Weighted windowed PLS models for virtual metrology of an industrial plasma etch process

Shane Lynn
and John V. Ringwood
Department of Electronic Engineering,
National University of Ireland,
Maynooth, Co. Kildare, Ireland.
Email: shane.a.lynn@eeng.nuim.ie
john.ringwood@eeng.nuim.ie

Niall MacGearailt
School of Electronic Engineering
Dublin City University,
Glasnevin, Dublin 9,
Ireland.
Email: niall.macgearailt@eeng.dcu.ie

Abstract—Virtual metrology is the prediction of metrology variables using easily accessible process variables and mathematical models. Because metrology variables in semiconductor manufacture can be expensive and time consuming to measure, virtual metrology is beneficial as it reduces cost and throughput time. This work proposes a virtual metrology scheme that uses sliding-window models to virtually measure etch rates in an industrial plasma etch process. The windowed models use partial least squares (PLS) regression and a sample weighting scheme to combat the effects of both process drifts due to machine conditioning and process shifts due to maintenance events. An industrial data set is examined and the weighted windowed PLS models outperform global models and non-weighted windowed models.

I. INTRODUCTION

Plasma etching is a semiconductor manufacturing process where material is removed in exact amounts from the surface of silicon wafers. Etchant gases in plasma form are directed towards the wafer surface using electric and magnetic fields, where they react and evaporate to remove matter. Plasma etching is preferred to wet etching methods as it is capable of producing a highly anisotropic etch, allowing deep and narrow trenches to be etched into the wafer surface [1].

Plasma etching is performed in etching chambers, where conditions are monitored carefully, and etchant gases are introduced in exact quantities to form plasma. The etch rate is the rate at which material is removed from the wafer surface. It is important to control and monitor the etch rate as precisely as possible since the ultimate trench depth and profile has a dramatic impact on the performance of the devices being produced. In-situ measurements of etch rate and etch depth are expensive, time consuming, and invasive, sometimes requiring destruction of the device. Hence, measurements are not available to machine operators without considerable metrology delay and it is not practical for fabrication plants to measure every wafer processed. Control is difficult to implement under such conditions and machines running out of specification without detection can lead to days of useless produce.

Fortunately a great deal of information is available from the etch chambers as each wafer is processed. This information includes the temperatures, pressures, gas flows, and electrical signals that are easily monitored during processing. The aim of virtual metrology (VM) techniques is to find mathematical relationships between these easily measured variables and the variables of interest such as the etch rate or etch depth.

Since it enables process monitoring with low overhead, VM is gaining increased interest in semiconductor manufacturing of late in an attempt to increase process control and reduce cycle time and cost. As Moore's law drives component dimensions smaller and smaller [2], there is an urgent requirement for more accurate control of fabrication processes. Fab wide VM schemes have been suggested that allow feed forward and feedback of information between different processes [3]. However, the individual VM schemes for each process are not trivial. Due to the complexities of the plasma etching process, it is notoriously difficult to model and predict, with more challenges arising with emerging technologies and techniques [4]. Some work has focussed on the prediction of contamination in plasma etching chambers [5], and a great deal of work has been completed on endpoint detection from ancillary variables [6] [7]. A comprehensive review of virtual metrology in plasma etch can be found in [8].

Plasma etch modelling is further complicated by regular preventative maintenance (PM) cycles that aim to maintain the equipment to a high standard. Unfortunately these maintenance events can have a pronounced effect on the behaviour of the etch chambers. Relationships between variables can change dramatically, and substantial changes and unpredictable shifts in etch rate are observed. These changes in system behaviour can invalidate models built upon previous measurements, increasing the difficulty of delivering accurate predications of etch rate. Along with this complication, the etch chambers exhibit considerable process drift during processing. The physical properties of the chamber are changed by chemicals deposited on the chamber walls after each wafer etched, and this ultimately affects the etch response. A sample variable that exhibits PM shifts and process drift is shown in Figure 1

Previous work has examined the disaggregation of datasets to combat the effect of PM events [9] and to identify key process variables from datasets [10]. This paper focuses on the VM of etch rate using sliding-window modelling methods to
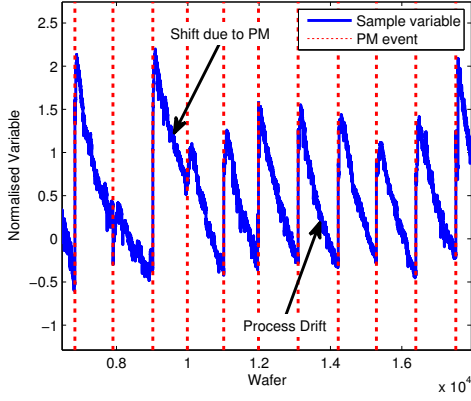
271

Fig. 1. Normalised variable value from etch process that exhibits shifts due to maintenance events and drift due to chamber conditioning.

achieve better results. A novel weighted windowed PLS technique is developed that incorporates maintenance information into the VM scheme.

The remainder of this paper is set out as follows: Section II describes the mathematical techniques used to model the plasma process, Section III examines the modelling schemes investigated, Section IV details structure of the data used for the investigations, Section V shows the modelling results achieved and finally the conclusions are given in Section VI.

II. MODELLING TECHNIQUES

Partial least squares (PLS) is chosen as the modelling technique for this analysis. PLS is a mathematical technique used for the modelling of input-output processes. Originally developed for the field of econometrics, PLS is finding uses in many industries as of late, and has advantages over other regression methods in applications where a large number of collinear regressor variables are available to model a system where a mathematical relationship cannot easily be described. In these cases with many variables, few samples, and ill-understood relationships, techniques such as multiple linear regression (MLR) can produce erroneous predictions [11] [12]. In our work, a weighting factor is incorporated to the modelling process to take machine maintenance events into account. This section describes PLS, weighted least squares, and the combination of these techniques to achieve the weighted PLS algorithm.

A. Partial Least Squares

Systems with a large number of regressor variables often contain only a small number of underlying or latent factors that account for variations in the output. The aim of PLS is to extract these latent factors and model the responses. This is similar to principal component analysis (PCA), an unsupervised statistical technique where the latent factors that describe the variance within a data set, known as principal components, are extracted. PCA decomposes a data matrix, $X_{n \times m}$, made up of n samples and m variables, as the sum of

the outer product of vectors $\mathbf{t}_i$ and $\mathbf{p}_i$ plus a residual matrix $\mathbf{E}$ [6].

$$\mathbf{X} = \mathbf{t}_1\mathbf{p}_1^T + \mathbf{t}_2\mathbf{p}_2^T + ... + \mathbf{t}_l\mathbf{p}_l^T + \mathbf{E} \quad (1)$$
$$= \mathbf{TP^T} + \mathbf{E} \quad (2)$$

where,

$$\mathbf{T} = [\mathbf{t}_1, \mathbf{t}_2, ... \mathbf{t}_l] \quad (3)$$
$$\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, ... \mathbf{p}_l] \quad (4)$$

and l is the number of principal components. The vectors $\mathbf{t}_i$ are known as the *scores* and $\mathbf{T} \in \mathbb{R}^{n \times l}$ the score matrix; the $\mathbf{p}_i$ vectors are the *loadings* and $P \in \mathbb{R}^{m \times l}$ the loadings matrix. For PCA, the decomposition of $\mathbf{X}$ is such that the columns of the loading matrix $\mathbf{P}$ are orthonormal and columns of the score matrix $\mathbf{T}$ are orthogonal. The first component is the linear combination of the m original variables that explains the greatest amount of variability ($\mathbf{t}_1 = \mathbf{Xp}_1$). For a matrix $\mathbf{X}$ of rank r, r principal components can be calculated. However, the first l ($l < r$) of these may be sufficient to explain the bulk of the variance in the data. If $l = rank(\mathbf{X})$, then $\mathbf{E} = 0$, and the representation of the data is exact for the new variables [13].

However, PCA is an unsupervised technique in which the data matrix is decomposed with no reference to any output variables. PLS attempts to remedy this situation by decomposing the output variables simultaneously and choosing the scores so that the relationship between successive pairs of scores is as strong as possible.

Hence, given a system with an input matrix $\mathbf{X}_{n \times m}$, and an output matrix $\mathbf{Y}_{n \times k}$, where k is the number of output variables, $\mathbf{X}$ and $\mathbf{Y}$ are decomposed simultaneously as

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E} \text{ and} \quad (5)$$
$$\mathbf{Y} = \mathbf{UQ}^T + \mathbf{F} \quad (6)$$

where $\mathbf{U}$, $\mathbf{Q}$, and $\mathbf{F}$ are the output matrix score, loading and residual matrices. These two decompositions are tied by an *inner relationship*, $u_i = b_i t_i$, such that the $\mathbf{X}$-loadings are rotated to maximise the covariance between the X-scores and the $\mathbf{Y}$-scores. The *mixed relationship* can be written as

$$\hat{\mathbf{Y}} = \mathbf{TBQ}^T + \mathbf{F}, \quad (7)$$

where $\mathbf{F}$ is to be minimized to predict the output as accurately as possible [14]. $\mathbf{T}$ are the latent vectors, decomposed from $\mathbf{X}$. If their number is equal to the rank of $\mathbf{X}$, they perform an exact decomposition of $\mathbf{X}$ as in PCA. It should be noted however, that they only estimate $\mathbf{Y}$, hence the notation $\hat{\mathbf{Y}}$. For the purposes of this work, the number of components retained in the model is optimised using a set of validation data not used in the determination of the PLS model. The validation data is a set that is randomly chosen from the training data before the PLS model is determined. The number of components that produces the lowest validation data error is kept in the PLS model. Error values for typical training and validation sets are shown in Figure 2.
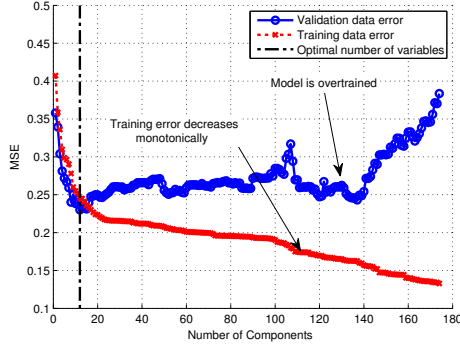
Fig. 2. Typical mean squared error (MSE) curves generated during PLS model training. Notice that as the model becomes overfitted on the training data, the validation error increases. The number of components retained in the model is at the point of minimum validation error.

There are a number of different techniques used to perform PLS on a given dataset. For the purposes of this paper, the non-iterative partial least squares (NIPALs) algorithms is used, as described in [15].

B. Weighted Linear Regression

Multiple Linear Regression (MLR) is a linear method that attempts to model the relationship between two or more regressor variables and an output variable by fitting a linear equation to the observed data. This leads to a model of the form

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_m x_m \qquad (8)$$

being fit to each data point. Here, y is a measured output and $x_1 \ldots x_m$ are system inputs that can be used as regression variables. We denote the data points to be

$$y_i, x_{i,1}, x_{i,2}, \ldots, x_{i,m}, i = 1, \ldots n$$

The output observations y_i will be represented by the vector $\mathbf{y}$, the unknown model parameters, $\beta_0, \beta_1, \ldots, \beta_m$ by the vector $\boldsymbol{\beta}$, and the data matrix $\mathbf{X}_{n \times (m+1)}$ takes the form:

$$\mathbf{X} = \begin{bmatrix} 1 & x_{1,1} & x_{1,2} & \ldots & x_{1,m} \\ 1 & x_{2,1} & x_{2,2} & \ldots & x_{2,m} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n,1} & x_{n,2} & \ldots & x_{n,m} \end{bmatrix} \qquad (9)$$

Hence for a given $\boldsymbol{\beta}$, the vector of predicted values $\hat{\mathbf{y}}$ is given by

$$\hat{\mathbf{y}} = \mathbf{X}\boldsymbol{\beta} \qquad (10)$$

We wish to find $\boldsymbol{\beta}$ such that the sum of squared residuals,

$$S = \sum_{i=1}^{n} e_i^2 = \mathbf{e}^T \mathbf{e} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}), \qquad (11)$$

is at a minimum. S is minimised when its gradient with respect to β is zero. Since

$$S(\boldsymbol{\beta}) = \mathbf{y}^T \mathbf{y} - 2\boldsymbol{\beta}^T \mathbf{X}^T \mathbf{y} + \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{X}\boldsymbol{\beta}, \qquad (12)$$

then

$$\frac{\delta S}{\delta \beta} = -2\mathbf{X}^T \mathbf{y} + 2\mathbf{X}^T \mathbf{X}\boldsymbol{\beta} = 0, \qquad (13)$$

which we can rearrange as

$$(\mathbf{X}^T \mathbf{X})\boldsymbol{\beta} = \mathbf{X}^T \mathbf{y}. \qquad (14)$$

These *normal equations* can be solved to give a unique solution for $\boldsymbol{\beta}$ provided that $\mathbf{X}^T \mathbf{X}$ is nonsingular [16].

In some applications, it can be shown that some of the observations used in a regression analysis are less reliable than others [17], e.g. the variances of all the observations are not equal or the observations are correlated. A modified approach to the MLR algorithm can be taken to ensure that the best $\boldsymbol{\beta}$ can be found for this situation.

In this case, the sum of squared residuals to the minimised is given by

$$S = \sum_{i=1}^{n} W_{ii} e_i^2 = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T W (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \qquad (15)$$

where W is a diagonal matrix of the weights applied to each of the samples $i = 1, 2, \cdots, n$. In many applications these weights are inversely proportional to the variance of the samples involved in the regression. Following the same calculations as before and expressing the equations in matrix form yields

$$(\mathbf{X}^T \mathbf{W} \mathbf{X})\boldsymbol{\beta} = \mathbf{X}^T \mathbf{W} \mathbf{y} \qquad (16)$$

which can again be solved for a unique solution of $\boldsymbol{\beta}$.

C. Weighted PLS

The weighted PLS algorithm employed in this work uses a combination of PLS and weighted least squares. For systems with a single output variable, PLS can be visualised as a regression procedure where the scores $\mathbf{T}$ generated from the input matrix $\mathbf{X}$ are used as regressor variables to predict the output variable contained in $\mathbf{Y}$. As $\mathbf{Y}$ contains only one variable, it is not decomposed into loading and score matrices. In this case, the inner-relation for the PLS algorithm can be simply described as

$$\mathbf{Y} = \mathbf{T}\bar{\boldsymbol{\beta}}, \qquad (17)$$

where $\bar{\boldsymbol{\beta}}$ is normally found using the linear least squares algorithm. A weighted PLS system can be developed by introducing weights to this least squares problem. Each sample score determined through the PLS algorithm can be assigned a weight depending on some criteria, and the parameters $\bar{\boldsymbol{\beta}}$ can be determined instead using the weighted linear regression method. Hence $\bar{\boldsymbol{\beta}}$ is given by

$$\bar{\boldsymbol{\beta}} = (\mathbf{T}^T \mathbf{W} \mathbf{T})^{-1} \mathbf{T}^T \mathbf{W} \mathbf{y} \qquad (18)$$

where $\mathbf{W}$ is a diagonal matrix of weights relevant to each sample contained in $\mathbf{T}$. The weights applied should be determined through engineering knowledge of the system or through knowledge of differing conditions under which samples were collected.

III. PLASMA ETCH MODELLING

Four different modelling methods are examined and compared in this work. This section explains the differences between each method investigated.

1) Global Modelling: Global models describe models that are built using a set of training data and are used to predict all future wafers. These models do not update over time, and are suitable for systems that do not vary in time or radically change behaviour. The PLS algorithm is applied only once at initialisation using the training dataset. Preferably, the training data set is chosen so that it captures the main behaviour of the system and future activity is assumed to operate in the same regime.

2) Sliding window model: Sliding window models begin with a training dataset and are used to predict future wafers until new information becomes available. When more measurements are made, the oldest information in the training set is disregarded and the model is retrained, taking the newest information into account. These models are trained using the PLS algorithm every time new information becomes available. Techniques exist to incorporate new information into PLS [12] and PCA [18] [19] matrices, and are often referred to as recursive or adaptive PCA and PLS methods. For the purposes of this work, the PLS models are retrained when a new sample is introduced.

3) Weighted sliding window model: In this case the weighted PLS technique is applied when training the models for each window. The weights applied to each sample are determined depending on their relevance to the sample at the front of the window. The sample weights vary in accordance with the tool maintenance history to satisfy two assumptions. Firstly, due to process drift, it is assumed that samples closer to the front of the window are more relevant for prediction. Secondly, it is assumed that samples contained within the same PM cycle as the target sample are more relevant than samples from previous PM cycles.

In order to fulfill these assumptions, samples are first assigned a linearly decreasing weight across the window length, with the most recent sample given a weight of 1 and the oldest samples a weight of 0. Next the samples weights are adjusted according to the number of PM cycles spanned by the window length. The samples contained in the most recent PM cycle are first incremented, and older PM cycles are incremented by progressively lesser amounts. For example, in a window that spans three PM cycles, 2 is added to the samples in the most recent PM cycle, 1 to the second most recent, and 0 to the oldest PM cycle. This scheme leads to a weighting profile shown in Figure 3. To increase the effect of the weighting scheme, the weights are exponentially transformed to the profile shown in Figure 4.

4) Weighted sliding window model with recursion: As a final addition to the modelling structure, an element of recursion is incorporated to the weighted regression. As there is some level of autocorrelation observed in the etch rate signal, previous values of etch rate y_{i-1} are added as regressor variable to the weighted regression that is performed during
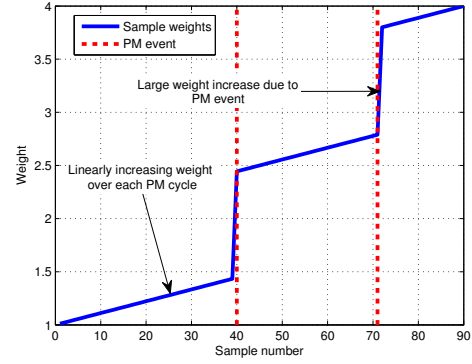

Fig. 3. Weighting profile for window size of 90 samples spanning three PM cycles. The most recent samples are on the right of the plot.
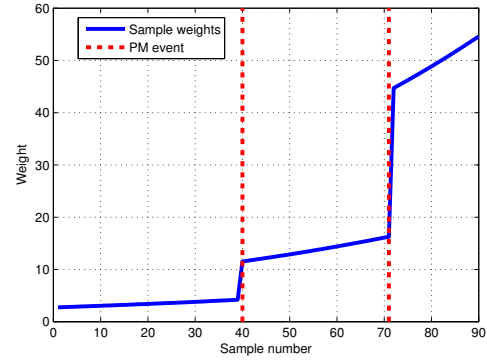

Fig. 4. Weighting profile for window size of 90 samples after exponential transformation.

model training for $\hat{y}_i$. Effectively the etch rate signal is shifted by one sample and used as another regressor, such that the etch rate measurement for the first sample y_1 is used to predict the second etch rate $\hat{y}_2$, the etch rate measurement from the second sample y_2 is used to predict the third $\hat{y}_3$ etc.

The PLS scores and the weighting scheme are calculated as before. However in this case, instead of simply regressing the PLS scores to the current etch rate, the PLS scores and the one sample shifted etch rates values are now used to model the new etch rate.

IV. DATASET DESCRIPTION

Over a period of six months, etch process (EP) data from a five-step industrial trench etch process was collected from an etch chamber. In total, process data for 18524 wafers was collected. This dataset spans 18 PM cycles in total. The EP data collected consists of over 90 different variables such as time series measurements of power delivered to the chamber, matchbox inductor positions, internal chamber temperatures, gas flow rates, and end point times. Time series measurements were reduced to mean and standard deviation values for each step. As the main trench etch occurred in one of the later steps of the etch process, the variables from these steps were used for the analysis. Through visual inspection and further engineering knowledge, the variables were reduced to a set of

28 variables that were thought to be of most importance to the etch process. All variables were mean centered and variance normalised before analysis.

Of the 18524 wafers collected, 794 wafers were measured optically downstream from the etch process. This corresponds to a measurement frequency of 4.3%. The etch rate for each wafer is determined by dividing the measured etch depth by the main trench etch time. The etch rate values are incorporated into the dataset, and are aligned with the respective EP data.

Due to errors with recording instruments and glitches with machinery, a number of erroneous values were recorded. Hotelling's T^2 statistic was used to identify clear outliers in the dataset and those points were removed. As a result, some gaps in the data exist where wafers were processed but no data is available, but unfortunately this is unavoidable.

V. RESULTS

The models described in Section III are applied to the etch dataset using window lengths between 30 and 350 samples. The window lengths describe the number of past samples used to train PLS models for the prediction of each wafer. In the case of global models, the window size specifies the number of wafers used initially to create the model.

As an error metric, the mean squared error (MSE) for the fitted model is given by

$$MSE = \frac{1}{n} \sum_{i=1}^{n} e_i^2 \qquad (19)$$

A larger MSE value for a model indicates a worse fit than one with a lower MSE. The MSE for a model has units equal to the square of the original quantity units.

The coefficient of determination, R^2, is also provided as a measure of how well future outcomes may be predicted by the developed model. The R^2 statistic has values between 0 and 1 and is defined as

$$R^2 = 1 - \frac{\sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}, \qquad (20)$$

with $\bar{y}$ denoting the mean of the observed output values. R^2 gives some information about the goodness of fit of a model. For a linear fit, an R^2 of 1 indicates that the regression line fits the data perfectly. The R^2 statistic is equivalent to the square of the correlation between the real and fitted values, y and $\hat{y}$.

The MSE for each model type over a range of window sizes is depicted in Figure 5. The large MSEs recorded for global models over the window lengths investigated show that the global models fail at successfully modelling the etch rate variations. A more detailed view of the MSE results for the other modelling techniques is provided in Figure 6. The R^2 values for each of the techniques are shown in Figure 7. We can see from these figures that the weighting scheme noticeably increases the accuracy of the windowed PLS models. Further improvements are observed with the incorporation of recursion in the model. The weighted windowed models that
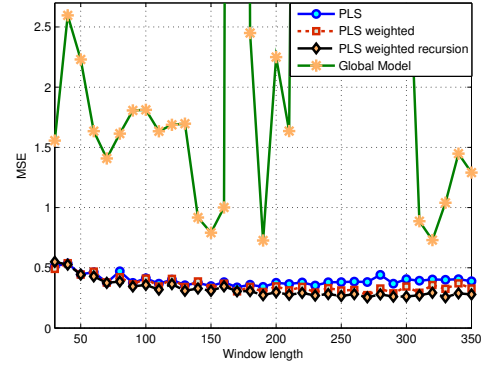

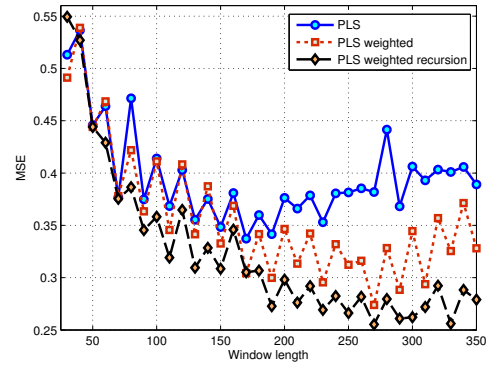Fig. 5. Mean squared error values for all model types over a range of window lengths.

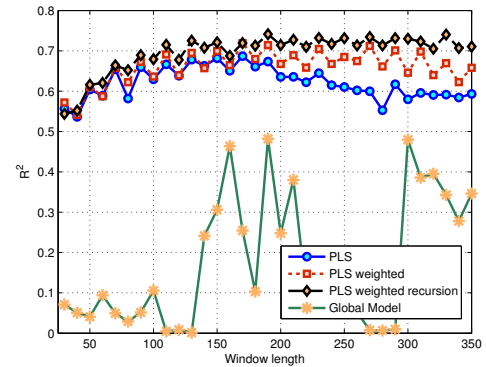Fig. 6. Mean squared errors for windowed models only.

Fig. 7. Variation of R^2 values for all models over range of window lengths.

incorporate recursion perform consistently better than all of the other model structures.

The best results were recorded for a window length of 270 wafers that spanned 7 PM cycles. This window length predicted the actual etch rate for the remaining 524 wafers with a MSE of 0.255 and R^2 of 0.734. The predictions and actual etch rate for this model and a similarly trained global model are shown on the same axis in Figure 8. A graphical analysis of the errors produced from this model is shown in Figure 9. Errors are found to obey a relatively normal distribution.
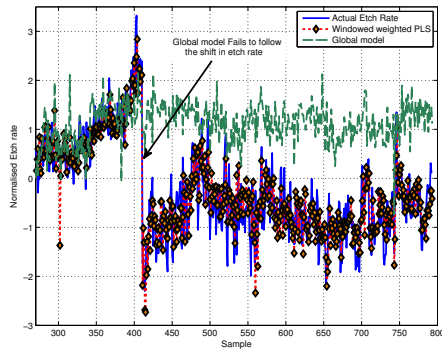
Fig. 8. Etch rate predictions from windowed weighted PLS model and global model using window length of 270 wafers.
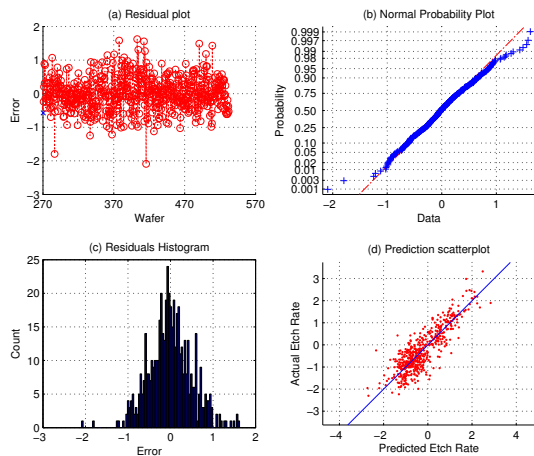

Fig. 9. Graphical analysis of model errors showing (a) error magnitudes, (b) normal probability plot, (c) histogram of error values, and (d) scatter plot of predicted and actual values.

VI. CONCLUSION

This paper has investigated the application of PLS based sliding window models to an industrial trench etch data set. Sliding windowed models outperform unchanging global models for all window sizes examined. This result is expected due to the effect of maintenance activities that can dramatically change the behaviour of the system and the relationships between the input and output variables. Windowed models take new information into account and as a result they are more successful at following new regimes of operation that arise after maintenance events.

An increase in accuracy is achieved through the application of a novel sample weighting scheme to the PLS algorithm, where the weights are determined in accordance with the maintenance history of the machine. The incorporation of this information leads to an decrease in prediction error and an increase in the coefficient of determination R^2 for almost all window sizes investigated.

Further increases in accuracy are achieved by incorporating some recursion in the models. In this case, the addition of previous etch rate measurements to the virtual metrology models aid predictions of future etch rates.

The predictions made during this work require only the etch process data collected from the manufacturing tool. No additional sensors were installed on the etch tool. The main advantage of this approach is that a low cost virtual metrology scheme is realisable with only the minimum amount of extra software infrastructure.

REFERENCES

[1] F. F. C. . J. P. Chang, *Lecture Notes on Principles of Plasma Processing*. Kluwer Academic/Plenum Publishers, 2003.

[2] G. E. Moore, "Cramming more components onto integrated circuits," *Electronics*, vol. 38, no. 8, pp. 114–117, 1965.

[3] A. Khan, D. Tilbury, and J. Moyne, "Fab-wide virtual metrology and feedback control," in *Asian AEC/APC Symposium*. NSF Engineering Reasearch Center for reconfigurable Manufacturing Systems, University of Michegan., 2006.

[4] Y. Yang, M. Wang, and M. J. Kushner, "Progress, opportunities and challenges in modeling of plasma etchingh," in *Interconnect Technology Conference, 2008. IITC 2008. International*, June 2008, pp. 90–92.

[5] S. Imai, "Virtual metrology for plasma particle in plasma etching equipment," in *Semiconductor Manufacturing, 2007. ISSM 2007. International Symposium on*, Oct. 2007, pp. 1–4.

[6] H. Yue, S. J. Qin, J. Wiseman, and A. Toprac, "Plasma etching endpoint detection using multiple wavelengths for small open-area wafers," *Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films*, vol. 19, no. 1, pp. 66–75, Jan/Feb 2001.

[7] D. White, B. Goodlin, A. Gower, D. Boning, H. Chen, H. Sawin, and T. Dalton, "Low open-area endpoint detection using a pca-based t^2 statistic and q statistic on optical emission spectroscopy measurements," *Semiconductor Manufacturing, IEEE Transactions on*, vol. 13, no. 2, pp. 193–207, May 2000.

[8] J. V. Ringwood, S. Lynn, G. Bacelli, B. Ma, E. Ragnoli, and S. McLoone, "Estimation and control in semiconductor etch: Practice and possibilities," *IEEE Transactions on Semiconductor Manufacturing*, submitted for publication.

[9] S. Lynn, J. Ringwood, E. Ragnoli, S. McLoone, and N. MacGearailt, "Virtual metrology for plasma etch using tool variables," in *Advanced Semiconductor Manufacturing Conference, 2009. ASMC '09. IEEE/SEMI*, May 2009, pp. 143–148.

[10] E. Ragnoli, S. McLoone, S. Lynn, J. Ringwood, and N. Macgearailt, "Identifying key process characteristics and predicting etch rate from high-dimension datasets," in *Advanced Semiconductor Manufacturing Conference, 2009. ASMC '09. IEEE/SEMI*, May 2009, pp. 106–111.

[11] R. D. Tobias, "An introduction to partial least squares regression," SAS Institute Inc., Tech. Rep., 1997.

[12] A. A. Khan, J. Moyne, and D. Tilbury, "Virtual metrology and feedback control for semiconductor manufacturing processes using recursive partial least squares," *Journal of Process Control*, vol. 18, no. 10, pp. 961–974, 2008.

[13] A. Afifi, V. A. Clarke, and S. May, *Computer-Aided Multivariate Analysis*, 4th ed., C. Chatfield, M. Tanner, and J. Zidek, Eds. Chapman & Hall/CRC, 2004.

[14] H. Abdi, *Partial Least Squares (PLS) Regression*. Sage, 2003, pp. 792–795.

[15] P. Geladi and B. R. Kowalski, "Partial least-squares regression: A tutorial," *Analytica Chimica Acta*, vol. 185, pp. 1–17, 1986.

[16] J. A. Rice, *Mathematical Statistics and Data Analysis*, third edition ed., J. A. Rice, Ed. Thomson Brooks Cole, 2007.

[17] N. Draper and H. Smith, *Applied Regression Analysis*, N. Draper and H. Smith, Eds. John Wiley & Sons, Inc, 1998.

[18] W. Li, H. H. Yue, S. Valle-Cervantes, and S. Qin, "Recursive pca for adaptive process monitoring," *Journal of Process Control*, vol. 10, pp. 471–486(16), Oct 2000.

[19] K. A. Chamness, "Diagnostics of plasma etch: Pca with adaptive centering and scaling," in *AEC/APC XV Symposium*, 2003.