Virtual Metrology Modeling for CVD Film Thickness

Jérôme Besnard[3], Dietmar Gleispach[1,*], Hervé Gris[3], Ariane Ferreira[2], Agnès Roussy[2], Christelle Kernaflen[2], Günter Hayderer[1]

[1]austriamicrosystems AG Unterpremstätten, Austria
[2]Department of Manufacturing Science and Logistics Ecole Nationale Supérieure des Mines de Saint-Etienne Gardanne, France
[3]PDF Solutions, Montpellier, France

Abstract   The semiconductor industry is continuously facing four main challenges in film characterization techniques: accuracy, speed, throughput and flexibility. Virtual Metrology (VM), defined as the prediction of metrology variables using process and wafer state information, is able to successfully address these four challenges. VM is understood as definition and application of predictive and corrective mathematical models to specify metrology outputs (physical measurements). These statistical models are based on metrology data and equipment parameters. The objective of this study is to develop a model predicting the CVD oxide thickness (average) for an IMD (Inter Metal Dielectric) deposition process using FDC data (Fault Detection and Classification) and metrology data. In this paper, two VM models are studied: one based on Partial Least Squares Regression (PLS) and one based on Tree ensembles. We will demonstrate that both models show good predictive strength. Finally, we will highlight that model update is key for ensuring a good model robustness over time and that an indicator of confidence of the predicted values is necessary too if the VM model has to be use on-line in a production environment.

Keywords   Advanced Process Control, CVD Oxide Thickness, Partial Least Squares Regression, Tree Ensembles, Semiconductor Manufacturing, Virtual Metrology, Model Update, Indicator of Confidence

1. Introduction

The semiconductor manufacturing industry has a large-volume multistage manufacturing system. To ensure high stability and high production yield, reliable and accurate process monitoring is required[1]. Advanced Process Control (APC) is currently deployed for factory-wide control of wafer processing in semiconductor manufacturing. The APC tools are considered to be the main drivers to guarantee a continuous process improvement[2].

However, most APC tools strongly depend on the physical measurement provided by metrology tools[3]. Critical wafer parameters are measured, such as, for example, the thickness and/or the uniformity of thin films. If a wafer is misprocessed in an early stage but detected at the wafer acceptance test, unnecessary resource consumption is unavoidable. Measuring every wafer's quality after each process step could avoid late wafer scraps but it is too expensive and time consuming. Therefore, metrology, as it is employed for product quality monitoring today, can only cover a small fraction of sampled wafers. Virtual metrology (VM) in contrast enables prediction of every wafer's metrology measurement based on production equipment data and previous

metrology results[4-7, 27]. This is achieved by defining and applying predictive models for metrology outputs (physical measurements) as a function of metrology and equipment data of current and previous steps of fabrica tion[8-10,28-31].Of course it is necessary to collect data from equipment sensors to characterize physical and chemical reactions in the process chamber. Sensor data will constitute the basis for the statistical models that will be developed. A typical Fault Detection and Classification (FDC) system collects on-line sensor data from the processing equipment by sensors for every wafer or batch. They are called process variables or FDC data. Reliable and accurate FDC data are essential in VM model [11]. The objective of a VM module is to develop a robust prediction that can provide estimation of metrology and which is able to handle process drifts whether they are induced by preventive maintenance actions or not.

This paper deals with the prediction of PECVD (Plasma Enhanced Chemical Vapor Deposition) oxide thickness for an Inter Metal Dielectric (IMD) layers using FDC and metrology data. Two types of mathematical models are studied to build VM modules for PECVD processes. Partial Least Squares Regression (PLS)[12-13] and a non-linear approach based on Tree ensembles[14-16] are considered. The technical challenge and innovation are to build a single robust model, either with PLS or Tree ensembles, which is valid for several products, different layers and two different chucks. The alternative would be to make a model per layer, chuck and product, but we strongly believe that the maintenance of

many single models, in our case 12 different models, is not compatible with the constraints of the industry.

Section II deals with fabrication process. In section III we present the mathematical background to build VM models. Results are described in sections IV. Some perspectives about what the next steps of this work will be are given in section V. Finally, section V concludes this paper with a summary.

1.1. Fabrication Process

The film layer under investigation for thickness modeling is part of the IMD used in the Back-End of Line (BEOL) of a 0.35μm technology process. This oxide layer is used three times during the production of a four metal layer device. PECVD USG (Undoped Silicon Glass) films are commonly used to fill the gaps between metal lines due to their conformal step coverage characteristics. However, as the device geometry is shrinking, the gap fill capability of USG films is no longer sufficient. State of the art technique is the combination of HDP (High Density Plasma) and USG films to provide a high-productivity and low-cost solution. HDP is used to fill the gap just enough to cover the top of the metal line and then the USG is used as a cap layer on top of the HDP oxide film[17].
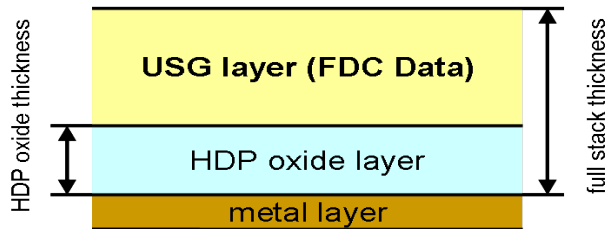

Figure 1. Layer structure for inter metal dielectric

Figure 1 shows the layer structure for one inter metal layer just before the Chemical Mechanical Polishing (CMP) step. The process steps are identical for all three stages. We use identical equipment production recipes, identical metrology setup and identical cleaning procedures for all three stages in the process flow. After metal deposition and structuring (lithography and etch) the HDP oxide is deposited. The HDP oxide film thickness is then measured by ellipsometry, using a 9-site template recipe. FDC data for VM modeling are collected during the USG deposition right after the HDP process. The full oxide stack is measured by the same ellipsometer tool also using a 9-site template recipe.

A schematic drawing of the process flow can be seen in figure 2. To guarantee the collection of a proper set of data within a few weeks, ten wafers per lot are measured before and after the USG process. The objective of the VM model is to use the predictive results as an input parameter for the following CMP process step. The CMP tool uses this input data for calculating the polishing time of each wafer and therefore the integrated layer thickness measurement could be skipped. The benefit of this VM implementation is an increased throughput and cost reduction.
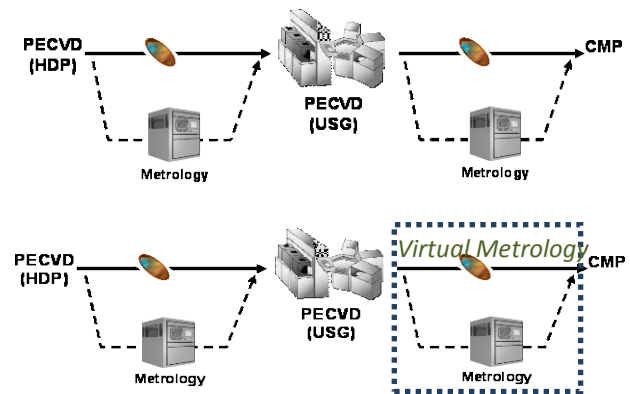

Figure 2. Process flow

Wafers are processed in a twin-chamber of a PECVD tool. The same deposition recipe is used for the deposition of different inter-metal layers and several products. During wafer processing, the relevant process parameters that characterize the PECVD process, such as gas flows, pressure, temperature plasma parameters, etc., are gathered. These temporal data are then consolidated with statistical methods. The temporal data (sensors) are collected at a sampling rate of 0.5 second. If too many samples are missing during the data collection, the data are discarded and the wafer is not used in the VM modeling. The temporal data are then transformed into the so-called FDC Indicators. A FDC Indicator is the summarization of temporal data into a single point, based on a given algorithm (mean, range, maximum, minimum, slope, etc). A data set consists of data from production equipment (input data X for VM modeling) and metrology equipment (output data Y for VM modeling). To assure the quality and effectiveness of VM models it is necessary to do preliminary quality studies of process and metrology equipments like variance analysis or repeatability and reliability studies (R&R studies). In addition, context information like layer, product and chuck is essential as categorical input for VM modeling.

Input data X consists of 24 indicators and three contextual variables. The output variable Y represents the average of the PECVD oxide thickness of each wafer.

2. Mathematical Models

2.1. Notation

The following notation conventions are used in this paper: scalars are designated using lowercase italics. Vectors are generally interpreted as column vectors and are designated using bold italic lowercase (i.e. x). Matrices are shown in bold italic uppercase (i.e. X), where x_{ij}, with ($i=1,\ldots, I$) and ($j=1,\ldots, J$), is the ij^{th} element of $X(I \times J)$. Let X of $\Re^p$ be an input data set and Y of $\Re^m$ be arranged in the following way:

$$X = \begin{pmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_I^T \end{pmatrix} \quad Y = \begin{pmatrix} y_1^T \\ y_2^T \\ \vdots \\ y_I^T \end{pmatrix}$$

where $x_i^T = (x_{i,1},\dots,x_{i,p})$ and $y_i^T = (y_{i,1},\dots,y_{i,m})$. The characters *I, J, N, p, q, m* and *n* are reserved for indicating the dimension of vectors and matrices of data.

2.2. VM Modeling

There are some important points when designing the mathematical models and a methodology that should be considered. In this section we propose two successive stages to deploy mathematical models in order to build a VM Module for an individual process:

2.2.1. Data partitioning: Training set and Test Set

Let X ($I \times p$) and Y ($I \times m$) be the available data set (cleaned and normalized) respectively from production and metrology process. The data set partitioning consist in the extraction of two units: a unit of 70% of the data set for the training-validation (training and cross validation) and a unit of 30% of the data set for the test. Let X_N ($N \times p$) and Y_N ($N \times m$) be the training-validation data set, and let X_N ($n \times p$) and Y_N ($n \times m$) be the test data set with $N+n=I$. It is possible to split the available data set in a temporal way (chronological selection) without loss of representativeness. In this case study we have chosen this type of data partitioning before the application of the three mathematical models.

Alternatively, the Kennard-Stone method [15] can be used to perform the data set partitioning. The inputs variables domain X of $\Re^p$, $x_i^T = (x_{i,1},\dots,x_{i,p})$ is considered for the Kennard-Stone method.. It is a sequential method to select a training set uniformly which covers the entire X variable's space. The selection criteria use the Euclidean distance.

2.2.2. Mathematical Modeling

A linear regression model of a given process can be written as:

$$Y = XB + E \qquad (1)$$

where X is the matrix of input data, Y is the matrix of output data, B is the matrix of regression coefficients and E is the matrix of errors whose elements are independently distributed with mean zero and variance σ^2[18-19]. Linear or non linear regression methods can be applied to the matrices X and Y to compute the coefficient matrix B. The regression model is built in two levels: the Training-validation level with the training-validation data set and the test level with the test data set. The training of models that are linear with respect to their parameters (such as linear regressions, polynomials models) can be performed easily with the traditional least-squares method, whereas the training of models that are nonlinear with respect to their parameters (such as neural networks) requires more complex methods. More details about the training of mathematical models can be found in[16].

Global approaches to model selection in the training-validation level are Cross-validation[20] and Leave-One-Out, methods for estimating generalization error based on resampling[21]. It is obvious to perform the model selection on the basis of the Validation Root Mean Square Error on the Training-validation data set (*VRMSE*). The *VRMSE* is given by equation (2):

$$VRMSE = 100 \times \left(\frac{1}{N} \sum_{i=1}^{N} \left(\frac{y_i - \hat{y}_i}{y_i} \right)^2 \right)^{1/2} \qquad (2)$$

where yi is the measured output value, ŷi is the estimated output value from the model, and N is the size of the training data set. The VRMSE is often used for comparing various models. In n-fold Cross-Validation the data are divided into n subsets of (approximately) equal size. The net is trained n times, where one of the training subsets is left out. Only the omitted subset is used to compute the error criterion of interest. If n is equal to the sample size it is called leave-one-out cross-validation. The prediction performance of the selected model is estimated using the test data set. The performance indicator is the Test Root Mean Square Error of Prediction (*TRMSE*) computed on the test data set:

$$TRMSE = 100 \times \left(\frac{1}{n} \sum_{i=1}^{n} \left(\frac{y_i - \hat{y}_i}{y_i} \right)^2 \right)^{1/2} \qquad (3)$$

where y_i is the measured output value, $\hat{y}_i$ is the estimated output value from the model, and n is the size of test data set.

2.3. PLS Models

Consider a set of historical process data consisting of an ($I \times p$) matrix of process variable measurements (FDC data) X and a corresponding ($I \times m$) matrix of metrology data Y. Projection to Latent Structures or Partial Least Squares (PLS) can be applied to the matrices X and Y to estimate the coefficient matrix B in (1).

$$\hat{Y}^{PLS} = X\hat{B}^{PLS} + E \qquad (4)$$

where $\hat{Y}^{PLS}$ is the PLS estimate of the process output Y. PLS modeling consists of simultaneous projections of both the X and Y spaces on low dimensional hyper planes of the latent components. This is achieved by simultaneously reducing the dimensions of X and Y, by seeking q ($< p$) latent variables which mainly explains covariance between X and Y. Therefore this method is useful to obtain a group of latent variables which explain the variability of both, Y and X. The latent variable models for linear spaces are given by Equations (5) and (6) [12]:

$$X = TP^T + E \qquad (5)$$
$$Y = TQ^T + F \qquad (6)$$

where E and F are error terms, T is an ($I \times A$) matrix of latent variable scores, and P ($p \times A$) and Q ($m \times A$) are loading matrices that show how the latent variables are related to the original X and Y variables. The sample covariance matrix is $X^T Y Y^T X$. The first PLS latent variable $t_1 = Xw_1$ is the linear combination of the X-variables that maximizes the covariance between t_1 and the Y space. The first PLS

weight vector w_1 is the first eigenvector of the sample co-variance matrix X^TYY^TX. After the scores for the first component have been computed, the columns of X are regressed on t_1 to give a regression vector:

$$\mathbf{p}_1 = \frac{\mathbf{X}^T\mathbf{t}_1}{\mathbf{t}_1{}^T\mathbf{t}_1} \tag{7}$$

In NIPALS (Nonlinear estimation by iterative Partial Least Squares) algorithm[13] the second latent variable t_2, orthogonal to t_1, is calculated from the new matrix of co-variance $X_2{}^TY_2Y_2{}^TX_2$, where X_2 and Y_2 are calculated by the equations (8) and (9):

$$\mathbf{X}_2 = \mathbf{X} - \mathbf{t}_1\mathbf{p}_1^T \tag{8}$$

$$\mathbf{Y}_2 = \mathbf{Y} - \mathbf{t}_1\mathbf{q}_1^T \tag{9}$$

q_1 is obtained by regression of the columns of Y in t_1, i.e.:

$$\mathbf{q}_1 = \frac{\mathbf{t}_1^T\mathbf{Y}}{\mathbf{t}_1{}^T\mathbf{t}_1} \tag{10}$$

The second latent variable is computed by the equation $t_2=Xw_2$, where w_2 is the first eigenvector of the sample co-variance matrix $X_2{}^TY_2Y_2{}^TX_2$, and so on. The new latent vectors or scores ($t_1, t_2,…$) and the weight vectors ($w_1, w_2,…$) are orthogonal. The final models for X and Y are given by Equations (5) and (6).

Latent variable models assume that both the process and metrology data spaces are observed with error and that both are effectively of very low dimension (i.e. non-full rank). The dimension A of the latent variable space is often quite small compared with the dimension of the process data space, and it is determined by cross-validation or some other procedure. Effectively, these models reduce the dimension of the problem through a projection of the high-dimensional X and Y spaces onto the low-dimensional latent variable space T, which contains most of the important information[12].

2.4. Tree Ensemble Models

It has been shown by Breiman et al.[22], in the classification case, that under reasonable assumptions, an ensemble procedure allows getting accurate models. Indeed, if the base model has a low-bias and high variance under some random perturbation of the learning conditions, then aggregating a large family of such models give birth to a low-bias, low variance aggregated model, that is more accurate than the individuals models[15].

To allow such results to hold, it is critical that the individual models are as independently built as possible, while maintaining low bias. Tree base learners, either based on algorithms such as CART[22] or C4.5[23], are known to have a low bias when fully learned (no pruning)[24]. In order to be able to build families of trees that have a low correlation to one another, from a finite dataset, several methods have been proposed: Bootstrapping the learning set (also known as bagging methods), Random splits, Injecting random noise in the response or building random artificial features as (linear) combinations of the existing ones. All these ideas aim at learning trees that are as uncorrelated as possible.

Following Breiman[22], we use here a combination of bagging from the base learning set, and random splits as our main ensemble method. Base learners are regression trees, following a modified CART algorithm for tree learning. Given X, a set of $(I \times p)$ FDC data, and a corresponding $Y (I \times m)$ metrology, and 2 parameters q (random selection among features at the individual split level) and *nTrees (number of trees grown and aggregated)*, the algorithm is as follows:

1. Iterating over the m responses:
2. Looping 1 -> *nTrees*:
a) Build a bootstrap sample $(I \times p)$ X^b and corresponding Y^b $(I \times 1)$ response
b) Build a fully-grown tree τ, following modified CART algorithm
i. randomly selecting $q<m$ candidates for a given split inside a node
ii. Select the best split among the q candidates as the one that reduces most the residual variances over the 2 children nodes
iii. Recursively until stopping criterion is reached, i.e. node is pure (internal variance equal 0)
3. Average the predictions, i.e.

$$\hat{y} = \frac{1}{nTrees} \sum_{i=1}^{nTrees} \tau_i(\mathbf{X}) \tag{11}$$

Bagging allows the calculation of the so-called out-of-bootstrap prediction, which is very similar to cross-validation or leave-one-out, since predictions on the learning set are derived by averaging, for each individual, the set of trees in which this individual is not in the bootstrap sample. Hence bagged ensembles have an internal estimation of their generalization error. A well-known property of trees is their inability to model linear effects, which is why, when a strong linear effect from one or several parameters is discovered in the data, we build the tree ensemble on the residual from the main linear effect.

Finally, tree ensembles have internal estimations of the importance of each of the feature that are calculated by averaging their out-of-bootstrap contribution in the prediction for each tree. More precisely, one can estimate the increase in out-of-boostrap error that scrambling one parameter would produce, over several repetition of the scrambling procedure. This allows dropping low importance features from the model by comparing features importance to probes (random features). In the end, the model will be:

$$\hat{y} = \mathbf{A}\mathbf{X}^{lin} + \frac{1}{nTrees} \sum_{i=1}^{nTrees} \tau_i(\mathbf{X}^{nlin}) \tag{12}$$

where *lin* and *nlin* are disjoint subsets of the initial set of indicators. R^2 is defined by equation (18):

$$R^2 = 1 - \frac{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{S_y^2} \tag{13}$$

where y_i is the measured output value, $\hat{y}_i$ is the estimate output value by model, N is the size of training data set and

S_y^2 is the variance of y. This metric is estimated from the out-of-bootstrap predictions.

3. Results

In this section we present the results of two different models (PLS and tree ensembles) for prediction of the PECVD oxide film thickness.

3.1. PLS Models with 3 Qualitative Variables

A PLS model is built without input variable selection. The model is calibrated with 306 wafers of the training validation data set. The 168 wafers of the test data set are used to validate the model. The input variables of the PLS model are the 24 quantitative FDC indicators and the three qualitative variables which are chuck, layer and product. The PLS model with four principal components is selected by cross-validation. Q^2(cum) is increasing for the first four principal components and decreasing for the fifth principal component. Actually, using the first four principal components, 46.5% of the X variability (quantitative and qualitative variables) can explain 89.7% of the output Y (average thickness) variability (see table I). Therefore the best statistical model is achieved by using only four principal components.

Table 1.   Result table for the principal components of PLS

Principal Component	$R^2\mathbf{X}$(cum)	$R^2\mathbf{Y}$(cum)	Q^2(cum)
1	16.2	75.6	74.2
2	29.5	85.5	83.7
3	39.0	88.3	85.7
4	46.5	89.7	86.1

Analyses have been done on each parameter to quantify its importance. In table II the five most important variables can be found.

Table 2.   Variable importance of PLS model

Variable	VIP (4PC)
X24	2.16
X10	1.89
X3	1.77
X17	1.61
X25	1.60

The trained model is applied to the test data set. In figure 3 the measured and the predicted average oxide film thickness for PLS model are shown. The *VRMSE* and *TRMSE* are around 0.53% and 0.58% of the average thickness, respectively. Figure 3 shows the predicted average oxide film

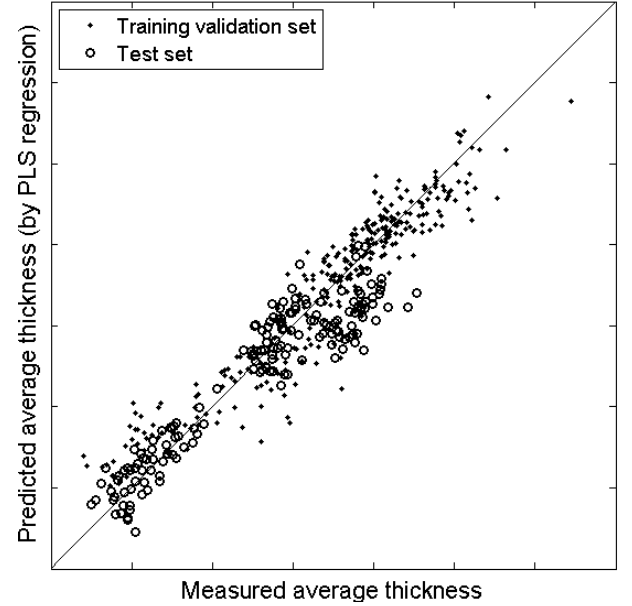thickness using the results of PLS model versus the measured average thickness.

Figure 3.   Predicted average thickness (by PLS model) versus measured average thickness representation

3.2. Tree Ensembles Model

Modeling is done using the learning set of 306 wafers. After one iteration of the algorithm, one indicator (X10) is selected for the prelinear part; the remaining 26 are left for tree ensemble modeling, including the three qualitative parameters chuck, product and layer. The second iteration of the algorithm provides a model that selected five parameters: X10 in the linear part of the model, and four in the tree ensemble model (X9, X8, X24 and X25). X24 and 25 are two qualitative parameters (see table III). R^2 is estimated to be 0.84, being defined as equation (13).

Table 3.   Ranking of indicators for the tree ensemble model

Rank	Variable	model
1	X10	Linear
2	X24	Tree ensemble
3	X25	Tree ensemble
4	X9	Tree ensemble
5	X8	Tree ensemble

This metric is estimated from the out-of-bootstrap predictions. Other model quality metric include VRMSE, estimated also from the out-of-boostrap predictions, measured at 0.69% of the average thickness for this model.

The model is then used to predict average oxide film thicknesses for the test set. Figure 4 shows the result of the tree ensembles model. The TRMSE is comparable to the VRMSE and is equal to 0.59% (see table IV). R^2 is calculated to be 0.84 also, for the test set.
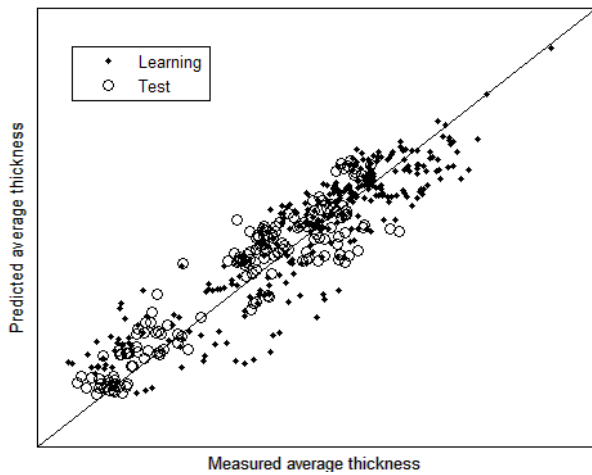
Figure 4. Predicted average thickness (by tree ensemble model) versus measured average thickness representation

Table 4. Tree ensemble model summary

Quant. variables	Qual. variables	Selected	R²	VRMSE (%)	TRMSE (%)
22	3	5	0.84	0.69	0.59

4. Perspectives

In this paper we have been comparing two different types of mathematical modeling (*PLS* vs. *Tree Ensembles*). This academic study can be considered as the first step for virtual metrology, and demonstrates that here, with reasonable parameter selection; different algorithms yield similar results in terms of prediction capacity. So modeling capacity will contribute a low part to the algorithmic choices to be made in virtual metrology, we think. However, two main challenges remain to be addressed before using online virtual metrology prediction in Fab environment. The first one is to ensure VM models will be robust enough with time; this has to deal with model update approaches. The second one is to guarantee that predicted values can be trustfully used; this will be addressed by developing an indicator of confidence for each predicted value. The ultimate goal is to provide a reliable prediction that can be used for the CMP step.

The model robustness over time is a key topic that must not be neglected. Many factors can impact the model validity such as the chamber aging, a sudden chamber malfunction, as well as the unscheduled and scheduled preventive maintenances [25]. All these events might lead to a change with time of some collected variables, used as model input on the form of FDC indicators. It is difficult assessing ahead the impact that such changes might have on the model validity. For all these reasons, a static model, built on a leaning dataset with no further updates, doesn't seem to be a sustainable solution. One should prefer an approach based on dynamic models. In that case, many possibilities exist such as a regular update, a data-driven update (based on estimated

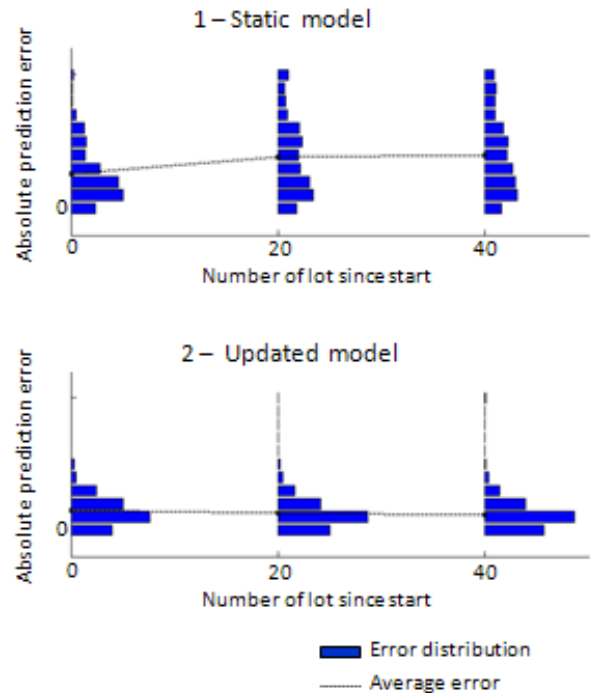quality of the model) and a chamber-driven update (based on maintenance events).

Figure 5. Comparing the evolution of errors distribution between updating and non-updating models

The second challenge is to provide a predictive quality index in order to guarantee the accuracy of the model prediction. This quality index could be compared to the GOF (goodness-of-fit) available for each measure done in a metrology tool. The quality Index will be a combination of two metrics. In the case of *Tree Ensembles* modeling, a pre-selection of all available indicators is done; only indicators having a major contribution in predicting the metrology output are kept in the model and use for calculated the predicted metrology values. The others with no or less influence on the metrology output are left aside. However, if one of them changes drastically suddenly, as it might happen when a chamber malfunction is detected, the VM predicted value might be questionable. Indeed, this parameter might have a strong influence on the metrology though this parameter was not kept in the model due to a constant value over time, for instance. Should this happen, the corresponding quality metrics should reflect it. The second metric is related to the quality of the model prediction itself and can roughly be describes as an R- squared (R2). The combination of these 2 metrics should give an indication precise enough to determine whether or not the predicted value can be used in the feed-forward control loop for CMP removal step [26]. If the confidence in the predicted value is too low, thus the wafer should go with no doubt through a real measurement in the metrology tool.

5. Conclusions

This paper presents two mathematical models that have been used to develop virtual metrology for predicting the average oxide film thickness deposited during a PECVD process. The two models have good predictive strength. PLS and tree ensemble show equivalent performance on the test set, but PLS shows slightly better results on the learning set. This could be explained by the elimination of an outlier value (point with highest measure thickness) in learning set of PLS model. PLS uses four principal components, which are based on all the variables. The tree ensemble model uses five variables only. Three out of the top five most important variables of PLS are used in the tree ensemble model.

The predictive results are in excellent agreement with the measured data. In addition, we have shown as a novelty in virtual metrology that it is possible to create a single model for different layer, different products and for one chamber with two different chucks.

The first results we have had on model update techniques, as well as on building a predictive quality index are very encouraging to reach our final goal which is to have Virtual Metrology running online in the CMP step.

ACKNOWLEDGMENTS

REFERENCES

[1]   A.C. Diebold, "Overview of metrology requirements based on the 1994 National Technology Roadmap for semiconductor". Advanced Semiconductor Manufacturing Conference and Workshop 1995, ASMC 95 Proceedings. IEEE/SEMI 1995, November 1995, pp. 50–60.

[2]   J.R. Moyne. "Making the move to fab-wide APC", Solid State Technology, vol. 47, no. 9, September 2004, pp. 47-52.

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

[4]   Y.-J. Chang, Y. Kang, C.-L. Hsu, C.-T. Chang and T. Y. Chan. "Virtual Metrology Technique for Semiconductor Manufacturing", Proceedings of the Conference on Neural Networks, Sheraton Vancouver Wall Center Hotel, 2006, pp. 5289–5293.

[5]   P. H. Chen, S. WU, J. Lin, F. Ko, H. Lo, J. Wang, C. H. Yu, and M. S. Liang. "Virtual Metrology: A solution for wafer to wafer advanced process control", Proceedings of the IEEE

International Symposium on Semiconductor Manufacturing, September 2005, pp.155–157.

[6]   F.-T. Cheng. "Researching Strategy and Development Proposal of e-Manufacturing". Automation Division of National Science Council, Taiwan, R.O.C, October 2004.

[7]   F.-T. Cheng, H.-C. Huang, and W.-M. Wu "Dual-Phase Virtual Metrology Scheme", IEEE Transactions on Semiconductor Manufacturing, vol. 20, no. 4, November 2007, pp. 566–571.

[8]   M.-H. Hung, T.-H. Lin, P.H. Chen and R.-C. Lin. "A novel virtual metrology scheme for predicting CVD thickness in semiconductor manufacturing", IEEE/ASME Transactions on Mechatronics, vol. 12, no. 3, June 2007, pp. 364–375.

[9]   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, November 2007, pp. 364–375.

[10]   T.-H. Lin, M.-H. Hung, R.-C. Lin and F.-T. Cheng. "A virtual metrology scheme for predicting CVD thickness in semiconductor manufacturing". Proceedings of the 2006 IEEE International Conference on Robotics and Automation, May 2006, pp. 1054–1059.

[11]   Y.-C. Su, T.-H. Lin, F.-T. Cheng, and W.-M. Wu. "Accuracy and Real-Time Considerations for Implementing Various Virtual Metrology Algorithms". IEEE Transactions on Semiconductor Manufacturing, n. 21, vol. 3, August 2008, pp. 426–434.

[12]   T. Kourti, "Application of latent variable methods to process control and multivariate statistical process control in industry," Int. J. Adapt. Contr. Signal Process, vol. 19, no. 4, 2005, pp. 213–246.

[13]   M. Tenenhaus, La Régression PLS Théorie et Pratique, Editions Technip, Paris, 1998.

[14]   L. Györfi, M. Kohler, A. Krzyzak, and H. Walk, A Distribution Free Theory of Nonparametric Regression, Springer-Verlag, New York 2002.

[15]   R.W.Kennard and L.Stone. "Computer Aided Design of Experiments", Technometrics, no. 11, 1969, p.137-148.

[16]   G. Dreyfus. Neural Networks Methodology and Applications. Hardcover, 2002.

[17]   S. P. Muranka, M. Eizenberg, A.K. Sinha. Interlayers Dielectrics for Semiconductor Technologies, Elsevier edition, 2003.

[18]   S. Chen, S. A. Billings, and W. Luo. "Orthogonal least squares methods and their application to non-linear system identification", International Journal of Control, vol. 50, no. 5, 1989, pp. 1873 – 1896.

[19]   A. C. Rencher, Methods of multivariate Analysis, Hardcover 2002.

[20]   M. Stone, "Cross-validatory choice and assessment of statistical predictions", Journal of the Royal Statistical Society, B 36, 1974, pp 111-147.

[21]   B. Efron and R.J. Tibshirani. "Improvements on cross-validation: The bootstrap method", Journal of the American Statistical Association, vol. 92, 1997,pp.548-560.

[22] L. Breiman, "Arcing classifiers", The Annals of Statistics, vol. 26, no.3, 1998, pp. 801-849.

[23] J. R. Quinlan, C4.5: Programs for Machine Learning, Morgan Kaufmann, 1993.

[24] J. H. Friedman and P. Hall. "On Bagging and Nonlinear Estimation", Journal of Statistical Planning and Inference, vol. 137, no.3, March 2007, 669 – 683.

[25] D. Zeng and C. J. Spanos. "Virtual Metrology Modeling for Plasma Etch Operations" IEEE Transactions on Semiconductor Manufacturing, n. 4, vol. 22, November 2009, pp. 419–431.

[26] C-A. Kao, F-T. Cheng, W-M. Wu. "Preliminary study of Run-to-Run Control utilizing Virtual Metrology with Reliance Index", IEEE Conference on Automation science and Engineering, 2011, Trieste, Italy

[27] F. T. Cheng, J. Chang, H. Huang, C. Kao, Y. Chen, and J. Peng. "Benefit model of virtual metrology and integrating AVM into MES." IEEE Transactions on Semiconductor Manufacturing, 24(2):261-272, 2011

[28] B. S. Gill, T. F. Edgar, and J. D. Stuber. "A novel approach to virtual metrology using Kalman Filtering." Future Fab International, 35:86-91, 2010

[29] P. Kang, H. J. Lee, S. Cho, D. Kim, J. Park, C. K. Park, and S. Doh. "A virtual metrology system for semiconductor manufacturing." Expert Systems with Applications, 36(10):12554-12561, 2009

[30] C. Shan, P. Tianhong, and J. ShiShang. "Development of a virtual metrology for high-mix TFT-LCD manufacturing processes." Journal of Semiconductors, 31(11):116006/1-116006/5, 2010

[31] E. Ragnoli, S. McLoone, S. Lynn, J. Ringwood and N.Macgearailt. "Identifying key process characteristics and predicting Etch Rate from High Dimensionality Datasets" Proc. IEEE/SEMI Advanced Semiconductor Manufacturing Conference (ASMC), Berlin, 2009