View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Kyoto University Research Information Repository

Kyoto University Research Information Repository

KYOTO UNIVERSITY

Title	Adaptive Virtual Metrology Design for Semiconductor Dry Etching Process Through Locally Weighted Partial Least Squares
Author(s)	Hirai, Toshiya; Kano, Manabu
Citation	IEEE Transactions on Semiconductor Manufacturing (2015), 28(2): 137-144
Issue Date	2015-05
URL	http://hdl.handle.net/2433/201404
Right	© 2015 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.
Type	Journal Article
Textversion	author

Kyoto University

Adaptive Virtual Metrology Design for Semiconductor Dry Etching Process through Locally Weighted Partial Least Squares

Toshiya Hirai and Manabu Kano, *Member, IEEE*

***Abstract*—In semiconductor manufacturing processes, virtual metrology (VM) has been investigated as a promising tool to predict important characteristics of products. Although partial least squares (PLS) is a well-known modeling technique that can cope with collinearity and therefore applied to construction of VM, its prediction performance deteriorates due to changes in process characteristics. In particular, maintenance of equipment strongly affects the process characteristics and the prediction performance. In the present work, VM was developed by using locally weighted PLS (LW-PLS), which is a type of Just-In-Time modeling technique, and it was used to predict the etching conversion differential of an actual dry etching process. The industrial application results have shown that the developed VM based on LW-PLS is superior to the conventional VM based on the sequential update model (SUM) and the artificial neural network (ANN) model. In particular, it has been confirmed that the LW-PLS-based VM can keep its high prediction performance even after the maintenance, i.e. replacement of parts.**

***Index Terms*—equipment engineering system, just-in-time modeling, locally weighted regression, semiconductor process, virtual metrology.**

I. Introduction

SEMICONDUCTORS are manufactured through more than 300 process steps. If an error is detected in the final electrical test (ETest), the cause is analyzed on the basis of the ETest result and process measurements. Since many semiconductor manufacturers execute quality control through sampling measurements, it is critical whether the target wafer is sampled or not. In recent dry etching processes, less than one wafer is sampled and measured when 25 wafers are processed in each lot; thus, cause analysis is difficult in many cases.

Due to increasing miniaturization of semiconductors, machining capacity of assuring product quality in manufacturing equipment is reaching its limit. To reduce the variability of wafer processing and assure product quality, wafer-to-wafer (W2W) level control is required. Consequently, all wafers need to be measured. However, the introduction of advanced process control (APC) based on measurements of all wafers requires costs of expensive metrology tools and longer production cycle time. Thus, virtual metrology (VM) has received much attention in order to predict important characteristics of products without installing additional metrology tools.

T. Hirai is with Sony Semiconductor Corporation. e-mail: Toshiya.Hirai@jp.sony.com

M. Kano is with the Department of Systems Science, Kyoto University, Kyoto 606-8501, Japan.

Manuscript received April 00, 2000; revised December 00, 2000.

Investigated in the present work is a dry etching tool, which is typical equipment drawing the circuit on wafers. Before each wafer is processed in a chamber of the dry etching tool, the chamber is evacuated, plasma is excited, and gases are flown in it. Monitoring the processing status of the dry etching is difficult because wafers are processed in the vacuum chambers. To visualize the vacuum process, optical emission spectroscopy (OES) technology has been developed. In addition, process data are accumulated and analyzed by an equipment engineering system (EES). Both OES and EES are useful to develop VM [1], [2]. However, prediction accuracy of VM decreases due to maintenance events or other external disturbances. In particular, replacement of parts strongly affects the process characteristics.

The present research aims to develop VM that can predict the etching conversion differential with high accuracy even when characteristics of the dry etching tool change significantly. Here, the etching conversion differential is defined as the difference between critical dimension (CD) at the end of a lithography process and CD at the end of the etching process. The design of VM for a dry etching process was investigated by Zeng and Spanos [3]; they proposed to adopt back-propagation neural networks (BPNNs) for modeling and compared its performance with several other techniques. To develop adaptive VM, locally weighted partial least squares (LW-PLS) is investigated in the present research. LW-PLS is a kind of just-in-time (JIT) modeling methods and has been successfully applied to pharmaceutical and petrochemical processes as an adaptive modeling technique that can cope with changes in process characteristics [4], [5], [6], [7]. LW-PLS and the conventional sequential update model (SUM) are applied to real industrial data of the dry etching tool, and their prediction performance is compared.

In the following section, EES and VM are introduced. Then, data preprocessing and modeling methods are explained in sections III and IV, respectively. The validation results using actual data are shown in section V, which is followed by the conclusion.

II. EES and VM

This section briefly introduces EES and VM, which play important role in semiconductor manufacturing processes.

A. Equipment Engineering System (EES)

In 2002, EES was proposed by Selete (Semiconductor Leading Edge Technologies) and others as a platform to pick

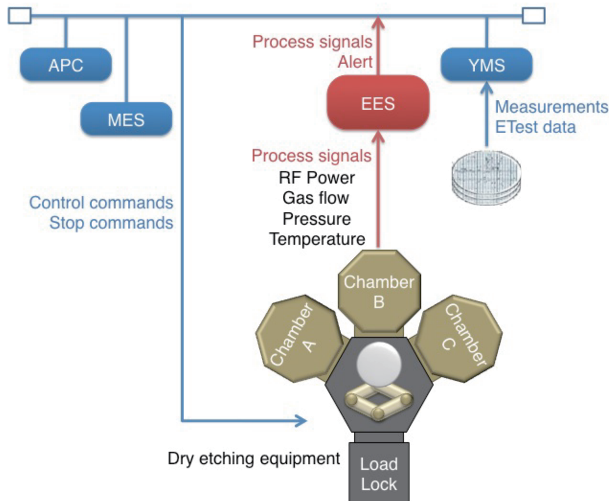

Fig. 1. Dry etching equipment controlled by EES-centered system. Equipment engineering system (EES), manufacturing execution system (MES), advanced process control (APC), and yield management system (YMS) are integrated.

up process signals and analyze data [8]. Today, equipment suppliers and system suppliers implement EES in dry etching equipment. Figure 1 shows the dry etching equipment monitored and controlled by an EES-centered system. Each arrow in this figure describes the flow direction of information or data. The equipment has multiple chambers. When a wafer is processed in the chamber, the state such as radio frequency (RF) power, gas flow, pressure, and temperature is sensed. EES collects and analyzes these process signals in real time to verify the equipment condition, and it notifies a manufacturing execution system (MES) if anomaly is detected. MES issues a stop command to the equipment when receiving the alert. The results of the analysis by EES are sent to APC, which calculates necessary changes of process parameters for each wafer and sends the parameters to MES; then MES issues a control command to the equipment. In addition, the process signals collected by EES are sent to a yield management system (YMS), which compares the signals, measurements, and ETest data.

B. Virtual Metrology (VM)

Virtual metrology (VM) is the prediction of metrology variables from process and wafer state information. VM plays an important role in the semiconductor manufacturing because the quality of all wafers, regardless of whether it is sampled or not, has to be controlled on the basis of equipment data. Such a prediction technique is called a soft-sensor or virtual sensing technology in other industries. Soft-sensors have been widely used for more than several decades to predict difficult-to-measure product quality or other key variables from easy-to-measure signals such as temperature, pressure, and flow rate. The history of soft-sensor development was summarized with several recent industrial applications by Kano and Fujiwara [4].

VM provides wafer level metrology data for wafer-to-wafer (W2W) level control without increasing the number of measured wafers [9]. W2W control requires timely measurement of the quality, which VM predicts on the basis of process signals collected by EES at the wafer level [10].

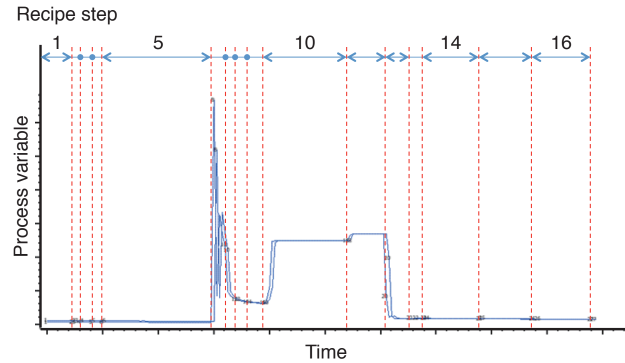

Fig. 2. Recipe steps of dry etching equipment. Wafers are processed in 16 recipe steps and about 400 kinds of signals are stored in EES.

In addition, a method of monitoring the whole factory by using VM has been proposed [11], [12]. A factory- or fab-wide supervisory control strategy coordinates individual run-to-run (R2R) process controllers utilizing the end-of-run or in situ metrology as well as the ETest data of the final product. The factory-wide controller could alter process target to achieve the factory-wide metrics such as overall factory yield and final ETest values [13]. Moreover, a VM model can be used to explain the variations in ETest data [14].

III. Data Preprocessing

To realize high-performance VM, appropriate preprocessing of measured data is crucial. In this section, derivation of representative variables used as input variables of VM, exclusion of outliers, and selection of input variables are explained.

A. Representative Variables

The dry etching tool is equipped with sensors to measure process variables such as RF power, gas flow, pressure, and temperature. These process signals are stored in EES together with OES signals at a high resolution of 100 msec or higher, and they are afterwards used to calculate representative variables.

In the dry etching tool investigated here, wafers are processed in 16 recipe steps as shown in Figure 2. About 400 kinds of signals stored in EES are converted into nine kinds of statistics: the mean, the median, the maximum, the minimum, the range, the standard deviation (StdDev), the integral, and the differential. These statistics are used as representative variables. Thus, the total number of representative variables becomes 57600, i.e. 400 signals $\times$ 16 recipe steps $\times$ 9 statistics for a single dry etching tool.

In the literature, monitoring of multiphase batch processes has been investigated [15]. In particular, an important issue is how to divide whole batch operation into several phases automatically on the basis of measured data so that better monitoring performance is achieved. In our application, however, such phase partition technique is not necessary because recipe steps (phases) are known in advance as shown in Fig. 2.

B. Outlier Exclusion

In EES, Hotelling's T^2 statistic [16] is used to detect outliers. The T^2 statistic is the scaled squared distance of a sample vector $\boldsymbol{x}$ from the center, where $\boldsymbol{x}$ is scaled, i.e. divided, by standard deviations in the direction of corresponding eigenvectors of the sample covariance matrix. The T^2 statistic indicates how far the sample is from the multivariate mean of the data.

It is important to distinguish between outliers and out-of-spec wafers. Outliers are caused mainly by failure of measurement; therefore outliers must be detected and properly treated before a statistical model is built or used. This preprocessing is crucial to develop a highly accurate model and to realize accurate and reliable estimation.

C. Input Variable Selection

Representative variables are used as input variables of VM. To improve prediction accuracy and reduce computational load, it is necessary to select important input variables that are related with the output. Although various input variable selection methods have been proposed [4], variable importance in the projection (VIP) [17], [18] was used for input variable selection in this work.

VIP is calculated when a PLS model is constructed. PLS is a linear regression method that can cope with the collinearity problem; thus it has been used as a modeling tool in various industries where process variables are highly correlated. PLS is now regularly employed in the area of semiconductor manufacturing [19]. In the following, PLS and then VIP are briefly explained.

1) Partial Least Squares (PLS): In PLS with one output variable, an input data matrix $\boldsymbol{X} \in \Re^{N \times M}$ and an output data vector $\boldsymbol{y} \in \Re^{N}$ are decomposed as follows:

$$\boldsymbol{X} = \boldsymbol{T}\boldsymbol{P}^T + \boldsymbol{E} \quad (1)$$

$$\boldsymbol{y} = \boldsymbol{T}\boldsymbol{b} + \boldsymbol{f} \quad (2)$$

where $\boldsymbol{T} \in \Re^{N \times R}$ is a latent variable matrix whose columns are latent variables $\boldsymbol{t}_r \in \Re^{N}$ $(r = 1, 2, \cdots, R)$, $\boldsymbol{P} \in \Re^{M \times R}$ is a loading matrix of $\boldsymbol{X}$ and its columns are loading vectors $\boldsymbol{p}_r$, $\boldsymbol{b} = [b_1, b_2, \cdots, b_R]^T$ is a loading vector of $\boldsymbol{y}$, and $\boldsymbol{E}$ and $\boldsymbol{f}$ are errors. N, M, and R denote the number of samples, that of input variables, and that of adopted latent variables, respectively.

The nonlinear iterative partial least squares (NIPALS) algorithm can be used to construct the PLS model [18]. Suppose that the first to $(r-1)$th latent variables $\boldsymbol{t}_1, \boldsymbol{t}_2, \cdots, \boldsymbol{t}_{r-1}$, the loading vectors $\boldsymbol{p}_1, \boldsymbol{p}_2, \cdots, \boldsymbol{p}_{r-1}$ and $b_1, b_2, \cdots, b_{r-1}$ are given. The rth residual input and output can be written as follows:

$$\boldsymbol{X}_r = \boldsymbol{X}_{r-1} - \boldsymbol{t}_{r-1}\boldsymbol{p}_{r-1}^T \quad (3)$$

$$\boldsymbol{y}_r = \boldsymbol{y}_{r-1} - b_{r-1}\boldsymbol{t}_{r-1} \quad (4)$$

where $\boldsymbol{X}_1 = \boldsymbol{X}$ and $\boldsymbol{y}_1 = \boldsymbol{y}$. The latent variable $\boldsymbol{t}_r$ is a linear combination of the columns of $\boldsymbol{X}_r$, that is, $\boldsymbol{t}_r = \boldsymbol{X}_r\boldsymbol{w}_r$ where $\boldsymbol{w}_r \in \Re^{M}$ is the rth weighting vector. PLS aims to maximize the covariance between $\boldsymbol{y}_r$ and $\boldsymbol{t}_r$ under the constraint $\|\boldsymbol{w}_r\| = 1$. By using the Lagrange multipliers method, $\boldsymbol{w}_r$ is derived as

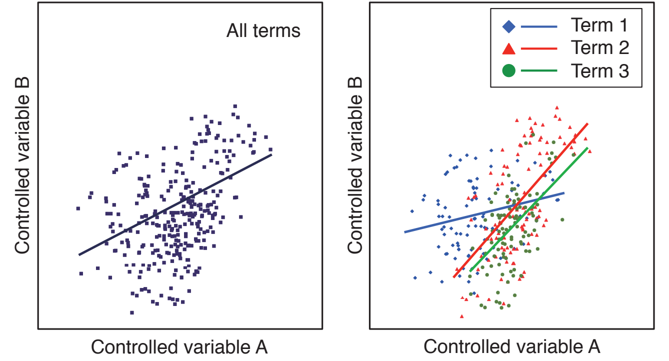

Fig. 3. Drift and shift of operation status of dry etching equipment.

$$\boldsymbol{w}_r = \frac{\boldsymbol{X}_r^T\boldsymbol{y}_r}{\|\boldsymbol{X}_r^T\boldsymbol{y}_r\|}. \quad (5)$$

The rth loading $\boldsymbol{p}_r$ and b_r are as follows:

$$\boldsymbol{p}_r = \frac{\boldsymbol{X}_r^T\boldsymbol{t}_r}{\boldsymbol{t}_r^T\boldsymbol{t}_r} \quad (6)$$

$$b_r = \frac{\boldsymbol{y}_r^T\boldsymbol{t}_r}{\boldsymbol{t}_r^T\boldsymbol{t}_r}. \quad (7)$$

Finally, the above procedure is repeated until the number of adopted latent variables R is achieved; R can be determined by cross validation.

2) Variable Importance in the Projection (VIP): VIP is also known as variable influence on projection. The VIP score of an input variable is a summary of the importance for the projection to find latent variables, expresses the contribution of the input variable to the output variable, and thus can be used for variable selection [17]. VIP is a weighted sum of squares of the PLS weights with the weights calculated from the amount of Y-variance of each PLS component [18]. The VIP score of the mth variable is defined as

$$V_m = \sqrt{\frac{M\sum_{r=1}^{R}\left(b_r^2\boldsymbol{t}_r^T\boldsymbol{t}_r(w_{mr}/\|\boldsymbol{w}_r\|)^2\right)}{\sum_{r=1}^{R}b_r^2\boldsymbol{t}_r^T\boldsymbol{t}_r}} \quad (8)$$

where w_{mr} is the mth element of the weighting vector of the rth latent variable $\boldsymbol{t}_r$. The variables satisfying $V_m > \mu$ should be selected as input variables. Since the average of squared VIP scores equals 1, $\mu = 1$ is suggested as a criterion for variable selection [20]. However, from the authors' experience, μ needs to be adjusted in order to improve the estimation performance [4].

IV. Prediction Method

Dry etching equipment needs to be maintained; for example, parts are replaced and chambers are cleaned repeatedly. Changes in equipment status caused by such maintenance deteriorate the prediction accuracy of VM. Figure 3 shows time-varying relationship between two controlled variables of the equipment. Even though the equipment is operated by using the same parameters determined in advance, the

equipment status suffers from drifts and shifts, which are caused by the replacement of parts and other factors. A drift of a variable is its continuous change in the same direction, and a shift of a variable is its large change between samples. Regardless of such changes in the equipment status, it is crucial to keep high prediction accuracy of VM.

The model maintenance is not a specific problem to VM, but it is a critical problem in various industries. According to the questionnaire survey of chemical process control in Japan [21], the most important problem of soft-sensors was how to cope with changes in process characteristics and maintain high prediction accuracy for a long period of time, i.e. model maintenance. When the prediction performance of a soft-sensor or VM becomes worse with time and, hence, laborious model reconstruction is frequently required, its applications cannot be widespread. The problem of model maintenance must be solved when VM is designed.

In the present work, VM is developed by using LW-PLS, which is one of just-in-time modeling methods. Since there are a number of signals that are mutually correlated in semiconductor manufacturing, PLS is used to build local models and to deal with collinearity.

A. Just-In-Time Modeling

Just-in-time (JIT) modeling can cope with changes in process characteristics as well as nonlinearity, and it has been used not only for virtual sensing [4] but also for process monitoring [22]. JIT modeling is also known as lazy learning [23], [24], [25], [26]. In JIT modeling, a local model is built from past data around a query, i.e. a new sample, only when an estimate is required. This approach is useful when a global model is difficult to build. Various JIT modeling methods were reviewed in [4].

In comparison with conventional modeling methods, JIT modeling has the following features. 1) When new input and output data become available, they are stored into a database. 2) A local model is constructed from samples located in the neighborhood around the query and output variables are estimated only when estimation is required. 3) The constructed local model is discarded after its use for estimation. In JIT modeling, samples for local modeling should be selected appropriately each time.

A simple approach to selecting neighbors of a query is k nearest neighbors (k-NN). Although k-NN works in practice, the prediction performance of JIT modeling can be improved by choosing neighbors in a more sophisticated way. Another simple approach to building a local model is a weighted average [27], which is preferable to other local modeling methods when the number of neighbors is limited. The prediction performance can be further improved by adopting local linear regression (LLR).

Instead of deciding to use or not to use each sample stored in a database, one can use all samples by prioritizing them according to their importance. The importance is evaluated by the similarity between the query and samples, e.g., the distance between the query and samples. The prioritizing strategy is more general because it is the same as deciding to use or not to use samples when the similarity is binary 0,1. Thus, JIT modeling is also known as locally weighted regression (LWR) [28]. LWR constructs a local model by prioritizing samples in a database according to the similarity between a query and the samples. Hence, LWR can cope with abrupt changes as well as gradual ones in contrast to recursive methods. In addition, it can cope with nonlinearity since it builds a piecewise model. LWR usually outperforms a typical weighted average method.

LWR has been successfully used to develop a quality design system for steel products [29]. The developed system based on LWR can predict the quality of various steel products and optimize operating conditions, and it has contributed to a reduction in quality deviations and production cost.

To build an accurate model with LWR, the similarity needs to be properly defined. In general, similarity is defined on the basis of the Euclidean distance or the Mahalanobis distance. Other similarity measures proposed so far are based on the distance and the angle [30], [31], [32], the distance and the output estimate derived by a global model [33], the correlation among input variables [34], [35], and the weighted Euclidean distance [5], [6], [29].

B. LW-PLS

LW-PLS investigated in the present work is an extension of LWR. LW-PLS uses PLS for local modeling while LWR uses multiple linear regression (MLR) [5]. LW-PLS is preferable to conventional LWR since robust local models can be constructed even when the number of neighbor samples is less than that of input variables. The basic algorithm of LW-PLS is explained below.

The nth sample ($n = 1, 2, \cdots, N$) of input and output variables is denoted by

$$\boldsymbol{x}_n = [x_{n1}, x_{n2}, \cdots, x_{nM}]^T \quad (9)$$

$$\boldsymbol{y}_n = [y_{n1}, y_{n2}, \cdots, y_{nL}]^T \quad (10)$$

where M and L are the numbers of input and output variables, respectively, and $\boldsymbol{X} \in \Re^{N \times M}$ and $\boldsymbol{Y} \in \Re^{N \times L}$ are the input and output variable matrices whose nth rows are $\boldsymbol{x}_n^T$ and $\boldsymbol{y}_n^T$, respectively.

In LW-PLS, $\boldsymbol{X}$ and $\boldsymbol{Y}$ are stored in a database. When an output estimation is required for a query $\boldsymbol{x}_q$, the similarity ω_n between $\boldsymbol{x}_q$ and $\boldsymbol{x}_n$ is calculated, and a local PLS model is constructed by weighting samples with a similarity matrix $\boldsymbol{\Omega} \in \Re^{N \times N}$.

$$\boldsymbol{\Omega} = \mathrm{diag}(\omega_1, \omega_2, \cdots, \omega_N) \quad (11)$$

The output estimate $\hat{\boldsymbol{y}}_q$ corresponding to the query $\boldsymbol{x}_q$ is calculated as follows.

1) Determine the number of latent variables R and set $r = 1$.
2) Calculate the similarity matrix $\boldsymbol{\Omega}$.

3) Calculate $\boldsymbol{X}_r$, $\boldsymbol{Y}_r$, and $\boldsymbol{x}_{q,r}$:

$$\boldsymbol{X}_r = \boldsymbol{X} - \mathbf{1}_N[\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_M] \tag{12}$$

$$\boldsymbol{Y}_r = \boldsymbol{Y} - \mathbf{1}_N[\bar{y}_1, \bar{y}_2, \cdots, \bar{y}_L] \tag{13}$$

$$\boldsymbol{x}_{q,r} = \boldsymbol{x}_q - [\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_M]^T \tag{14}$$

$$\bar{x}_m = \sum_{n=1}^{N} \omega_n x_{nm} / \sum_{n=1}^{N} \omega_n \tag{15}$$

$$\bar{y}_l = \sum_{n=1}^{N} \omega_n y_{nl} / \sum_{n=1}^{N} \omega_n \tag{16}$$

where $\mathbf{1}_N \in \Re^N$ is a vector of ones.

4) Set $\hat{\boldsymbol{y}}_q = [\bar{y}_1, \bar{y}_2, \cdots, \bar{y}_L]^T$.

5) Derive the rth latent variable of $\boldsymbol{X}$

$$\boldsymbol{t}_r = \boldsymbol{X}_r \boldsymbol{w}_r \tag{17}$$

where $\boldsymbol{w}_r$ is the eigenvector of $\boldsymbol{X}_r^T \boldsymbol{\Omega} \boldsymbol{Y}_r \boldsymbol{Y}_r^T \boldsymbol{\Omega} \boldsymbol{X}_r$, which corresponds to the maximum eigen value.

6) Derive the rth loading vector of $\boldsymbol{X}$ and the regression coefficient vector.

$$\boldsymbol{p}_r = \frac{\boldsymbol{X}_r^T \boldsymbol{\Omega} \boldsymbol{t}_r}{\boldsymbol{t}_r^T \boldsymbol{\Omega} \boldsymbol{t}_r} \tag{18}$$

$$\boldsymbol{q}_r = \frac{\boldsymbol{Y}_r^T \boldsymbol{\Omega} \boldsymbol{t}_r}{\boldsymbol{t}_r^T \boldsymbol{\Omega} \boldsymbol{t}_r} \tag{19}$$

7) Derive the rth latent variable of $\boldsymbol{x}_q$.

$$t_{q,r} = \boldsymbol{x}_{q,r}^T \boldsymbol{w}_r \tag{20}$$

8) Update $\hat{\boldsymbol{y}}_q = \hat{\boldsymbol{y}}_q + t_{q,r} \boldsymbol{q}_r$.

9) If $r = R$, end. Otherwise, set

$$\boldsymbol{X}_{r+1} = \boldsymbol{X}_r - \boldsymbol{t}_r \boldsymbol{p}_r^T \tag{21}$$

$$\boldsymbol{Y}_{r+1} = \boldsymbol{Y}_r - \boldsymbol{t}_r \boldsymbol{q}_r^T \tag{22}$$

$$\boldsymbol{x}_{q,r+1} = \boldsymbol{x}_{q,r} - t_{q,r} \boldsymbol{p}_r \ . \tag{23}$$

10) Set $r = r + 1$ and go to step 5.

In step 3, the weighted mean of each variable is subtracted from each column of $\boldsymbol{X}$, $\boldsymbol{Y}$, and $\boldsymbol{x}_q^T$ to make the query near to the origin of the multidimensional space.

The prediction performance of LW-PLS depends on the definition of similarity ω_n between the query $\boldsymbol{x}_q$ and the samples $\boldsymbol{x}_n$. Here, the similarity is defined on the basis of the weighted Euclidean distance d_n.

$$\omega_n = \exp\left(-\frac{d_n}{\sigma_d \varphi}\right) \tag{24}$$

$$d_n = \sqrt{(\boldsymbol{x}_n - \boldsymbol{x}_q)^T \boldsymbol{\Theta} (\boldsymbol{x}_n - \boldsymbol{x}_q)} \tag{25}$$

$$\boldsymbol{\Theta} = \mathrm{diag}(\theta_1, \theta_2, \cdots, \theta_M) \tag{26}$$

where σ_d is a standard deviation of $\{d_n\}$, φ is a localization parameter, $\boldsymbol{\Theta} \in \Re^{M \times M}$ is a weighting matrix, and θ_m is a weight for the mth input variable.

The similarity decreases steeply when φ is small and gradually when φ is large. The prediction performance of LW-PLS is higher than, or at least the same as, that of PLS as long as the localization parameter φ is tuned properly, because LW-PLS becomes equivalent to PLS when $\varphi = \infty$. In other words, LW-PLS includes PLS as a special case.

An inferential control system, which integrated a soft-sensor based on LW-PLS with a commercial model predictive control (MPC) software, was developed and applied successfully to petrochemical processes [7]. The operation cost and environmental burden were significantly reduced with the developed system. In addition, it was reported that the implementation of the developed inferential control system had contributed toward reducing the burden of model maintenance.

In general, to realize accurate prediction by using LW-PLS, the weights θ_m need to be properly determined. For example, absolute values of input variables' regression coefficients of an LW-PLS model constructed without weights were used as the weights [5]. Another technique is to use the variance of regression coefficients of multiple LW-PLS models as the weights, because the weights should correspond to the strength of the nonlinear effect of the inputs on the output around a query [36]. Various weighting techniques have been reviewed in [4]. In the present work, regression coefficients of a global PLS model were used as weights because the prediction performance was improved by using this weighting technique.

Users can adopt various weighting strategies when using LW-PLS. For example, users can emphasize recent samples by using time information as an input variable and can adjust the balance between distance and time by using weights when defining the similarity. In general, users can achieve desirable performance by adopting a proper weighting strategy.

V. VERIFICATION OF VM WITH ACTUAL DATA

In the present work, VM was developed and validated by using historical data of the actual dry etching equipment. Product etching conversion differential was defined as the response variable that needs to be predicted. On the basis of the physical findings of etching engineers, 50 signals were selected from about 400 process signals collected by EES and they were used for VM design. This selection relies on qualitative models that the engineers have, and the models can describe the relationship among variables of the dry etching process. In this work, all the signals selected by the engineers were used so that the developed model was consistent with the engineers' knowledge even though they might not be significant from the statistical viewpoint. Then, nine statistics were calculated for each of the selected signals. After outliers were detected and removed by using the T^2 statistic, variable selection was conducted on the basis of VIP. Table I shows the selected 10 representative variables, which are combinations of signals, statistics, and recipe steps, together with the corresponding VIP. Here the information on the recipe step is concealed. These 10 representative variables were selected by engineers on the basis of VIP as well as their knowledge on the process.

In the dry etching equipment under investigation, etching conversion differential has been predicted by the sequential update model (SUM). SUM generates a MLR model recursively by using data of 100 wafers processed in the most recent period in order to cope with changes in equipment status. Figure 4 illustrates how SUM and LW-PLS prioritize data. In LW-PLS, a total of 200 samples are stored in the database, and samples with smaller d_n are prioritized through the similarity

TABLE I
VIP OF SELECTED REPRESENTATIVE VARIABLES

Signal	Statistic	VIP
Exposure	Mean	3.47
Temperature	StdDev	1.25
RF Power	StdDev	1.21
Flow	Mean	1.08
Gas Pressure	Mean	1.02
Flow	StdDev	0.94
Gas Pressure	Range	0.91
Voltage	Mean	0.72
Pressure	Mean	0.51
Position	Mean	0.46

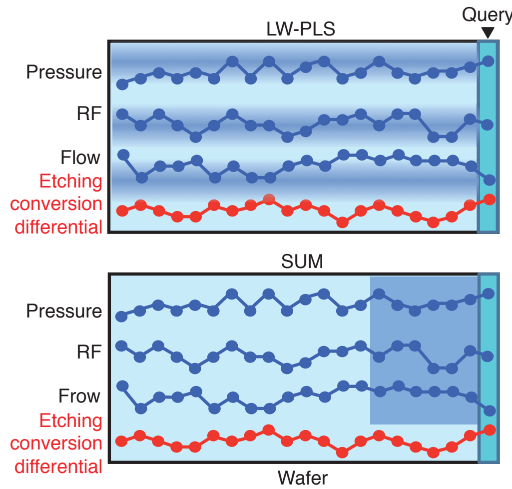

Fig. 4. Weighting strategies of LW-PLS and SUM. LW-PLS prioritizes samples with smaller distance to the query, while SUM prioritizes recent samples regardless of the similarity between the sample and the query.

ω_n, which is expressed as gradations of color in Fig. 4. In SUM, on the other hand, recent samples are prioritized regardless of the similarity between the sample and the query as shown by deep color in Fig. 4. Whereas SUM sequentially updates models relying on the most recent data, LW-PLS emphasizes samples similar to the query. Hence, LW-PLS can cope with not only gradual changes in process characteristics but also abrupt changes, which deteriorate the prediction accuracy of SUM. In addition, LW-PLS is more robust against outliers than SUM because outliers are automatically ignored due to their excessive distance from the query.

Wafers are processed under the conditions of high temperature and high voltage, using a variety of reactant gases, to make the circuitry on them. While wafers are processed, the excited plasma damages equipment parts. Consequently, these parts need to be replaced for the equipment maintenance. In addition, materials removed from wafers may block an exhaust line and be deposited inside the chamber. The deposited materials change the operating conditions. Thus, they must be removed during equipment maintenance.

Figure 5 shows actual operation data of the dry etching equipment. The data in the left-hand period are samples stored in the database for modeling, and the other data in the right-hand period are used to validate the developed VM. Both periods, i.e. the database period and the test period, include the maintenance, at which the RF power and other process signals change drastically. VM is required to cope with such changes in process signals caused by maintenance and to keep high prediction accuracy.

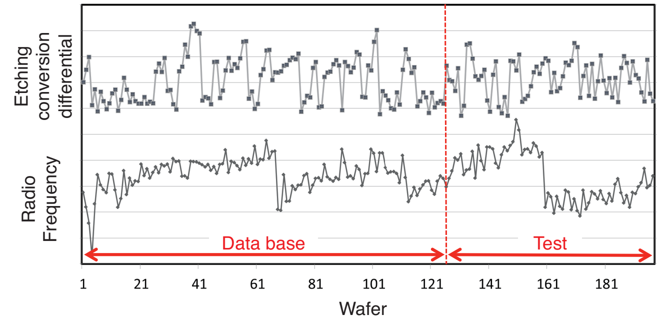

Fig. 5. Operation data used for model construction and model validation. Modeling data is not fixed during the test period. SUM uses data of 100 wafers processed in the most recent period for modeling, while LW-PLS stores data of 200 wafers processed in the most recent period in the database.

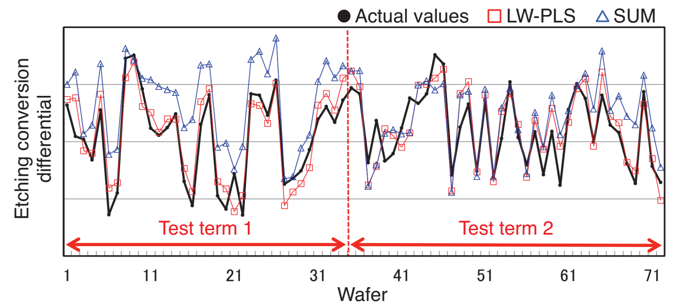

Fig. 6. Prediction results of etching conversion differential through LW-PLS and SUM.

LW-PLS and SUM were applied to actual dry etching equipment data in order to develop VM of the etching conversion differential. Two tuning parameters of LW-PLS, i.e., the localization parameter and the number of latent variables, were determined by leave-one-out cross validation (LOOCV) and set at $\varphi = 1.2$ and $R = 6$. The prediction performance of LW-PLS and SUM were compared with the data in the test period. Figure 6 shows the prediction results, in which the broken line indicates the maintenance. Test terms 1 and 2 are before and after the maintenance, respectively. This result has clearly shown that the proposed VM based on LW-PLS outperforms the conventional VM based on SUM. The VM based on LW-PLS was able to predict the etching conversion differential with accuracy in both periods. In other words, LW-PLS was applicable to the equipment with different parts while SUM was not.

Table II compares the prediction accuracy of LW-PLS, SUM, and also an artificial neural network (ANN) model. Here the ANN model was developed to compare the JIT modeling approach and the global nonlinear modeling approach. The developed three-layer ANN had five hidden nodes, in which sigmoid functions were used. All parameters of ANN were

TABLE II
COMPARISON OF LW-PLS, SUM, AND ANN IN THE PREDICTION ACCURACY: CORRELATION COEFFICIENT R BETWEEN THE ACTUAL VALUES AND THE PREDICTED VALUES AND ROOT MEAN SQUARE ERROR (RMSE)

Method	Term	R	RMSE
LW-PLS	Test 1	0.92	0.0015
	Test 2	0.86	0.0018
	1 + 2	0.90	0.0017
SUM	Test 1	0.88	0.0037
	Test 2	0.77	0.0024
	1 + 2	0.79	0.0031
ANN	Test 1	0.88	0.0019
	Test 2	0.76	0.0022
	1 + 2	0.83	0.0020

tuned by cross validation. It is confirmed that the RMSE (Root Mean Square Error) of LW-PLS is about 55% and 85% of the RMSE of SUM and ANN, that is, the prediction performance of LW-PLS is significantly higher than that of SUM and ANN. In addition, variations in the prediction performance at the different terms, i.e. before and after the maintenance, are reduced by adopting LW-PLS. When LW-PLS was used, correlation coefficient R between the actual values and the predicted values changed 0.06 and RMSE changed 0.0003. When SUM was used, on the other hand, R changed 0.11 and RMSE changed 0.0013. When ANN was used, R changed 0.12 and RMSE changed 0.0003. The prediction performance of SUM and ANN was strongly influenced by the equipment maintenance, i.e. parts replacement, while LW-PLS predicted the etching conversion differential with high accuracy regardless of the maintenance. Consequently, LW-PLS is more robust against the equipment maintenance than SUM and ANN, and VM based on LW-PLS can achieve the higher prediction performance.

VI. CONCLUSION

The equipment maintenance, in particular, parts replacement, frequently changes characteristics of dry etching equipment and thus deteriorates the predictive accuracy of VM. In this work, VM was developed by using LW-PLS in order to realize the high prediction accuracy even when maintenance is conducted. The prediction performance of LW-PLS was compared with that of the conventional SUM and ANN through their applications to the real operation data of the dry etching equipment. It was confirmed that the prediction error was reduced by half by adopting LW-PLS in comparison with SUM and also that LW-PLS was more robust against the equipment maintenance than SUM and ANN. The developed VM based on LW-PLS is shown to be effective in preventing deterioration of the prediction accuracy due to equipment maintenance.

LW-PLS is a kind of locally weighted regression methods, thus its performance depends on the definition of similarity. In this work, similarity was defined on the basis of the Euclidean distance, in which input variables were weighted by regression coefficients of the global PLS model. This approach improved the prediction performance of VM. But there might be a room for further improvement. The possible techniques include: 1) to update similarity according to the process characteristics, 2) to use time information as an input variable and adjust the balance between distance and time, and 3) to use correlation-based similarity.

In addition to predicting product characteristics, VM has a potential for reducing the frequency of measurement. This would bring a significant profit to manufacturing companies because the number of expensive metrology tools can be reduced without sacrificing product quality. Hence we are developing VM for other processes and expect to significantly improve the productivity.

ACKNOWLEDGMENT

This study is partially supported by Japan Society for the Promotion of Science (JSPS), Grant-in-Aid for Scientific Research (C) 24560940.

REFERENCES

[1] M. Fukasawa, A. Kawashima, N. Kuboi, H. Takagi, Y. Tanaka, H. Sakayori, K. Oshima, K. Nagahata, and T. Tatsumi, "Prediction of fluctuations in plasma wall interactions using an EES," in *International Symposium on Dry Process (DPS)*, pp. 247–248, 2008.

[2] J. Ringwood, S. Lynn, G. Bacelli, E. Ragnoli, and S. McLoone, "Estimation and control in semiconductor etch: Practice and possibilities," *IEEE Transactions on Semiconductor Manufacturing*, vol. 23, no. 1, pp. 87–98, 2010.

[3] D. Zeng and C. J. Spanos, "Virtual metrology modeling for plasma etch operations," *IEEE Transactions on Semiconductor Manufacturing*, vol. 22, no. 4, pp. 419–431, 2009.

[4] M. Kano and K. Fujiwara, "Virtual sensing technology in process industries: Trends and challenges revealed by recent industrial applications," *Journal of Chemical Engineering of Japan*, vol. 46, no. 1, pp. 1–17, 2013.

[5] S. Kim, M. Kano, H. Nakagawa, and S. Hasebe, "Estimation of active pharmaceutical ingredients content using locally weighted partial least squares and statistical wavelength selection." *International Journal of Pharmaceutics*, vol. 421, no. 2, pp. 269–274, 2011.

[6] H. Nakagawa, T. Tajima, M. Kano, S. Kim, S. Hasebe, T. Suzuki, and H. Nakagami, "Evaluation of infrared-reflection absorption spectroscopy measurement and locally weighted partial least-squares for rapid analysis of residual drug substances in cleaning processes." *Analytical Chemistry*, vol. 84, no. 8, pp. 3820–3826, 2012.

[7] S. Kim, M. Kano, S. Hasebe, A. Takinami, and T. Seki, "Long-term industrial applications of inferential control based on just-in-time soft-sensors: Economical impact and challenges," *Industrial & Engineering Chemistry Research*, vol. 52, no. 35, pp. 12346–12356, 2013.

[8] International SEMATECH-JEITA/Selete Collaboration, *Equipment engineering functions guidebook Version 2.5*, 2002.

[9] P. Chen, S. Wu, J. Lin, F. Ko, H. Lo, J. Wang, C. Yu, and M. Liang, "Virtual metrology: a solution for wafer to wafer advanced process control," in *IEEE International Symposium on Semiconductor Manufacturing (ISSM)*. IEEE, pp. 155–157, 2005.

[10] M.-F. Wu, C.-H. Lin, D.-H. Wong, S.-S. Jang, and S.-T. Tseng, "Performance analysis of ewma controllers subject to metrology delay," *IEEE Transactions on Semiconductor Manufacturing*, vol. 21, no. 3, pp. 413–425, 2008.

[11] J. Moyne, "Making the move to fab-wide apc," *Solid State Technology*, vol. 47, no. 9, pp. 47–52, 2004.

[12] S. J. Qin, G. Cherry, R. Good, J. Wang, and C. A. Harrison, "Semiconductor manufacturing process control and monitoring: A fab-wide framework," *Journal of Process Control*, vol. 16, no. 3, pp. 179–191, 2006.

[13] A. A. Khan, J. R. Moyne, and D. M. Tilbury, "An approach for factory-wide control utilizing virtual metrology," *IEEE Transactions on Semiconductor Manufacturing*, vol. 20, no. 4, pp. 364–375, 2007.

[14] T.-H. Pan, B.-Q. Sheng, D. S.-H. Wong, and S.-S. Jang, "A virtual metrology system for predicting end-of-line electrical properties using a mancova model with tools clustering," *IEEE Transactions on Industrial Informatics*, vol. 7, no. 2, pp. 187–195, 2011.

[15] Y. Yao and F. Gao, "A survey on multistage/multiphase statistical modeling methods for batch processes," *Annual Review in Control*, vol. 33, no. 2, pp. 172–183, 2009.

[16] H. Hotelling, "The generalization of student's ratio," *The Annals of Mathematical Statistics*, vol. 2, no. 3, pp. 360–378, 1931.

[17] S. Wold, E. Johansson, and M. Cocchi, *3D QSAR in Drug Design; Theory, Methods, and Applications*. ESCOM, pp. 523–550, 1993.

[18] S. Wold, M. Sjöström, and L. Eriksson, "PLS-regression: a basic tool of chemometrics," *Chemometrics and Intelligent Laboratory Systems*, vol. 58, no. 2, pp. 109–130, 2001.

[19] S. A. Lynn, J. Ringwood, and N. MacGearailt, "Global and local virtual metrology models for a plasma etch process," *IEEE Transactions on Semiconductor Manufacturing*, vol. 25, no. 1, pp. 94–103, 2012.

[20] I.-G. Chong and C.-H. Jun, "Performance of some variable selection methods when multicollinearity is present," *Chemometrics and Intelligent Laboratory Systems*, vol. 78, no. 1-2, pp. 103–112, 2005.

[21] M. Kano and M. Ogawa, "The state of the art in chemical process control in Japan: Good practice and questionnaire survey," *Journal of Process Control*, vol. 20, no. 9, pp. 969–982, 2010.

[22] M. Kano and Y. Nakagawa, "Data-based process monitoring, process control, and quality improvement: Recent developments and applications in steel industry," *Computers & Chemical Engineering*, vol. 32, no. 1-2, pp. 12–24, 2008.

[23] C. G. Atkeson, A. W. Moore, and S. Schaal, "Locally weighted learning," *Artificial Intelligence Review*, vol. 11, no. 1-5, pp. 11–73, 1997.

[24] O. Maron and A. W. Moore, "The racing algorithm: Model selection for lazy learners," in *Lazy Learning*, D. W. Aha, Ed. Dordrecht: Springer Netherlands, pp. 193–225, 1997.

[25] G. Bontempi, M. Birattari, and H. Bersini, "Lazy learning for local modelling and control design," *International Journal of Control*, vol. 72, no. 7-8, pp. 643–658, 1999.

[26] M. Birattari, G. Bontempi, and H. Bersini, "Lazy learning meets the recursive least squares algorithm," in *Advances in Neural Information Processing Systems*. MIT Press, pp. 375–381, 1999.

[27] D. Kibler, D. W. Aha, and M. K. Albert, "Instance-based prediction of real-valued attributes," *Computational Intelligence*, vol. 5, no. 2, pp. 51–57, 1989.

[28] W. S. Cleveland, "Robust locally and smoothing weighted regression scatterplots," *Journal of the American Statistical Association*, vol. 74, no. 368, pp. 829–836, 1979.

[29] H. Shigemori, M. Kano, and S. Hasebe, "Optimum quality design system for steel products through locally weighted regression model," *Journal of Process Control*, vol. 21, no. 2, pp. 293–301, 2011.

[30] S. Chang, E. Baughman, and B. McIntosh; "Implementation of Locally Weighted Regression to Maintain Calibrations on FT-NIR Analyzers for Industrial Processes," *Applied Spectroscopy*, vol. 55, pp. 1199–1206, 2001.

[31] C. Cheng and M. Chiu; "A New Data-Based Methodology for Nonlinear Process Modeling," *Chemical Engineering Science*, vol. 59, pp. 2801–2810, 2004.

[32] Z. Ge and Z. Song; "Online Monitoring of Nonlinear Multiple Mode Processes Based on Adaptive Local Model Approach," *Control Engineering Practice*, vol. 16, pp. 1427–1437, 2008.

[33] Z. Wang, T. Isaksson, and B. Kowalski; "New Approach for Distance Measurement in Locally Weighted Regression," *Analytical Chemistry*, vol. 66, pp. 249–260, 1994.

[34] K. Fujiwara, M. Kano, S. Hasebe, and A. Takinami; "Soft-Sensor Development Using Correlation-Based Just-in-Time Modeling," *AIChE Journal*, vol. 55, pp. 1754–1765, 2009.

[35] K. Fujiwara, M. Kano, and S. Hasebe; "Development of Correlation-Based Clustering Method and Its Application to Software Sensing," *Chemometrics and Intelligent Laboratory Systems*, vol. 44, pp. 130–138, 2010.

[36] S. Kim, R. Okajima, M. Kano, and S. Hasebe, "Development of soft-sensor using locally weighted pls with adaptive similarity measure," *Chemometrics and Intelligent Laboratory Systems*, vol. 124, pp. 43–49, 2013.

Toshiya Hirai Toshiya Hirai received B.E. and M.E. degrees in precision engineering from Ibaraki University, Ibaraki, Japan, in 1984 and 1986, respectively. In 1986, he joined Sony Corporation. His background is in process control technologies for semiconductor manufacturing. He participated in the development of the Equipment Engineering System to Selete (Semiconductor Leading Edge Technologies, Inc.) in 2003 and 2006, respectively.

Manabu Kano Dr. Manabu Kano graduated from Department of Chemical Engineering, Kyoto University, in 1992. After he got master degree in 1994, he took a position of instructor at Kyoto University. In 1999, he received Ph.D. from Kyoto University. From 1999 to 2000, he was a visiting scholar at the Ohio State University, US. Since 2012, Dr. Manabu Kano is a professor of Department of Systems Science, Kyoto University. His research interest has focused on process data analysis for quality improvement, process monitoring, and process control. He has conducted researches in collaboration with various industries such as semiconductor, chemical, steel, and pharmaceutical. Dr. Manabu Kano has received many awards including Best Paper Award, Technology Award, and also Pioneer Technology Award from The Society of Instrument and Control Engineers; Instrumentation, Control and System Engineering Research Award from The Iron and Steel Institute of Japan; Research Award for Young Investigators from The Society of Chemical Engineers, Japan.