8th International IFAC Symposium on
Dynamics and Control of Process Systems
Preprints Vol.1, June 6-8, 2007, Cancún, Mexico

PROCESS MONITORING USING KEY SENSITIVITY INDEX: APPLICATIONS TO SEMICONDUCTOR MANUFACTURING

Jyh-Cheng Jeng, An-Jhih Su, Cheng-Ching Yu†, H. P. Huang

Dept. of Chem. Eng., National Taiwan Universit, Taipei 106, TAIWAN

Abstract: Process monitoring is essential to maintain product quality in semiconductor manufacturing and the batch-wise processing nature makes data-based monitoring attractive. However, unlike chemical processes, the semiconductor manufacturing process exhibits the following characteristics: (1) much shorter (minutes) and often variable (deliberately adjusted) batch time, (2) multiple processing steps (10-20) in each batch, (3) only some particular processing steps (not the entire trajectory) constituting the quality determining steps, (4) mixed products for the same batch processing. In this work, instead of incorporating large number of trajectory data with variable batch time and possibly "missing" data for some process variables using multivariate statistics, a process-insight based approach, key sensitive index (KSI), is taken. From process knowledge, the key sensitive time-slot (KST) in the recipe is identified. Next, possible key sensitive process variables (KSV) are selected and validated according to the process trend correlation. Then, an index for these variables (key sensitive index, KSI) is sought. The KSI's can be classified into: variable-based, model-based, and data-based measures. Two integrated circuit processing examples from real fab data are used to illustrate the KSI-based approach and results clearly indicate that process trend is captured using KSI-based approach.

Keywords: MPCA, PCA, process monitoring, fault detection, batch process, semiconductor manufacturing

1. INTRODUCTION

Batch processes play an important role in the production and processing of chemicals, pharmaceutical, and semiconductor devices. Generally, a batch process is characterized by prescribed processing of raw materials for a finite duration to convert them to products. A high degree of reproducibility is necessary to obtain successful batches. Some batch processes include a single step, whereas many others are carried out in a sequence of discrete steps, which are usually referred as recipes in semiconductor manufacturing. Events taking place in each step have impacts on the final product yield and quality. For chemical process industry (CPI), upon completion of a batch, a range of quality measurements is usually made at the quality control laboratory, often hours later. For semiconductor industry, the quality measurements are usually not available at the end of a single processing unit until the processes in several processing units have been completed. Therefore, monitoring and control of the intermediate stages of operation and intermediate product quality is as important as monitoring and control at the final stage.

Online process performance monitoring and product quality prediction in real time can reduce quality variations. Multivariate statistical projection methods such as principal component analysis (PCA) and partial least squares (PLS) are capable of utilizing massive amounts of process data and compress the information data down into a lower dimensional latent space in which process monitoring and results interpreting are much easier. These methods are first utilized in developing multivariate statistical process monitoring

† to whom all correspondence should be addressed.
E-mail: ccyu@ntu.edu.tw
Fax: +886-2-2362-3040

and quality prediction techniques for continuous processes (Kourti et al., 1996). Nomikos and MacGregor (1994) pioneer the use of the large number of trajectory measurements collected from batch process and develop multi-way projection methods for statistical process monitoring. Batch data are recorded in terms of batch runs, variables, and time. They are arranged into a three-dimensional array in the multi-way projection methods and, therefore, analysis of batch processes data arouses a variety of challenges. Due to the nature of batch processes, every batch may come to completion at a different termination time, resulting in unequal batch data lengths. The unequal length batch data should be equalized prior to forming a three-dimensional array. The simplest way for equalizing batch lengths is cutting batch data lengths to the length of the variable with the shortest sequence, but this is not recommended because of the possible significant information loss generated by discarding data. Several methods for equalizing batch lengths, such as the indicator variable technique (IVT) (Nomikos and MacGregor, 1995; Neogi and Schlags, 1998), dynamic time warping (DTW) (Kassidas et al., 1998), and curve registration (Williams and Cinar, 2000), have been suggested in the literature. IVT is based on selecting a process variable to indicate the progress of the batch instead of the time axis. This variable should be chosen such that it progresses monotonically in time and has the same starting and ending value for each batch. But IVT does not account for the locations and the alignment of critical local features (landmarks) of each trajectory. DTW is used to synchronize two trajectories by appropriately translating, expanding, and contracting localized segments within the trajectories to achieve a minimum distance between the two trajectories. However, warping the batch trajectories to have the minimum distance with the reference batch trajectory may distort their critical features or fault patterns and, hence, reduce the monitoring ability. Curve registration is a two-step process of identifying landmarks within a trajectory (or set of trajectories) and then warping the test trajectory to the reference trajectory containing reference landmarks. Identifying multivariate landmarks is challenging because the number and location of landmarks may be different for different process variables. In addition to unequal batch data lengths, another difficulty for implementing multi-way projection methods is that the process variables are not necessarily recorded at regular intervals. Thus, interpolation is often needed to find measurements at regular time instances. This has to be done properly; otherwise, the auto-correlation and cross-correlation might be seriously affected (Kourti, 2003).

Multi-way PCA (MPCA) has been successfully applied for batch process analyses and monitoring in chemical process industry (e.g. Nomikos and MacGregor, 1995, MacGregor and Kourti, 1995). However, implementing multi-way PCA in the field of semiconductor manufacturing will encounter some difficulties. The main reason is that, unlike chemical processes, the IC processing has much shorter (seconds-minutes for a wafer as opposed to minutes-hours for chemicals) time duration. Preprocessing and arrangement of process data to deal with the problem of unequal batch data lengths using aforementioned methods will impose a relatively high computational burden and may significantly distort the critical features in the batch trajectories. Moreover, not the entire batch trajectory, but some particular processing sequence constitutes the quality determining steps, and this is especially true for crystallization processes in CPI and thermal processes in IC industries. Using the entire batch trajectory for analyses might cloud the relationship between variables in these critical steps and the final product quality and, subsequently, reduces the resolution in process trend monitoring. These characteristics make the data-based approaches difficult in semiconductor manufacturing industry.

In this work, an alternative for batch process monitoring is sought. Instead of incorporating large number of trajectory data with variable batch time and possibly "missing" data for some process variables using MPCA, a key sensitive index (KSI) based approach is proposed. From process insight or the experience of a process operator, a certain period time within a batch where measurements have significant effects on product quality, the key sensitive time-slot (KST), is identified. Next, based on the KST, possible key sensitive process variables (KSV) are chosen. The KSV may not be measured values themselves in KST, but some quantity, such as area, slope, maximum, etc., computed from the raw measurements. Once a KSV is computed from each batch (wafer-to-wafer) under normal operation, its autocorrelation function is calculated as the batch process progresses. If significant autocorrelation is found, a time-series model could be established from the selected KSV; if not, a different KSV is sought. With the time-series model, the process trend can thus be forecasted and the index for process operating status (key sensitive index, KSI) is defined and computed. By monitoring the KSI, possible maintenance action can therefore be called for, whenever necessary. This provides dynamical capability for process trend monitoring while maintaining the simplicity of single-variate analyses, if applicable. Two IC processing examples are used to illustrate the KSI-based approach.

2. KEY SENSITIVE TIME-SLOT AND KEY SENSITIVE VARIABLES

2.1 Key sensitive time-slot (KST).

In one process tool, there are many measurements, such as temperature, pressure, composition, flow rate, RF power etc., which will be acquired into database systems. Many engineers elaborate on mining useful knowledge or information from this huge amount of data. Figure 1 shows a simplified version of typical process raw data in semiconductor manufacturing. Because of the batch process characteristic, the techniques of process/equipment monitoring usually treats the data as three dimensions: batches, measurements, and time. Consequently, the multi-way PCA could be applied onto expanded 2-D data (Qin et al., 2005). When fetching raw data of a process tool, there are 10-90 available measured variables (depend on equipment). Many variables could be excluded because of the constant behavior or little influence on

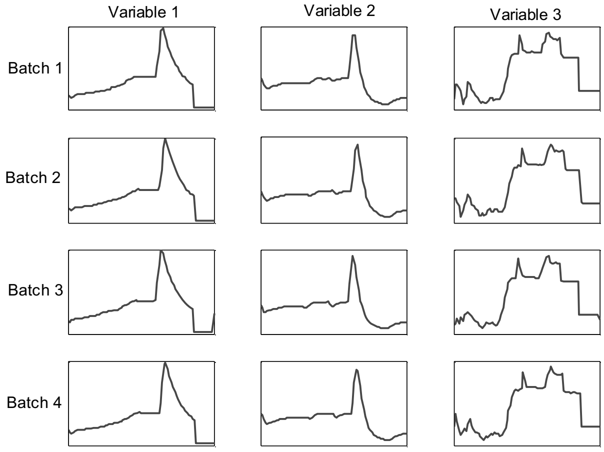


Fig. 1. Sketch of general process raw data in semiconductor manufacturing.

qualities from engineer's experience. It is the first step for dimension reduction and also the most important step to earn successful data analyses.

Despite of many successful applications of MPCA, several factors may impose difficulties in applications. First, the step length of each batch might be different. It results from that the operation time between batches is not coincident and this is quite true in tool level operation in IC processing. Second, the sampling interval may be different, and it leads to asynchronous data sequences. Third, it always encounters a problem about missing data, outlier, or other miscellanea when one does the data analysis. These difficulties naturally require substantial engineering effort on data pre-processing, to satisfy the argument format of modeling algorithms. In the worst case scenario, with excessive engineering effort, it may be not easy to restore the true system behavior.

After the selection of important measured variables, the next step is to cut each operation steps and filter them out. In practice, there are also many operation steps, typically 10-20, in a recipe (i.e., the entire batch run). Some steps are considered as pre-treatment, e.g., warming-up, introducing gas flows, etc., until the chamber condition is stabilized for subsequent processing. These kinds of minor steps could be neglected. The key sensitive time-slot is defined as the steps that have significant influence on quality or healthy condition such as the thermal budget period in RTP or particles-generation related steps in etching processes. For example, there are 6 steps in Fig.2, but only 3 critical steps, B, D, and F which are key sensitive time-slots. Usually, KSTs are the longest steps within a recipe. The minor steps, such as steps A, C, and E could be treated as transition steps. Moreover, experienced engineers can identify critical steps with little difficulty.

2.2 Key sensitive variable (KSV)

After the description of the KST, we introduce the concept of key sensitive variable (KSV) which is evaluated in the KST's. The KSV means certain features of process variables, and these features may have physical meanings which will affect product

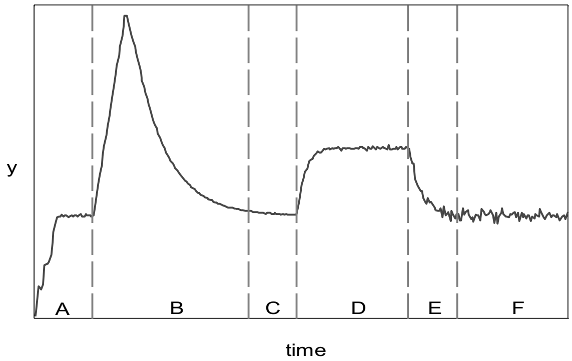


Fig. 2. Schematic of key sensitive time-slot (KST).

specifications. The KSV can be classified as: variable-based KSV and feature-based KSV.

1. Variable-based KSV: This is rather straightforward that the KSV is the process variable itself and it can be directly read from raw data. The KSV can be the variable at certain snapshot such as the initial value, final value, maximum value, and minimum value etc. For example, the peak value in the step B of Fig.2 may be used as a KSV. One should note that raw data may not precisely capture true behavior with a low sampling frequency or fast response processes.
2. Computed feature-based KSV: Some computed values from raw measurement, such as slope, mean, standard deviation, may have important physical implications. Again, let us use the step F in Fig.2 to illustrate this. The mean value or standard deviation can be a sensitive measure for process trend. Another possible example is the area of the peak or the ramp up slope in Step B (Fig. 2). This type of KSV, to a degree, reflects some physical meaning, e.g., integrated area for effective energy input.
3. Regressed feature-based KSV: Parameters of a prescribed model can be used extract process feature. These quantities can describe process condition as batch progresses. Simple models, steady-state polynomial model or dynamical transfer function model. For example, one could use initial value, gain, and time constant to capture the behavior of a first order response at the current stage (step D in Fig.2). This is effective for degrading process dynamics.

Thus a time sequence has been reduced several quantities. The second and third categories, however, may not suitable for on-line monitoring if complicated calculation is attempted.

Besides, any scalar value evaluated based on some concept could be a KSV. Actually, the fundamental of the KST and KSV is doing the dimension reduction for consequent analyses. For one batch, the two dimensional raw data matrix (the number of observation by the number of variables) has been reduced to many KSVs. Thus, it is easier for applying modeling of relationship with several qualities or fault detection.

2.2.2 Selection of the KSV

After obtaining many KSV candidates as possible, the even more reduction is possible. It is not guaranteed that all KSVs will have any relationship with qualities or healthy condition, although the KSV is evaluated from some features. A KSV candidate is a batch-wise sequence data. To realize the process trend, the instinctive way to check this subject is applying the autocorrelation analysis on all KSV candidates. The most satisfactory estimate of the kth lag autocorrelation of a time series is (Box et al., 1994):

$$r_k = \frac{c_k}{c_0} \tag{1}$$

where

$$c_k = \frac{1}{N}\sum_{t=1}^{N-k}\left(z_t - \bar{z}\right)\left(z_{t+k} - \bar{z}\right) \tag{2}$$

(k=0, 1, 2…., k) is the estimate of the autocovariance, and $\bar{z}$ is the sample mean of the time series.

If the autocorrelation function for a KSV candidate is found significant, it is then selected as KSV. Otherwise, if a candidate has weak correlation with itself, it can be treated as a white noise sequence while the equipment has been continuously operated for many batches. Thus, a part of KSV candidates can be removed. If, unfortunately, no one has significant expression, one should rediscover KSVs or even KST. Obtaining successful KSVs depends on the understanding and experience on process/equipment.

2.2.3 Modification on KSV via product index

In a foundry fab, the mix-product effect is frequently encountered. The modification on KSV according to product index is possible, so that a KSV can be applied over different product types. The product index could be linewidth, pattern density, layer material properties, and etc. If a KSV has unusual behaviors such as many shifts or spikes within a trend, usually it results from different product effects. In addition, one also can verify this issue by the analysis of the correlation between KSV and product index. To know the selected KSV having product effects is a challenging task, but it still can be figured out by some physical knowledge. For example, the removal rate of CMP will differ in with linewidth (Chiu et al, 2004), and the geometry of a wafer surface will influent the view factor of radiation related processes.

2.2.4 Generalized procedure

The generalized procedure for obtaining KSVs is given as follow:

S1: Select possible measurements based on experience.
S2: Define the key sensitive time-slots in a recipe.
S3: Obtain possible key sensitive variables.
S4: Check autocorrelation for each KSV. If no significant KSV is found, return to step 2 or 3.
S5: Take out the product effects, if information on product type and corresponding is available.

After this procedure, raw data of a batch process has been reduced into a vector in which each element is a KSV. Therefore, the data-drive analyses such as PCA, PLS and ANN could be applied straightly. In following section, we will propose a method to construct a time series based healthy index from these KSVs.

3. KEY SENSITIVE INDEX

In the case of extreme dimension reduction, a scalar index could be constructed as a function of many KSVs for one batch. This index should reveal the current status or critical feature of a batch trajectory, and it is very convenient for monitoring. This kind of index is termed as key sensitive indices (KSI) here which is used to describe the behavior of process. The choice of KSI may depend on the analysis done on the KSVs aforementioned. The possible ways for constructing KSI are described, but not limited to, as the following.

Variable-based KSI. The KSI is the KSV itself. If only one KSV is found and it exhibits significant autocorrelation, the simplest way is to take this KSV as KSI. The upper and lower limits for the KSI can be established to monitoring the batch process. In addition, the batch process trend can be realized by inspecting the KSI with batch process progressing.

Model-based KSI. The KSI is the prediction residual from time series models of KSVs. For a KSV with significant autocorrelation, it is suggested to identify time series models of this KSV for forecasting its future behavior. When the time series model is built using normal operating data, the difference between its predicted value and actual KSV of a batch can serve as the KSI. A large KSI (i.e. residual) means the process to deviate from its normal operating condition significantly. Moreover, we can identify different kind of time series models (e.g. ARMA and ARIMA) and use them for prediction to check the dynamic behavior (e.g. drifting) of the batch process progressing.

Data-based KSI. The KSI is a statistic (e.g. T^2, Q) resulted from model of PCA on KSVs. In some cases, no significant autocorrelation can be found in any possible KSVs computed. Since a batch trajectory is represented by several KSVs, multivariate projection methods such as PCA can be applied on these KSVs for batch process monitoring. Thus, the KSI could be the statistics, T^2 or Q, frequently used in PCA monitoring. Such KSI not only can be used for batch process monitoring, but also is suitable for fault diagnoses because it is derived based on PCA.

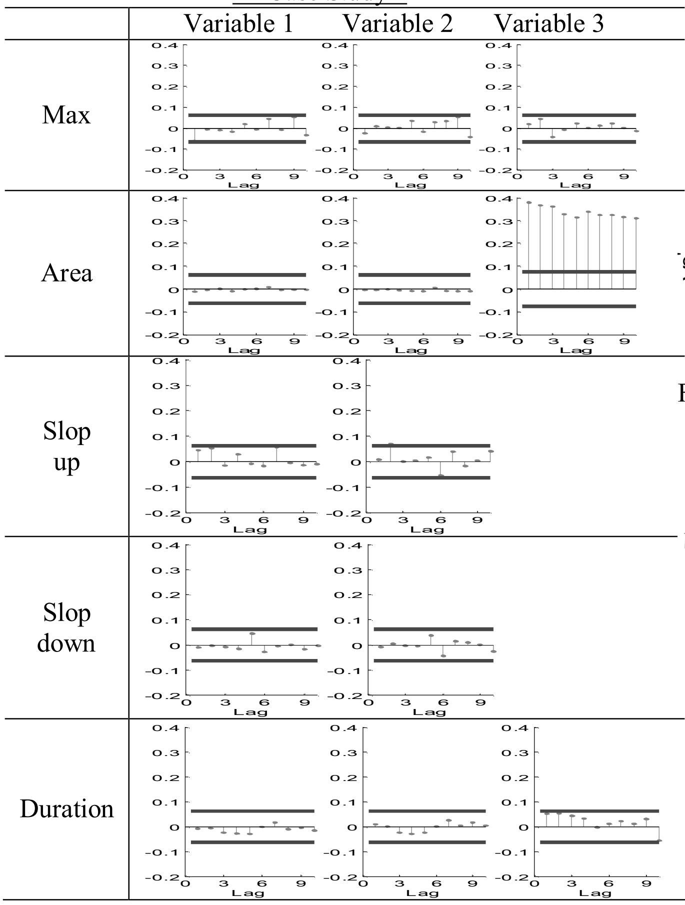
4. CASE STUDY

4.1 Case Study 1- Thermal Process

Filtering KSVs with Product Effects. Here comes a case study for illustration. A PM index is desired for a process tool, and the engineer provides three most important measurements. The autocorrelation analysis of possible KSV candidates is summarized in Table 1. The variable 3 does not have features of slopes, and has the only one significant autocorrelation on maximum value which is termed as $V3_{max}$ as plotted in Fig. 3 (a). As this tool has been proceed with almost five thousand wafers, this window size covers two completed PM cycles which are respectively located at vertical dash lines. Obviously, this KSV follows the PM behavior, because of abrupt shifts just only behind each PM.

The appearance of many spikes in $V3_{max}$ seems not appropriate for any sequential application. The first

Table 1 Autocorrelation analysis of possible KSVs in Case Study 1



instinctive reason resulted in these spikes comes out with that there are many different product in a foundry fab. The available information describing different product has been collected as product index 1, 2 and 3 (PI_1, PI_2, and PI_3). Consequently, we check the correlation coefficient between product indices and $V3_{max}$. The result indicates that the combination of PI_1 and PI_2, PI_{1+2}, has the strongest relationship with $V3_{max}$. In Fig.4, the spikes of $V3_{max}$ occur when the product index has higher values. Based on this observation, we do the some modifications including the least square regression, in which, if PI_{1+2} is beyond a threshold value, $V3_{max}$ will be reduced a values: $K{\cdot}PI_{1+2}$. The threshold value and parameter K are obtained in least square regression. The modified value, $V3'_{max}$, is demonstrated in Fig.3 (b). Therefore, this KSV is much clearer to indicate the PM behavior by taking out product effects.

A Time Series Based KSI. Roughly, $V3'_{max}$ can be treated as a monitoring index with given upper and lower control limits. However, if the tool's behavior (e.g., health index) can be captured, it is more reliable to make a decision of PM at which timing. The time series analysis (Box et al., 1994) is helpful for modeling a discrete sequence. An autoregressive moving average (ARMA) model of the following is built for $V3'_{max}$ based on measurements from 500 wafers.

$$\begin{aligned}&\left(1-1.744\,q^{-1}+0.776\,q^{-2}\right)V3'_{\max}(t)\\&=\left(1-1.346\,q^{-1}+0.476\,q^{-2}\right)e(t)\end{aligned} \tag{3}$$

where q^{-1} is the backward shift operator and $e(t)$ is white noise. It is found that one root of the

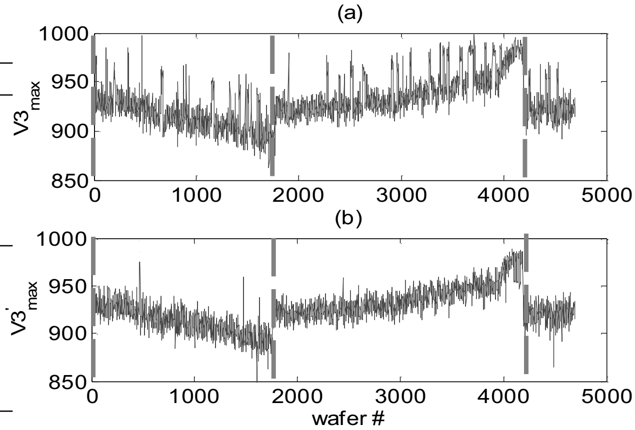


Fig. 3. The Trend of (a) KSV and (b) modified KSV in case study 1

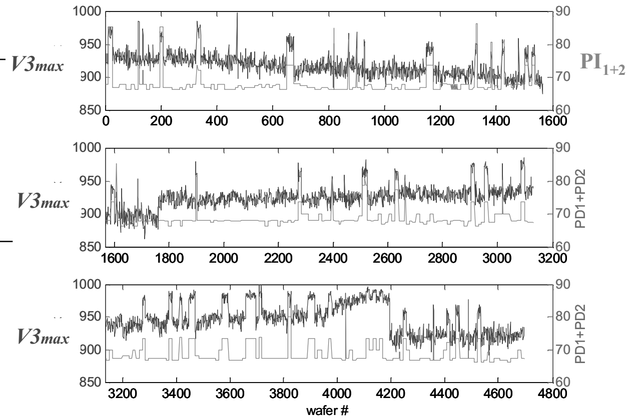


Fig. 4. The influence of product index in case study 1

autoregressive polynomial is close to unity, which means that the time series of $V3'_{max}$ exhibits nonstationary behavior. For this reason, an autoregressive integrated moving average (ARIMA) model is then built as the following to describe this behavior.

$$\begin{aligned}&\left(1+0.942\,q^{-1}\right)\nabla V3'_{\max}(t)\\&=\left(1+0.452\,q^{-1}-0.553\,q^{-2}\right)e(t)\end{aligned} \tag{4}$$

where $\nabla=(1-q^{-1})$. These two time-series models are then been used for forecasting the values of $V3'_{max}$ as the batch process progresses. Initially, both the forecasts of ARMA and ARIMA models can follow the process trend well. However, as the batch process progresses, the forecast of ARMA model starts to deviate from the actual $V3'_{max}$ more and more, while the forecast of ARIMA model keeps following. This phenomenon disappears after PM and then can be observed again as the batch process progresses. In order to capture the drifting behavior of this batch process, the KSI is thus defined as the absolute value of difference between residuals of these two models.

$$KSI = |\ \text{Residual}_{\text{ARMA}} - \text{Residual}_{\text{ARIMA}}\ | \tag{5}$$

The computed KSI is shown in Fig. 5. The results clearly indicate that the process trend can be realized using the proposed KSI and tool maintenance is required once this KSI is greater than a prescribed limit. Therefore, this KSI-based approach not only can be used for batch process trend monitoring, but also it is

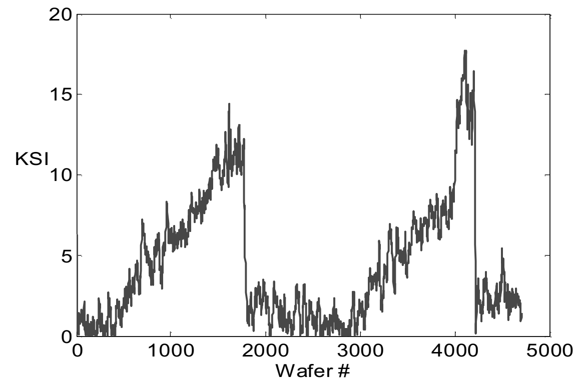


Fig. 5. The resultant key sensitive index (KSI) for case study 1

helpful for the engineers to decide when to call for tool maintenance.

4.2 Case Study 2- Thin Film Process

The concept of KSV is also been applied to another thin film process tool. There are three issues in this case study. One is virtual metrology which relates tool raw data and quality index, and the other two are fault detection and process trend monitoring.

The engineers provide totally 23 variables raw data. After filtering out some constant, meaningless, and poor resolution variables, we attempt to obtain KSVs from two key sensitive time-slots. Because the data of quality is given, we search many numbers of KSV to match quality as possible. Applying the least square model obtained from the training set to predict the test set, the accuracy is R^2=0.65 in which 41 KSVs are used. Although, this result may not be outstanding, it's value comes from that the simple concept of KSV and simple least square modeling can hit the greater part of quality behavior.

Besides, this process tool may produce defects that will no be detected until several sequential process steps have been proceed. A period of tool raw data that contains several scrapped wafers is provided. We obtain 16 KSVs, including wafer temperature, chamber side temperature, and other variables. A PCA model is built based on normal (no defects) wafers, and consequently use *Q* statistic for detection as shown in Fig. 6. At the end of *Q* line, the last 4 wafers soar abnormally, and these wafers have also been classified as scrapped wafers. For diagnosis, the contribution analysis shows several significant weighting KSVs which are obtained from chamber temperature. According to the process engineer, there was a leakage on a gas tube at that time, and it resulted in chamber side temperature increasing abnormally. It is coincident with our contribution analysis. Hence, the concept also can be applied to fault detection and diagnosis successfully.

Finally, for process trend monitoring, we also use 16 KSVs as described previously. This KSI grows up with the progress of batch process and return to its normal value immediately after PM. The results again indicate that the process trend and the healthy condition of the tool can be realized using the proposed KSI.

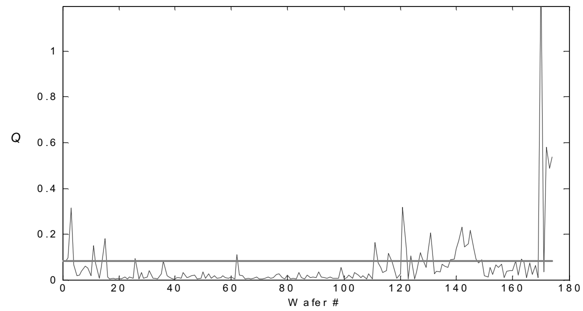


Fig. 6. Fault detection of case study 2

5. CONCLUSION

The concept of using KSV to construct a KSI is easy and straightforward, but with great benefit. It captures the key features of measurements for consequential analyses without complicated data pre-processing. We have demonstrated two successful real cases by proposed method. On the contrary, simplification also could possible lose some important information. The understanding of processes and equipment make KSV or other analysis technique working out.

REFERENCES

Box, G. E. P., G. M. Jenkins, and G. C. Reinsel (1994), *Time Series Analysis: Forecasting and Contrlo*, Prentice Hall

Chiu, J.B., A.J. Su, C.C. Yu, and S.H. Shen (2004), Planarization Strategy of Cu CMPInteraction between Plated Copper Thickness and Removal Rate, *Journal of The Electrochemical Society*, **151** (4), G217-G222

Kassidas, A., J. F. MacGregor and P. A. Taylor (1998), Synchronization of Batch Trajectories Using Dynamic Time Warping, *AIChE J.*, **44**, 846.

Kourti, T. (2003), Multivariate Dynamic Data Modeling for Analysis and Statistical Process Control of Batch Processes, *J. Chemom.*, **17**, 93.

Kourti, T., J. Lee and J. F. MacGregor (1996), Experience with Industrial Applications of Projection methods for Multivariate Statistical Process Control, *Computers Chem. Engng.*, **20**, S745-S750.

MacGregor, J. F. and T. Kourti (1995), Statistical Process Control of Multivariate Processes, *Control Engineering Practice*, **3**, 403-414.

Neogi, D. and C. E. Schlags (1998), Multivariate Statistical Analysis of an Emulsion Batch Process, *Ind. Eng. Chem. Res.*, **37**, 3971.

Nomikos, P. and J. F. MacGregor (1994), Monitoring of Batch Processes Using Multi-way Principle Component Analysis, *AIChE J.*, **40**, 1361.

Nomikos, P. and J. F. MacGregor (1995), Multivariate SPC Charts for Monitoring Batch Processes, *Technometrics*, **37**, 41-59.

Qin, S. J., G. Cherry, R. Good, J. Wang, and C. A. Harrison (2006), Semiconductor manufacturing process control and monitoring: A fab-wide framework, *J. of Process Control,* **16**, pp. 179-191

Williams, B. A. and A. Cinar (2000), An On-line Monitoring and Curve Registration System for Multivariable Batch Processes, *AIChE Annual Meeting*, Los Angeles, CA.