$

where $\vec{X}$ is a 31×1 vector of coordinates and $\vec{S}$ is a 31×31 covariance matrix. Then, the yield can be viewed as an implicit function of the Mahalanobis distance. The distribution of this yield can be calculated by bootstrap sampling of the original wafers within the given Mahalanobis distance. A graphical interpretation of this relationship is shown in Fig. 7. In this figure, the solid line is the relationship between the mean of yield distribution estimated by bootstrap sampling and the Mahalanobis distance. The dashed lines are its 95% confidence intervals of the yield estimated by bootstrap sampling. The horizontal line is the original yield 0.865.

It is obvious that the estimated 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 the maximum of lower bound estimate of 95% confidence interval which is marked as a dot in Fig. 7 is used. The mean estimate is 98.3%. Once c_i is determined, the joint boundary of these variables is also determined. However, the joint boundary which is a function of 31 independent variables cannot be easily monitored online simultaneously. Therefore, the projections of the joint boundary onto the axes of coordinates are used as the new specifications of the five most influential variables. The new yield can be estimated again by bootstrap method. The results are listed in Table III. The new upper specification limit (USL) and lower specification limit (LSL) are presented in units to the original targets. It can be observed from Table III that the yield increased greatly when the upper and lower bounds of the first representative variable are designated. The increases of the yield are not obvious after the designation of the specification of the third representative variable. A graphical interpretation of the increase of the yield when the USL and LSL of key variable 1 are designated is shown in Fig. 8. The yield for the wafers fall into the regions defined by USL and LSL is about 95% and the yield for the wafers outside the USL and LSL is 54% which indicates reasonable specifications.

It is interesting to note that according to the engineers' experience, the manufacturing on chamber B is crucial for the final quality; however, the key variable 1 found by the statistical analysis is actual temperature measurement of the last step of chamber A. After a careful examination, it was found that the waiting time before a wafer was sent to chamber B is different. This made the temperature of wafers different when they were fabricated on chamber B. After the variation source was found, the final product quality was improved greatly.

TABLE III
THE USLs AND LSLs FOR THE TOP FIVE VARIABLES AND THE INCREASES OF YIELD

Key variables	USL	LSL	Yield
1	1.023	0.984	0.95323
2	1.003	0.991	0.98276
3	1.012	0.959	0.98465
4	1.002	0.783	0.98656
5	1.270	0.313	0.98652

(Note: The key variables could be single process variable or linear composite.)

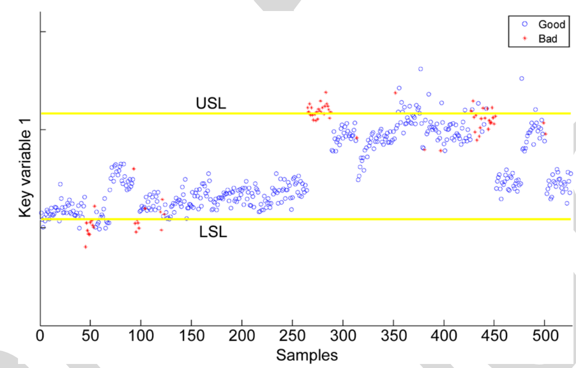


Fig. 8. USL and LSL for key variable 1.

IV. Conclusion

In the two given examples, we have presented a method of finding the key variables from a large set of variables recorded in a limited number of runs through a multistage multistep manufacturing process. It should be pointed out that although the statistical methods employed, e.g., Wilcoxon rank-sum test in the first example, clustering, regression and bootstrap estimates in the second are not new, however, the summary of analysis results, presented in a colored map of p-values which is very similar to a DNA microarray, allowed engineers to visualize them most intuitively. In an error of data explosion, it is important to develop industrial informatics tools that mine out important information from large amount of data and present them in succinct manners.

References

[1] S. Sharma, *Applied Multivariate Techniques*. New York: Wiley, 1996.

[2] B. E. Goodlin, D. S. Boning, H. H. Sawin, and B. M. Wise, "Simultaneous fault detection and classification for semiconductor manufacturing tools," *J. Electrochem. Soc.*, vol. 150, no. 12, pp. 778–784, 2003.

[3] D. C. Montgomery, E. A. Peck, and G. G. Vining, *Introduction to Linear Regression Analysis*, 3rd ed. New York: Wiley, 2001.

[4] B. M. Wise, N. B. Gallagher, S. W. Butler, D. D. White, Jr., and G. G. Barna, "A comparison of principal component analysis, multiway principal component analysis, trilinear decomposition and parallel factor analysis for fault detection in a semiconductor etch process," *J. Chemometrics*, vol. 13, pp. 379–396, 1999.

[5] H. H. Yue, S. J. Qin, R. J. Markle, C. Nauert, and M. Gatto, "Fault detection of plasma etchers using optical emission spectra," *IEEE Trans. Semiconduct. Manuf.*, vol. 13, no. 3, pp. 374–385, 2000.

[6] G. Spitzlsperger, C. Schmidt, G. Ernst, H. Strasser, and M. Speil, "Fault detection for a via etch process using adaptive multivariate methods," *IEEE Trans. Semiconduct. Manuf.*, vol. 18, no. 4, pp. 528–533, 2005.

[7] P. Baldi and S. Brunak, *Bioinformatics: The Machine Learning Approach*. Cambridge, MA: MIT Press, 2001.

[8] National Institute of Health, 2008, "DNA Microarry Technology." [Online]. Available: http://www.genome.gov/page.cfm?pageID=10000533

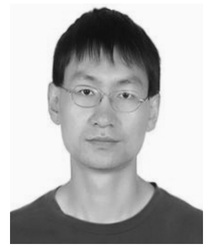
[9] J. J. Higgins, *Introduction to Modern Nonparametric Statistics*. Pacific Groove, CA: Brooks, 2004.

[10] G. P. McCabe, "Principal variables," *Technometrics*, vol. 26, no. 2, pp. 137–144, 1984.

[11] J. A. Cumming and D. A. Wooff, "Dimension reduction via principal variables," *Comput. Statist. Data Analy.*, vol. 52, no. 1, pp. 550–565, 2007.

[12] R. Srinivasan and M. S. Qian, "State-specific key variables for monitoring multi-state processes," *Chem. Eng. Res. Design*, vol. 85, no. 12, pp. 1630–1644, 2007.

[13] J. F. Cadima and I. T. Jolliffe, "Loadings and correlations in the interpretation of principal components," *J. Appl. Stat.*, vol. 22, no. 2, pp. 203–214, 1995.



Ming-Da Ma received the B.S. and M. S. degrees in control engineering from North China Electric Power University, Beijing, China, and the Ph.D. degree in control engineering from Shanghai Jiao Tong University, Shanghai, China, in 2000, 2003, and 2006, respectively.

He has been a Lecturer in the Center for Control and Guidance Technology, Harbin Institute of Technology, Harbin, China, since 2006. His research interests include advanced process control, process monitoring, statistical process control and robust control.

David Shan Hill Wong received the B.S. degree in chemical engineering from the California Institute of Technology, Pasadena, CA, in 1978 and the Ph.D. degree in chemical engineering from the University of Delaware, Newark, DE, in 1982.

He has been a faculty member of the Chemical Engineering Department, National Tsing-Hua University, Hsinchu, Taiwan, since 1983. His research interests in the area of process system engineering includes design and control of energy efficient separation processes and advanced process control in batch-based manufacturing. He is actively involved in consulting and collaborative research with chemical process and semiconductor manufacturing industries.

Shi-Shang Jang received the M.S. degree in chemical engineering from the National Taiwan University, Taipei, Taiwan, R.O.C., in 1980 and the Ph.D. degree in chemical engineering from Washington University, St. Louis, MO, in 1986.

Since 1992, he has been a Professor in the Chemical Engineering Department, National Tsing-Hua University, Hsinchu, Taiwan. He was the Chairman of the department from 2000 to 2004. He has been involved in many projects granted by Taiwan Semiconductor Corporation (TSMC) for the last five years. He is particularly interested in run-to-run control theory and practice.

Sheng-Tsaing Tseng (M'92) received the B.S. degree in business mathematics from Soochow University, the M. S. degree in applied mathematics from Tsing-Hua University, Hsinchu, Taiwan, and the Ph.D. degree in management science from Tamkang University, Taiwan.

Currently, he is a Professor at the Institute of Statistics, Tsing-Hua University. His research interests include quality and productivity improvement, reliability lifetime analysis, and statistical decision methodology. His articles have appeared in numerous technical journals, such as *Technometrics, Journal of Quality Technology*, the IEEE Transactions on Semiconductor Manufacturing, IEEE Transactions on Reliability,*IIE Transactions*, and etc.

Prof. Tseng received the Outstanding Research Award from the National Science Council (NSC) of R.O.C., Taiwan, in 1993, 1999, and 2004, respectively. Currently, he serves as an Associate Editor of *Journal of Statistical Planning and Inference* and he is an elected member of ISI.

Fault Detection Based on Statistical Multivariate Analysis and Microarray Visualization

Ming-Da Ma, David Shan-Hill Wong, Shi-Shang Jang, and Sheng-Tsaing Tseng, *Member, IEEE*

***Abstract*—In this work, a statistical method is proposed to mine out key variables from a large set of variables recorded in a limited number of runs through a multistage multistep manufacturing process. The method employed well-known single variable or multivariable techniques of discrimination and regression but also presented a synopsis of analysis results in a colored map of p-values very similar to a DNA microarray. This framework provides a systematic method of drawing inferences from the available evidence without interrupting the normal process operation. The proposed concept is illustrated by two industrial examples.**

***Index Terms*—Fault detection, microarray, quality improvement, semiconductor manufacturing, Wilcoxon rank-sum test.**

I. 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) methods such as run-to-run control (RtR) 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 to mine out key variables from a large set of variables recorded in a limited number of runs for sophisticated multistage multistep semiconductor manufacturing process.

Detection of faults in the shortest possible time is critical for minimizing scrap wafers and improving product yields for semiconductor manufacturing. One can monitor wafer-state data such as critical dimension, uniformity, and thickness, etc., and compare the result with a specified target and control limits using statistical process control (SPC). 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. Meanwhile, more and more real-time measurements of process variables such as temperature, pressure, power and flow rate, etc., are available as the manufacturing tool become more sophisticated and expensive. 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.

Traditionally, statistical process control using single variable control charts or multivariate statistical analysis (MVA) such as factor analysis, principal component analysis (PCA), and partial least square (PLS) could be applied to monitor tool status. For example, Wise *et al.* [4] have compared a wide assortment of factor analysis techniques for use in fault detection of plasma etcher. Spitzlsperger *et al.* [6] used Hotelling T^2 for fault detection of a via etch process by compressing multivariate data into a small number of latent variables or statistical parameters that can be monitored using control chart. Goodlin *et al.* [2] proposed a method to simultaneously detect and classify faults in a single-step using fault-specific control charts which are designed to discriminate between specific fault classes and the normal process operation. Yue *et al.* [5] applied multiway PCA method to optical emission spectra for plasma etchers.

Despite the effectiveness of the PCA/PLS method in dealing with the problems of high dimension and collinearities, each PC or latent variable is a linear combination of all variables, it is often difficult to interpret it. Selecting the variables with high loadings on the PCs as key variables is error prone as highlighted by Cadima and Jolliffe [13]. Variable selection techniques as an alternate approach to dimensionality reduction which seeks to identify a subset of measurement space that contains as much information as possible was proposed by McCabe [10]. The most informative variables are termed as principle variables. The variable selection techniques are extended to Longitudinal data and multistage processes recently [11], [12].

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 the development phase of a product, one has to find out the causes of unqualified wafers with limited amount of operating data. Furthermore, although these real-time sensors provide valuable information about the tool status but their relations to final wafer characteristics are unknown. Engineers are eager to know which variable in which step in a multistage-, multistep-, and multivariable production line is likely the key for inferior product quality. A quick diagnosis of possible cause of unqualified product is more desirable.

Manuscript received December 31, 2008; revised July 06, 2009; accepted July 19, 2009. Paper no.TII-08-12-0235.R1.

M.-D. Ma is with the Center for Control and Guidance Technology, Harbin Institute of Technology, Harbin 150001, China (e-mail: mamingda@hit.edu.cn).

D. S.-H. Wong and S.-S. Jang are with the Department of Chemical Engineering, National Tsing-Hua University, Hsin-Chu, Taiwan 30013 (e-mail: dshwong@che.nthu.edu.tw; ssjang@mx.nthu.edu.tw).

S.-T. Tseng is with the Institute of Statistics, National Tsing-Hua University, Hsin-Chu, Taiwan 30013 (e-mail: sttseng@stat.nthu.edu.tw).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TII.2009.2030793

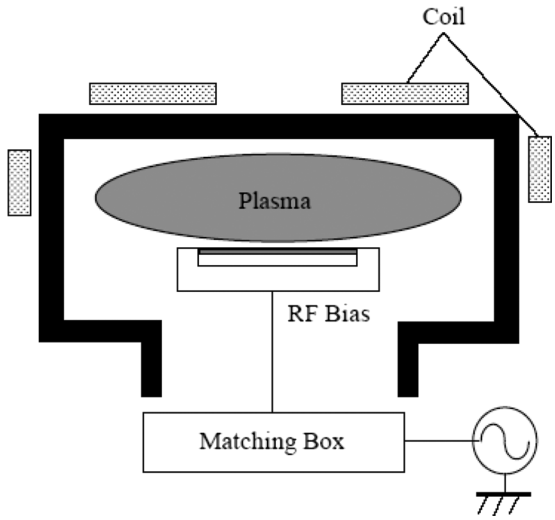


Fig. 1. Schematic diagram of the HDP-CVD reactor.

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 [7]. 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 [8]. 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.

II. Visualization of Simple Fault Detection

Consider a high-density plasma chemical vapor deposition (HDP-CVD) process, as shown in Fig. 1. There are 33 process variables for this manufacturing process and 9 steps with 193 measurement times. The quality data of 25 wafers are collected, among which 21 wafers are qualified and 4 wafers are defective.

More generally, assume that quality data of n wafers are collected from a tool, n_1 wafers are qualified, and n_2 wafers are unqualified, hence $n_1 + n_2 = n$. Let X_{ijk} be value of j^{th} ($j = 1 \dots J$) variable at the k^{th} time points ($k = 1 \dots K$) of the i^{th} wafer. Now, the question becomes what is the probability that the value of X_{ijk} measured for the qualified wafers is different from unqualified wafers. We can calculate the p-value of testing the mean of two groups at each batch time under the following hypothesis:

$$\begin{aligned} H_0 &: \mu^1_{jk} = \mu^2_{jk} \\ H_a &: \mu^1_{jk} \neq \mu^2_{jk} \end{aligned} \tag{1}$$

where H_0 is the null hypothesis and H_a is the alternative hypothesis, μ^1_{jk} and μ^2_{jk} are the means of qualified and unqualified wafers of the j^{th} variable at k^{th} time point, superscript 1 and 2 stand for qualified and unqualified wafers, respectively. A two-side test is implemented here since there is no prior knowledge of whether that μ^1_{jk} should be greater or smaller than μ^2_{jk}. This p-value can be used to determine whether variable j at time k is a good candidate for discriminating qualified and unqualified wafers.

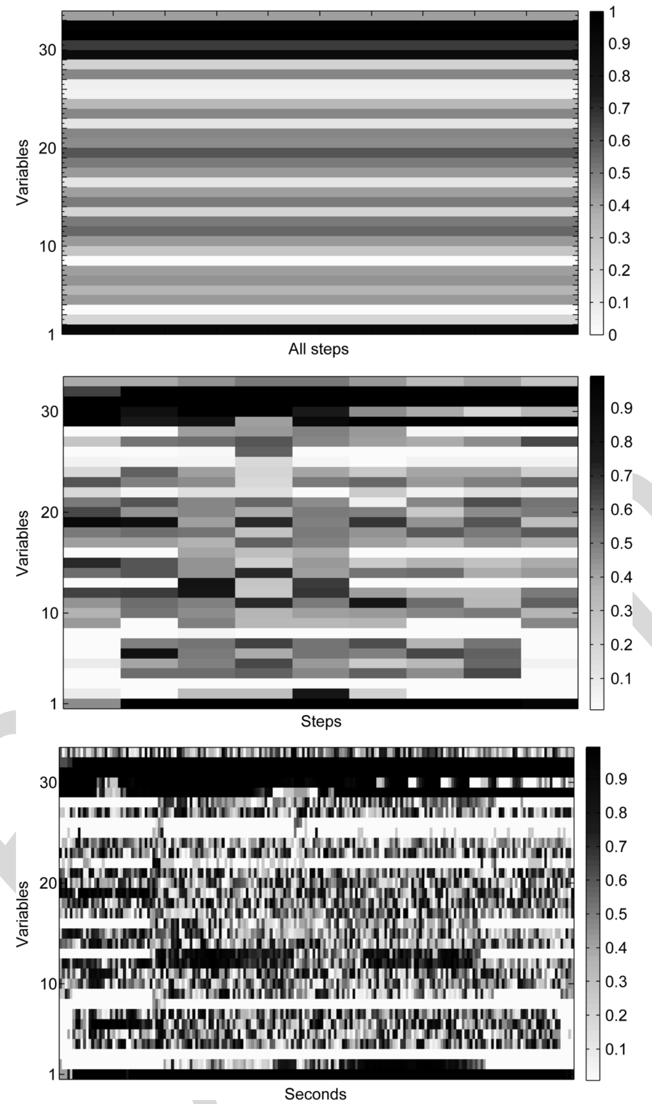


Fig. 2. Colored map synopsis images of different variables, different variables at different steps, and different variables at different time points.

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 two sets of data may not be normally distributed. The Wilcoxon rank-sum test is a nonparametric alternative which is based solely on the order in which the observations from the two samples fall. 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 rank-sum test statistic is the sum of the ranks for all observations from each set (qualified or unqualified) of samples [9].

Let p_{jk} be the p-value of the hypothesis (1) for the j^{th} variable at k^{th} time point. Note that the above analysis can be used

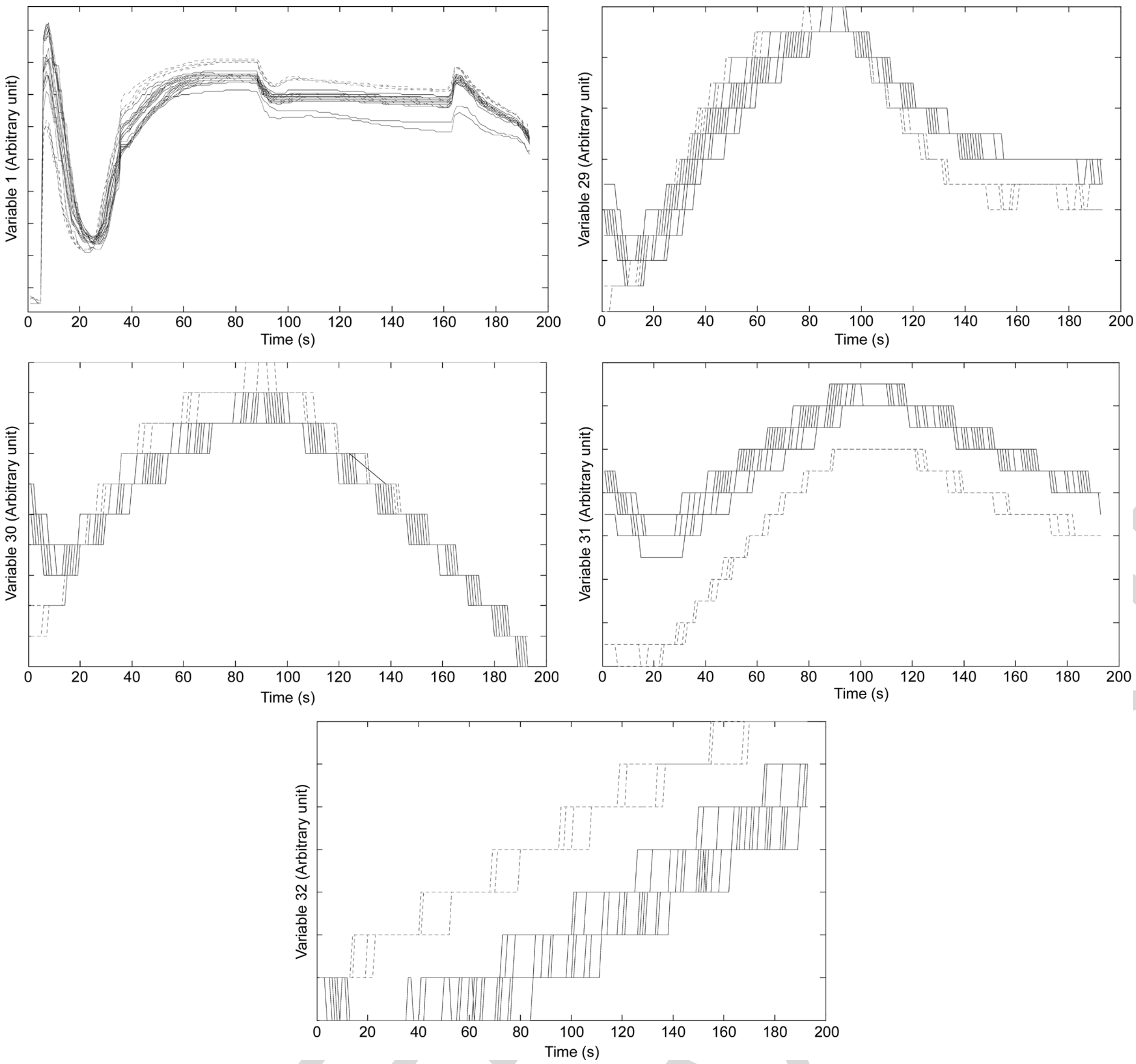


Fig. 3. Profiles of the top five influential process variables (solid line, qualified; dashed line, unqualified).

for different step s $(s = 1 \ldots S)$ of the process or the entire process by taking the appropriate averages of p_{jk}

$$P_{js} = \frac{\sum_{k=1}^{K} p_{jk}\delta_{sk}}{\sum_{k=1}^{K} \delta_{sk}}$$

$$\delta_{sk} = \begin{cases} 1, & \text{time point } k \text{ belongs to step } s \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

$$P_j = \frac{1}{K}\sum_{k=1}^{K} p_{jk} \tag{3}$$

where δ_{sk} is Kronecker delta. The results of above discrimination analysis can be easily visualized in a microarray-type expression shown in Fig. 2. Engineers can easily learn that variables 1, 29, 30, 31, and 32 are crucial to the wafer quality. It also can be learned from these figures which step or which time point is important for the final product quality. These are valuable information for engineers. The profiles of these five variables for qualified and unqualified wafers are plotted in Fig. 3.

The key variables found by statistical analysis are dome temperatures and wafer temperature. Fluctuation in dome temperature may generate particles that deposit on the wafer and may result in nonuniformity and lack of repeatability of a deposited film of material. In addition, variation in temperature over regions of the dome may result in excessive mechanical stress that can ultimately result in dome fracture. The result of the above statistical analysis is tested and confirmed by operating engineers.

III. Variation Reduction for Multistage CVD Process

Consider a chemical vapor deposition CVD process in which every wafer must be processed in 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 can be either film thickness measurements or wafer electrical measurements. There are 13, 5, 12 steps and

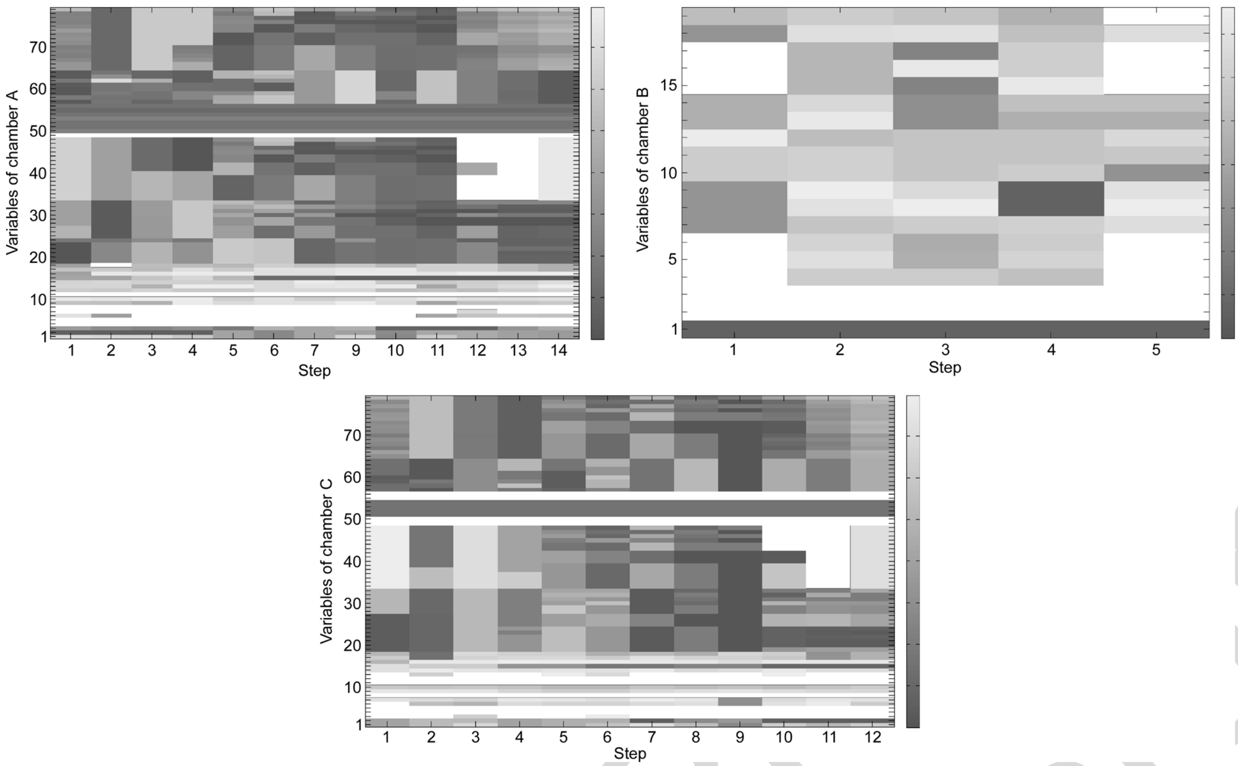


Fig. 4. Grouping of feature variables for the three chambers.

TABLE I
VARIABLES IN GROUP 8

Group 8	Total R^2 = 0.36113			
	Individual R^2	Correlation	Variable 1	Variable 2
	0.31203	Variable 1	1	0.97619
	0.35208	Variable 2	0.97619	1

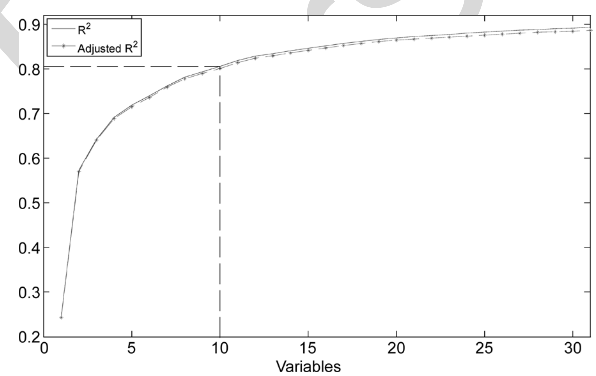


Fig. 5. Changes of R^2 and adjusted R^2 of stepwise regression.

79, 19, and 79 process variables for chamber A, B and C, respectively. The data set includes measurements of 526 wafers from 22 batches. A feature variable, e.g., average, max, min, range, is provided for each variable in each step. In this analysis, we want to know which of these feature variables will have significant effect on final quality and how the specifications of these process variables can be tightened or changed to improve final product quality.

To pick out the most influential variables, the first step is to reduce the dimension of the original data set. Process variables are usually highly correlated because of physical and chemical principles governing the process operation. The correlation is assumed to be linear in the following analysis. Cluster analysis is a useful technique used for combining variables into groups or clusters such that each group or cluster is homogeneous with respect to certain characteristics; while each group should be different from other groups with respect to the same characteristics. 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

$$d_{ij} = 1 - |r_{ij}| \tag{4}$$

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 [1].

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. 4. In these figures, variables that are filled with the same color are of the same group. The blank areas or the white color

TABLE II
VARIABLES IN GROUP 1

Group 1	Total $R^2 = 0.6425$							
	Individual R^2	Corr	Variable 1	Variable 2	Variable 3	Variable 4	Variable 5	Variable 6
	0.033207	Variable 1	1	0.97299	0.974	0.99135	0.95736	0.97312
	0.063261	Variable 2	0.97299	1	0.96825	0.98674	0.95048	0.98961
	0.017806	Variable 3	0.974	0.96825	1	0.97822	0.97716	0.95183
	0.03761	Variable 4	0.99135	0.98674	0.97822	1	0.97323	0.98846
	0.0010343	Variable 5	0.95736	0.95048	0.97716	0.97323	1	0.93974
	0.085809	Variable 6	0.97312	0.98961	0.95183	0.98846	0.93974	1

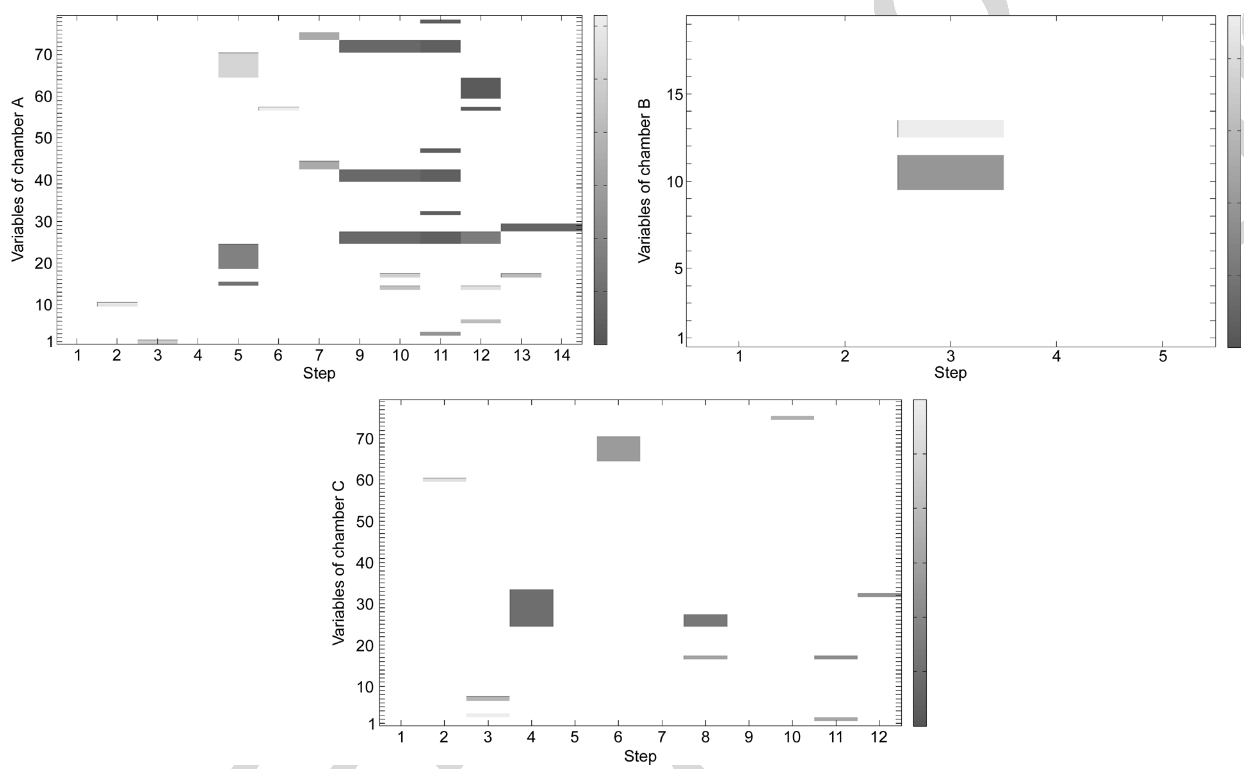


Fig. 6. The distributions of key clusters for the three chambers.

areas are caused by the result that some variables do not have observations at some steps.

The next step is to select representative variables from each group. The variables picked out should give good variance explanation of the quality index y which is usually evaluated by the R^2 statistics. For example, in Table I, the R^2 statistics of the variable 2 selected from group 8 is 0.352 and the total R^2 of the whole group is 0.361. In such a case, the variable 2 of group 8 is capable of representing the group. 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 which is the case of group 1 shown in Table II. 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 quality index and linear composites maximum. In this application, the following rules are adopted: if the R^2 of individual variable is more than 80% 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 until the R^2 of the linear composite is more than 80% of the total R^2 of all the variables in the group.

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

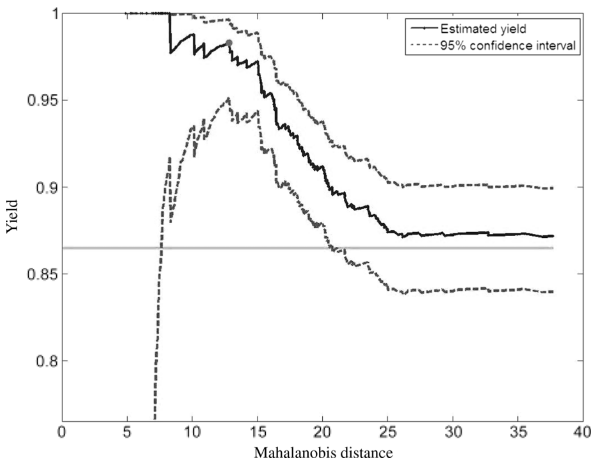


Fig. 7. Relations between the mean and 95% confidence estimates of yield and Mahalanobis distance from the center point.

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 [3]. The changes of R^2 and adjusted R^2 of stepwise regression are shown in Fig. 5. Thirty-one representative variables were selected by the stepwise regression algorithm. The distributions of the corresponding key clusters are shown in Fig. 6. In these figures, the clusters are labeled and colored by the rank of their order of selection. It is observed that most of the highly ranked variables are found in Chamber A.

In this preliminary experiment, all the 526 wafers are qualified wafers. To reduce the variance of quality characteristics, we define $[\overline{y} - 1.5s_y, \overline{y} + 1.5s_y]$ as the acceptable region for the wafer thickness. Here, $\overline{y}$ is the average value of y and s_y is the standard deviation of y, respectively. The wafers fall out of this region are treated as "unqualified" 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 selected key variables to improve the product yield. First, the center point for all the qualified wafers $\vec{\mu}$ in a space defined by the 31 important representative variables is determined. The Mahalanobis distance of each qualified wafer from the center point is calculated as

$$MD_i = (\vec{X}_i - \vec{\mu})^T \vec{S} (\vec{X}_i - \vec{\mu}) = c_i \tag{5}$$

where $\vec{X}$ is a 31×1 vector of coordinates and $\vec{S}$ is a 31×31 covariance matrix. Then, the yield can be viewed as an implicit function of the Mahalanobis distance. The distribution of this yield can be calculated by bootstrap sampling of the original wafers within the given Mahalanobis distance. A graphical interpretation of this relationship is shown in Fig. 7. In this figure, the solid line is the relationship between the mean of yield distribution estimated by bootstrap sampling and the Mahalanobis distance. The dashed lines are its 95% confidence intervals of the yield estimated by bootstrap sampling. The horizontal line is the original yield 0.865.

It is obvious that the estimated 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 the maximum of lower bound estimate of 95% confidence interval which is marked as a dot in Fig. 7 is used. The mean estimate is 98.3%. Once c_i is determined, the joint boundary of these variables is also determined. However, the joint boundary which is a function of 31 independent variables cannot be easily monitored online simultaneously. Therefore, the projections of the joint boundary onto the axes of coordinates are used as the new specifications of the five most influential variables. The new yield can be estimated again by bootstrap method. The results are listed in Table III. The new upper specification limit (USL) and lower specification limit (LSL) are presented in units to the original targets. It can be observed from Table III that the yield increased greatly when the upper and lower bounds of the first representative variable are designated. The increases of the yield are not obvious after the designation of the specification of the third representative variable. A graphical interpretation of the increase of the yield when the USL and LSL of key variable 1 are designated is shown in Fig. 8. The yield for the wafers fall into the regions defined by USL and LSL is about 95% and the yield for the wafers outside the USL and LSL is 54% which indicates reasonable specifications.

It is interesting to note that according to the engineers' experience, the manufacturing on chamber B is crucial for the final quality; however, the key variable 1 found by the statistical analysis is actual temperature measurement of the last step of chamber A. After a careful examination, it was found that the waiting time before a wafer was sent to chamber B is different. This made the temperature of wafers different when they were fabricated on chamber B. After the variation source was found, the final product quality was improved greatly.

TABLE III
THE USLS AND LSLS FOR THE TOP FIVE VARIABLES AND THE INCREASES OF YIELD

Key variables	USL	LSL	Yield
1	1.023	0.984	0.95323
2	1.003	0.991	0.98276
3	1.012	0.959	0.98465
4	1.002	0.783	0.98656
5	1.270	0.313	0.98652

(Note: The key variables could be single process variable or linear composite.)

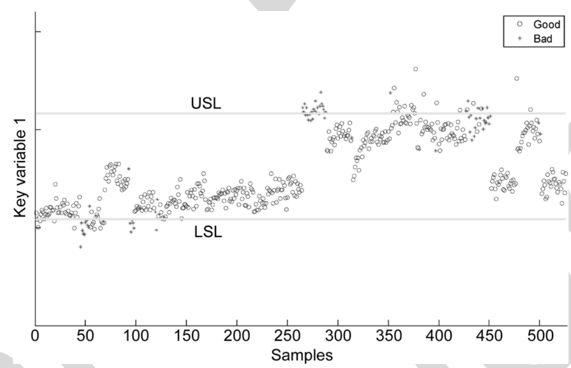


Fig. 8. USL and LSL for key variable 1.

IV. Conclusion

In the two given examples, we have presented a method of finding the key variables from a large set of variables recorded in a limited number of runs through a multistage multistep manufacturing process. It should be pointed out that although the statistical methods employed, e.g., Wilcoxon rank-sum test in the first example, clustering, regression and bootstrap estimates in the second are not new, however, the summary of analysis results, presented in a colored map of p-values which is very similar to a DNA microarray, allowed engineers to visualize them most intuitively. In an error of data explosion, it is important to develop industrial informatics tools that mine out important information from large amount of data and present them in succinct manners.

References

[1] S. Sharma, *Applied Multivariate Techniques*. New York: Wiley, 1996.

[2] B. E. Goodlin, D. S. Boning, H. H. Sawin, and B. M. Wise, "Simultaneous fault detection and classification for semiconductor manufacturing tools," *J. Electrochem. Soc.*, vol. 150, no. 12, pp. 778–784, 2003.

[3] D. C. Montgomery, E. A. Peck, and G. G. Vining, *Introduction to Linear Regression Analysis*, 3rd ed. New York: Wiley, 2001.

[4] B. M. Wise, N. B. Gallagher, S. W. Butler, D. D. White, Jr., and G. G. Barna, "A comparison of principal component analysis, multiway principal component analysis, trilinear decomposition and parallel factor analysis for fault detection in a semiconductor etch process," *J. Chemometrics*, vol. 13, pp. 379–396, 1999.

[5] H. H. Yue, S. J. Qin, R. J. Markle, C. Nauert, and M. Gatto, "Fault detection of plasma etchers using optical emission spectra," *IEEE Trans. Semiconduct. Manuf.*, vol. 13, no. 3, pp. 374–385, 2000.

[6] G. Spitzlsperger, C. Schmidt, G. Ernst, H. Strasser, and M. Speil, "Fault detection for a via etch process using adaptive multivariate methods," *IEEE Trans. Semiconduct. Manuf.*, vol. 18, no. 4, pp. 528–533, 2005.

[7] P. Baldi and S. Brunak, *Bioinformatics: The Machine Learning Approach*. Cambridge, MA: MIT Press, 2001.

[8] National Institute of Health, 2008, "DNA Microarry Technology." [Online]. Available: http://www.genome.gov/page.cfm?pageID=10000533

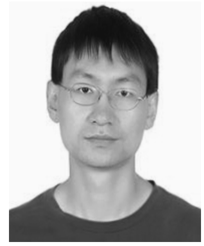
[9] J. J. Higgins, *Introduction to Modern Nonparametric Statistics*. Pacific Groove, CA: Brooks, 2004.

[10] G. P. McCabe, "Principal variables," *Technometrics*, vol. 26, no. 2, pp. 137–144, 1984.

[11] J. A. Cumming and D. A. Wooff, "Dimension reduction via principal variables," *Comput. Statist. Data Analy.*, vol. 52, no. 1, pp. 550–565, 2007.

[12] R. Srinivasan and M. S. Qian, "State-specific key variables for monitoring multi-state processes," *Chem. Eng. Res. Design*, vol. 85, no. 12, pp. 1630–1644, 2007.

[13] J. F. Cadima and I. T. Jolliffe, "Loadings and correlations in the interpretation of principal components," *J. Appl. Stat.*, vol. 22, no. 2, pp. 203–214, 1995.



Ming-Da Ma received the B.S. and M. S. degrees in control engineering from North China Electric Power University, Beijing, China, and the Ph.D. degree in control engineering from Shanghai Jiao Tong University, Shanghai, China, in 2000, 2003, and 2006, respectively.

He has been a Lecturer in the Center for Control and Guidance Technology, Harbin Institute of Technology, Harbin, China, since 2006. His research interests include advanced process control, process monitoring, statistical process control and robust control.

David Shan Hill Wong received the B.S. degree in chemical engineering from the California Institute of Technology, Pasadena, CA, in 1978 and the Ph.D. degree in chemical engineering from the University of Delaware, Newark, DE, in 1982.

He has been a faculty member of the Chemical Engineering Department, National Tsing-Hua University, Hsinchu, Taiwan, since 1983. His research interests in the area of process system engineering includes design and control of energy efficient separation processes and advanced process control in batch-based manufacturing. He is actively involved in consulting and collaborative research with chemical process and semiconductor manufacturing industries.

Shi-Shang Jang received the M.S. degree in chemical engineering from the National Taiwan University, Taipei, Taiwan, R.O.C., in 1980 and the Ph.D. degree in chemical engineering from Washington University, St. Louis, MO, in 1986.

Since 1992, he has been a Professor in the Chemical Engineering Department, National Tsing-Hua University, Hsinchu, Taiwan. He was the Chairman of the department from 2000 to 2004. He has been involved in many projects granted by Taiwan Semiconductor Corporation (TSMC) for the last five years. He is particularly interested in run-to-run control theory and practice.

Sheng-Tsaing Tseng (M'92) received the B.S. degree in business mathematics from Soochow University, the M. S. degree in applied mathematics from Tsing-Hua University, Hsinchu, Taiwan, and the Ph.D. degree in management science from Tamkang University, Taiwan.

Currently, he is a Professor at the Institute of Statistics, Tsing-Hua University. His research interests include quality and productivity improvement, reliability lifetime analysis, and statistical decision methodology. His articles have appeared in numerous technical journals, such as *Technometrics, Journal of Quality Technology*, the IEEE Transactions on Semiconductor Manufacturing, IEEE Transactions on Reliability, *IIE Transactions*, and etc.

Prof. Tseng received the Outstanding Research Award from the National Science Council (NSC) of R.O.C., Taiwan, in 1993, 1999, and 2004, respectively. Currently, he serves as an Associate Editor of *Journal of Statistical Planning and Inference* and he is an elected member of ISI.