Plasma Etch Process Virtual Metrology using Aggregative Linear Regression

PKS Prakash
The Irish Centre for Manufacturing Research
National University of Ireland, Maynooth,
Co. Kildare, Ireland
Email: prakash@eeng.nuim.ie

S. F. McLoone
Department of Electronic Engineering
National University of Ireland, Maynooth,
Co. Kildare, Ireland
Email: sean.mcloone@eeng.nuim.ie

Abstract—**To enhance product quality semiconductor manufacturing industries are increasing the amount of metrology information collected during manufacturing processes. This increase in information has provided companies with many opportunities for enhanced process monitoring and control. However, the increase in information also posses challenges as it is quite common now to collect many more measurements than samples from a process leading to ill-conditioned datasets. Ill-conditioned datasets are very common in semiconductor manufacturing industries where infrequent sampling is the norm. It is therefore critical to be able to quantify virtual metrology models developed from such data sets. This paper presents an aggregative linear regression methodology for modeling that allows the generation of confidence intervals on the predicted outputs. The aggregation enhances the robustness of the linear models in terms of process variation and model sensitivity towards prediction. Also, to deal with the large number of candidate process variables, variable selection methods are employed to reduce the dimensionality and computational efforts associated with building virtual metrology models. In the paper three methods for variable selection are evaluated in conjunction with aggregative linear regression (ALR). The proposed methodology is tested on a benchmark semiconductor plasma etch process dataset and the results are compared with state-of-art multiple linear regression (MLR) and Gaussian Process Regression (GPR) VM models.**

Keywords: Virtual metrology, Aggregative Linear Regression, Forward Stepwise Regression, Decision Trees.

I. INTRODUCTION

Plasma etching is a key process in modern semiconductor manufacturing facilities used to achieve precise control of wafer etching. Plasma etching offers process simplification and improved dimensional tolerances compared to wet-chemical-etching technology. The flowchart of a plasma etching process is shown in Fig. 1. The process involves introducing an appropriate mixture of gases such as $AlCl_3$ and Cl_2 into a vacuum chamber followed by ionization of gases using high power microwave frequencies (MF) to generate a plasma [1]. Radio frequencies (RF) accelerate the generated plasma toward the electrode where it interacts with the masked wafer surface both chemically and mechanically to etch away the exposed surface. The chemical composition of the exhaust gases from the chamber can be monitored using optical emission spectroscopy (OES) [2]. The challenge with operating plasma etchers is maintaining a consistent etch rate spatially and temporally for a given wafer and for successive wafers processed in the same etch tool. Etch rate variations occur for a variety of reasons including chamber seasoning effects due to chemical interactions with the chamber wall, temperature changes in the chamber during the etching step, non-uniformity in the composition of the plasma gases and variability in the RF current discharge [3]. The complex nonlinear behaviour of plasma etch processes and their sensitivity to disturbances makes them very difficult to model and control. Further, etch rate measurements are not accessible in real-time. They can only be obtained through a costly non-value added post-etch metrology step leading to significant delays before they are available for process adjustment. Consequently, in practice, plasma etch processes are operated in open-loop fashion using pre-determined fine-tuned recipes with a small number metrology measurements performed to facilitate process monitoring and statistical process control.

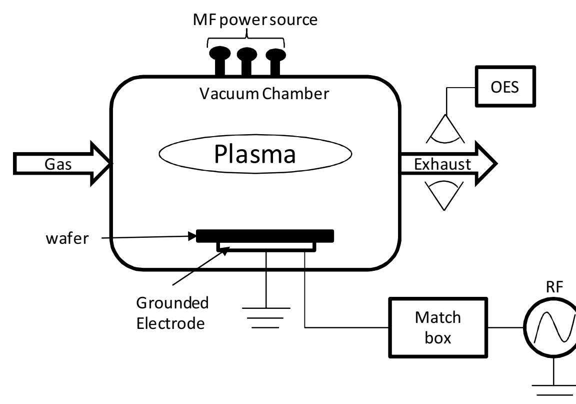

Figure 1: Flow diagram of a plasma etch process

Recently, virtual metrology (VM) has gained attention in the semiconductor manufacturing community as a cutting edge solution for estimating critical process quality parameters from other more accessible in-line process measurements [4]. This approach, also referred to as soft-sensing or inferential estimation, offers the potential to significantly enhance yield and improve process capability in semiconductor

manufacturing. PingHsu *et al.* [5] have demonstrated the application of VM tools for advance process control in wafer manufacturing and have shown a process capability (C_p) improvement of more than 65% using a VM model integrated with advance process control (APC).

Many researchers have used data mining based approaches to build virtual metrology (VM) models for estimating critical parameters in manufacturing processes. These methods use historical process measurements to build the relationship between the process variables and critical parameters. The trained models are then used to estimate the critical parameters on-line in a real-time control environment. All the methods used in VM models can be broadly classified as either linear or non-linear. Linear models include methods such as linear regression, principle component regression (PCR) and partial least square regression (PLS) [6][7]. Non-linear VM models involve methods such as non-linear regression and neural networks (NN) [8]. Ringwood *et al.* [9] summarizes different methods for building VM models for semiconductor etch processes based on statistical and machine learning techniques. Most of the research carried out to-date on VM focuses on determining estimates of critical parameters such as etch rate, with little consideration given to quantifying the reliability of the estimates produced. To understand and take account of the estimation accuracy of VM models, methodologies that generate estimates with confidence intervals (CI) are needed.

This paper proposes an aggregative linear regression (ALR) methodology for generating VM estimates of etch rate with associated confidence intervals. ALR is a multi-model approach for building VM models that helps to reduce the sensitivity of models to process variation and inherent noise. Two aggregation methods: (i) bagging; and (ii) Adaboost are explored in this research. Both these aggregation strategies are tested using linear regression as the basic component model. In modern semiconductor etching processes optical emission spectroscopy (OES) is common place as a source of in-line process information on plasma and exhaust gas composition leading to many thousands of candidate input variables for VM models. Thus, identification of a small number of key variables which can model the etch rate accurately is also a critical issue. This paper evaluates three different methods for selecting input variables for ALR, namely, correlation, forward stepwise regression and decision tree analysis. The developed methodology is evaluated on a benchmark semiconductor plasma etch process dataset and the results obtained with ALR are compared with state-of-art multiple linear regression (MLR) and Gaussian Process Regression (GPR) VM models.

The remainder of the paper is organized as follows: Section II presents the variable selection methods. Section III details the ALR methodology for etch rate estimation and also provides an overview of MLR and GPR. Confidence interval determination and performance metrics are then presented in Section IV. To illustrate and validate the proposed methodology comparative results are presented for a benchmark plasma etch process dataset in Section V. Finally, conclusions are summarized in Section VI.

II. VARIABLE SELECTION

The quality of the final product from a manufacturing process is usually dependent on many process variables. Identification of a few variables which are most influential and which can be used to model the system is a challenging problem, especially if a large number of process variables are collected during processing and only a limited number of post-process sample measurements are available leading to multi-collinearity issues with the dataset. Variable selection methods are needed to address this issue. This paper evaluates three different variable selection methods for determining input variables associated with a target response.

A. *Correlation based variable selection*

The correlation based methodology selects critical variables based on linear correlation of the input vector $\mathbf{x}_i$ with response vector $\mathbf{y}$ where i =1, 2, 3... r and r is the number of input variables. The correlation between input vector $\mathbf{x}_i$ and response $\mathbf{y}$ is defined as

$$\rho_i = \frac{E((\mathbf{x}_i - \mu_{\mathbf{x}_i})(\mathbf{y} - \mu_{\mathbf{y}}))}{\sigma_{\mathbf{x}_i}\sigma_{\mathbf{y}}}, \quad (1)$$

where, $\rho_i \in [-1 \;\; 1]$ represents the correlation coefficient for the ith input vector and $\mu_{\mathbf{x}_i}$ and $\mu_{\mathbf{y}}$ are the mean of the input and response vectors, respectively.

Since correlation only measures the linear relationship between the input and response vector, it is possible that there may be a stronger non-linear relation between them. This can be investigated by applying non-linear transformations to the data before conducting the correlation analysis. These transformations may be as simple as powers and cross-products or as complicated as spline relationships. In this research log and power transformations were performed when testing the correlation between the input and response vectors. The input variables with the highest correlation with the response are selected for building VM models.

B. *Stepwise Regression based variable selection*

Stepwise regression can be performed using three different approaches: Forward selection, Forward stepwise regression (FSR), and backward elimination [10]. Among all the approaches, forward stepwise regression is the most popular in the literature [11]. Forward stepwise regression is a combination of both the forward selection and backward elimination methods. Thus, variables selected into a model do not necessary stay there. As in the forward selection method variables in FSR are added one by one into the model based on a selection criteria such as the F-test, mean square error (MSE), residual square (R^2), adjusted coefficient of multiple determination ($Ajd R^2$), Akaike information criterion (AIC), Hannan and Quinn information criterion (HQ), and the Bayesian information criterion [12]. However, unlike the forward selection method, FSR also looks at all the variables already included in a model and deletes any variables that are no longer significant based on the selection criteria.

In this research, F-test is used as the selection criteria. The F-statistic is computed using Eq. (2) when assessing the significance of a variable being added to a model,

$$F = \frac{RSS_1 - RSS_2}{RSS_2} \times (n - p + 1), \quad (2)$$

where, RSS_i is the residual sum of squares of model i and (n-$p+1$) represents the degrees of freedom associated with the larger model with n representing sample size and p the number of variables in the previous model. The p-value is computed based on an F-distribution with degrees of freedom (1, n-$p+1$) and a variable is selected based on a 95% confidence interval, *i.e.*, p=0.05.

C. Decision tree based variable selection

Piecewise constant regression decision tree analysis is used to identify variables which can predict the response **y** [13]. The steps involved in constructing a decision tree are: (i) fitting a model to the training data; (ii) identifying the variable with the most significant chi-square statistic based on the residue of each predictor; (iii) determining the best split for the selected variable based on a suitable loss function; and (iv) pruning the developed tree on the basis of 10-fold cross validation. The loss function used in this research to fit the model is the least-squared deviation. The least square deviation LSD(η) at node η is computed as:

$$LSD(\eta) = \frac{1}{N_w(\eta)} \sum_{i \in \eta} w_i f_i (y_i - \bar{y})^2, \quad (3)$$

where, $N_w(\eta)$ represents the weighted number of cases at node η, w_i and f_i represent the weight and frequency, respectively associated with case i and response y_i. The weighted mean for node η, is denoted by $\bar{y}$. The variables identified by the regression tree are used as the inputs to a VM model.

III. MODELLING ETCHING DATA

This section discusses the 3 modelling techniques for modelling the etching data: (i) Multiple linear regression; (ii) Aggregative linear regression; and (iii) Gaussian process regression.

A. Multiple Linear Regression

In multiple linear regression a linear relationship is assumed to exist between the response **y** and input variables **X** and is defined as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad (4)$$

where $\mathbf{y} \in \Re^n$ is the response for n samples and $\mathbf{X} \in \Re^{n \times v}$ represents the input vectors with v variables and n samples. The vector $\boldsymbol{\beta} \in \Re^v$ represents the regression coefficients for the input variables. The regression coefficients $\boldsymbol{\beta}$ are determined based on minimization of the distance between the predicted and actual etch rate as shown in Eq. (5)

$$\min_{\beta} \|\mathbf{X}\boldsymbol{\beta} - \mathbf{y}\|_2 \quad (5)$$

However, the objective function defined in Eq. (5) becomes ill-posed when **X** is ill-conditioned or singular, *i.e.*, $v > n$ leading to imprecise results [14]. To deal with this scenario the pseudo-inverse is used in which $\boldsymbol{\beta}$ is selected such that it is the minimum norm solution to Eq. (5), i.e., $\min \|\boldsymbol{\beta}\|_2$ subject to satisfying Eq. (5) [14].

To enhance the robustness of linear regression prediction with respect to process variation and model sensitivity aggregation methods can be employed to improve the prediction accuracy and confidence in the solution. The approach referred to as aggregative Linear Regression (ALR) is discussed in the next sub-section.

B. Aggregative Linear Regression (ALR)

The ALR methodology utilizes multiple instances of the linear model to enhance the model robustness to noise and reduce the impact of data dependency. Two widely used aggregation methods are Bagging [15] and AdaBoost [16].

Bagging (also known as Bootstrap aggregation) uses bootstrapping on the training dataset to create many overlapping instances with equal probability of selection. The algorithm then creates a different model instance for each bootstrap sample and the predicted response $\bar{\bar{\mathbf{y}}}$ is the average of the predictions of all model instances. For a given input vector $\mathbf{x} \in \Re^v$ this is computed as

$$\bar{\bar{y}} = \frac{1}{m} \sum_{i=1}^{m} \hat{y}_i, \quad (6)$$

where $\hat{y}_i$ is the predicted response from the ith regression model for the input data **x**.

AdaBoost (adaptive boosting) based model aggregation is similar to the bagging approach except that a weighting is applied to the individual model predictions to obtain the overall predicted response. Thus, AdaBoost prediction $\bar{\bar{y}}$ can be written as

$$\bar{\bar{y}} = \sum_{i=1}^{m} w_i \hat{y}_i, \quad (7)$$

where, w_i represents the weight assigned to i^{th} model. The weights w_i are selected to reflect the overall confidence in the individual models based on their training set performance. Machine learning techniques can be used to generate weights that vary as a function of the input variables (local weights). Alternatively input-independent ‘global’ weights can be employed.

In this research, the Bagging strategy for aggregation is used to determine the etch rate estimates and build confidence interval around the estimates.

C. Gaussian process regression

Williams and Rasmussen [17] proposed Gaussian process regression (GPR) for high dimensional problems. GPR does not model for a specific structure, *f(x),* instead a Gaussian prior is placed on a range of possible functions that could map input **X** to **y** as

$$\mathbf{y} = f(x_1, x_2 .. x_v) + \varepsilon \quad (8)$$

where $\varepsilon \sim N(0, \sigma_\varepsilon^2)$ is Gaussian white noise with zero mean and σ_ε^2 variance. The range of possible functions for **y** are evaluated based on a covariance function as

$$y_i \sim N(0, \mathbf{K})$$

where $\mathbf{K} = \mathbf{\Sigma} + \sigma_\varepsilon^2 \mathbf{I}$ and the covariance for the input variables is determined using Eq. (9)

$$k(x_i, x_j) = m^2 e^{(\frac{-\|x_i - x_j\|^2}{2\tau^2}}, \quad (9)$$

where, m and τ are the hyperparameters of the covariance function. The hyperparameters are identified by maximizing the log marginal likelihood given by

$$\log(P(\mathbf{y} \mid \mathbf{X})) = -\frac{1}{2}\mathbf{y}^T \mathbf{K}^{-1}\mathbf{y} - \frac{1}{2}\log(|\mathbf{K}|) - \frac{1}{2}\log(2\pi) \quad (10)$$

Based on the optimized hyperparameters the posterior distribution of $\mathbf{y}_*$ for given $\mathbf{X}_*$, i.e., $P(\mathbf{y}_* | \mathbf{X}_*, \mathbf{y}, \mathbf{X})$ is determined using Eq. (11-12). Here $\{\mathbf{y}_*, \mathbf{X}_*\}$ are unseen data and $\{\mathbf{y}, \mathbf{X}\}$ is the dataset used to train the Gaussian model.

$$\mu(y_*) = \mathbf{k}_*^T \mathbf{K}^{-1}\mathbf{y} \quad (11)$$

$$\sigma^2(y_*) = k_{**} - \mathbf{k}_*^T \mathbf{K}^{-1}\mathbf{k}_* + \sigma_\varepsilon^2 \quad (12)$$

Column vector $\mathbf{k}_*$ is the covariance between the training and testing data points and k_{**} is the auto covariance of the given input. The mean and variance of the posterior distribution of $\mathbf{y}_*$ is given by $\mu(y_*)$ and $\sigma^2(y_*)$.

IV. Building Confidence Intervals (CI)

This section describes methods for building confidence intervals around the estimates determined using the methods discussed in Section III.

A. CI using Multiple Linear Regression (MLR)

This confidence interval is built using the standard error estimates obtained from the training and test data. The standard error (σ_ε) is determined as

$$\sigma_\varepsilon = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_{tst,i} - y_{tst,i})^2}{n-1}}, \quad (13)$$

where, $\mathbf{y}_{tst}$ and $\hat{\mathbf{y}}_{tst}$ are the true and estimated values of the test dataset, respectively. Using σ_ε the CI around the estimates is defined as

$$CI = \hat{y}_* \pm Z_{1-\alpha}\sigma_\varepsilon \quad (14)$$

where, $\hat{\mathbf{y}}_*$ represents the estimated response for given $\mathbf{X}_*$ for unseen data. Scalar $Z_{1-\alpha}$ is the upper critical value of the normal distribution for given confidence level α.

B. CI using Aggregative Linear Regression (ALR)

The confidence interval for ALR is estimated as

$$CI = \bar{\bar{y}} \pm Z_{1-\alpha}\hat{\sigma}(\mathbf{x}_*) \quad (15)$$

where $\hat{\sigma}(\mathbf{x}_*)$ is defined as

$$\hat{\sigma}^2(\mathbf{x}_*) = \sigma_{\text{ALR}}^2(\mathbf{x}_*) + \sigma_\varepsilon^2$$

The term $\sigma_{\text{ALR}}^2(\mathbf{x}_*)$ represents the variance of the prediction y_* for the given input vector $\mathbf{x}_*$ across the m ALR models, that is;

$$\sigma_{\text{ALR}} = \sqrt{\frac{1}{m-1}\sum_{j=1}^{m}(\hat{y}_{*j} - \mu_{\hat{y}_*})^2}$$

and σ_ε is the standard error as determined using Eq. (13).

C. CI using Gaussian Process Regression (GPR)

The CI predicts the distribution of y_* for given $\mathbf{x}_*$ based on the training samples. The mean ($\mu(y_*)$) and variance ($\sigma^2(y_*)$) predicted by the GPR can be used to build the confidence interval around y_* as

$$CI = \mu(y_*) \pm Z_{1-\alpha}\sigma(y_*) \quad (15)$$

D. Performance metrics for CI

The confidence intervals determined using Eqs (12-15) are evaluated on unseen data using two criteria: (i) percentage of actual data points violating the confidence interval; and (ii) the conservativeness of the confidence interval as measured by the confidence interval performance index (CIPI),

$$CIPI = \frac{1}{2}\sum_{i=1}^{p}\frac{|UCI_i - LCI_i|}{|\hat{y}_{*,i} - y_{*,i}|} \quad (16)$$

Here $\hat{y}_{*,i}$ and $y_{*,i}$ are the estimated and actual response, respectively for the ith data point and UCI_i and LCI_i are the upper and lower limits of the corresponding confidence interval. If CIPI<1 the model has under estimated the CI, while a value greater than 1 indicates that the model has over estimated the CI (*i.e.* wider than required).

V. Case Study

The case study employed is a VM dataset consisting of OES signatures and corresponding actual etch rates for a sequence of wafers processed on an industrial plasma etch tool. The OES dataset consists of 6 statistical moments, mean,

variance, skewness, kurtosis, maximum and minimum amplitude for the time series for each of 2048 wavelengths recorded during the etch step. For the most part, sampling of wafers processed on the tool was performed uniformly for each lot of 25 wafers. A total of 2194 sampled wafers are considered in this study for training and testing of VM models. A subset of 1894 samples are used for training and validation of the models with the remaining 300 samples retained for testing performance of the VM model confidence intervals.

The goal is to develop a model that can predict etch rate using only the OES measurements. For the considered dataset, estimation of the number of input variables required by each model is carried out using the variable selection approaches discussed in Section II. An extensive study was carried out to understand the affect of the number of variables selected on VM model performance, the results of which are reported in Table 1-3. These show the normalized mean square error (NMSE) performance of MLR, ALR and GPR as function of the number of input variables selected for each of the three variable selection methods considered.

Table 1: Training and test dataset NMSE performance for Correlation based variable selection

	Linear		ALR		Gaussian Regression	
Nv	**Training**	**Validation**	**Training**	**Validation**	**Training**	**Validation**
1	16.05	15.93	16.00	16.00	16.41	15.43
50	9.21	11.12	9.46	10.89	6.71	12.05
100	6.02	8.55	6.25	8.57	1.19	10.08
150	5.00	7.38	4.84	7.77	0.12	8.98
200	3.64	6.63	3.58	6.91	0.15	9.34
250	2.80	6.93	2.79	6.71	0.16	8.79
285*	2.46	6.02	2.34	6.61	0.10	8.95
296*	2.15	6.70	2.25	6.50	0.10	8.80
300*	1.99	7.36	2.21	6.66	0.12	7.91
350	1.96	7.49	1.78	6.92	0.11	8.34

Table 2: Training and test dataset NMSE performance for FSR based variable selection

	Linear		ALR		Gaussian Regression	
Nv	**Training**	**Validation**	**Training**	**Validation**	**Training**	**Validation**
1	15.64	16.57	16.20	16.28	16.69	15.43
50	3.59	3.59	3.30	3.78	0.20	4.95
51*	3.32	3.73	3.23	3.75	0.23	4.35
100	1.77	2.83	1.97	2.70	0.16	6.10
150	1.34	2.37	1.41	2.27	0.16	5.06
200	1.04	2.08	1.08	2.09	0.12	6.45
250	0.82	1.95	0.83	1.95	0.11	6.34
300	0.64	1.89	0.64	1.89	0.13	6.34
342*	0.56	1.56	0.50	1.84	0.14	6.8
346*	0.47	1.74	0.49	1.82	0.11	6.41
350	0.49	1.71	0.48	1.85	0.09	6.74

Table 3: Training and test dataset NMSE performance for decision tree based variable selection

	Linear		ALR		Gaussian Regression	
Nv	**Training**	**Validation**	**Training**	**Validation**	**Training**	**Validation**
1	88.40	87.79	88.17	88.02	80.51	84.44
50	10.73	13.06	10.95	12.74	0.08	13.48
100	7.54	10.23	7.32	10.34	0.08	10.33
150	5.98	10.50	5.61	9.54	0.11	9.69
200	4.96	9.59	4.96	9.92	0.11	10.15
250	4.23	9.41	4.04	9.85	0.14	10.04
300	3.16	10.14	3.06	9.42	0.13	9.42
332*	2.30	7.41	2.35	8.47	1.31	9.56
340*	2.16	7.71	2.11	8.01	0.14	8.9
341*	2.29	7.50	2.10	7.99	0.12	9.67
350	1.98	8.93	2.05	8.08	0.15	9.29

The number of variables that yields the minimum validation error for each model is highlighted in each Table. An example of the trend in validation error for different VM models where variables are selected using FSR is shown in Fig. 1. As can be seen MLR and ALR begin to over fit after 340 variables, while the more flexible GP model over fits the dataset at 50 variables.

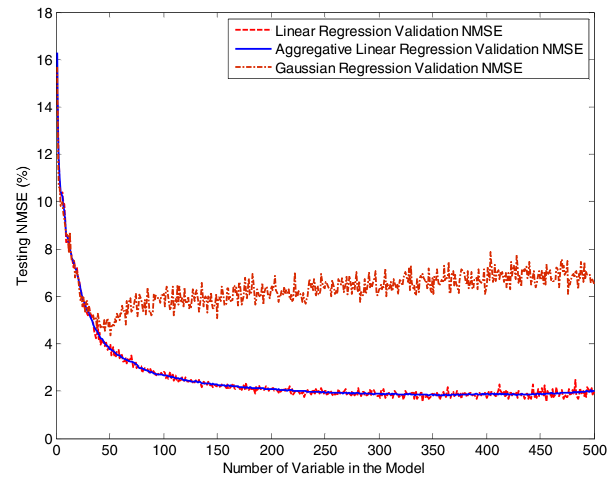

Figure 1: Validation error as a function of the number of FSR input variables for different VM models

Table 4: Performance of confidence interval estimates

Variable Selection Method	VM Models	Performance			
		Nv	**% violation**	**CIPI**	**% NMSE**
Correlation	**Linear**	285	6	2.49	9.6
	ALR	296	1.33	3.06	10.33
	GPR	300	5.33	2.27	9.69
FSR	**Linear**	342	5.33	2.39	2.55
	ALR	346	0	3.75	2.2
	GPR	50	10.33	1.99	9.02
Decision Tree	**Linear**	332	4.33	2.41	10.52
	ALR	350	1.33	3.13	14.75
	GPR	340	19.33	1.69	13.68

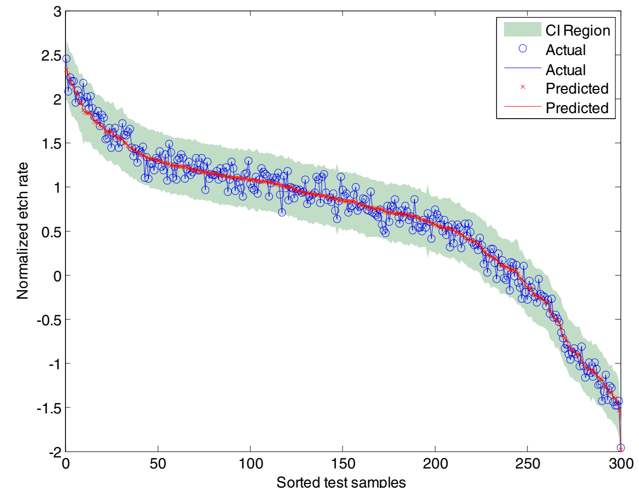

Figure 2: ALR predictions and confidence intervals for the optimum FSR-ALR VM model.

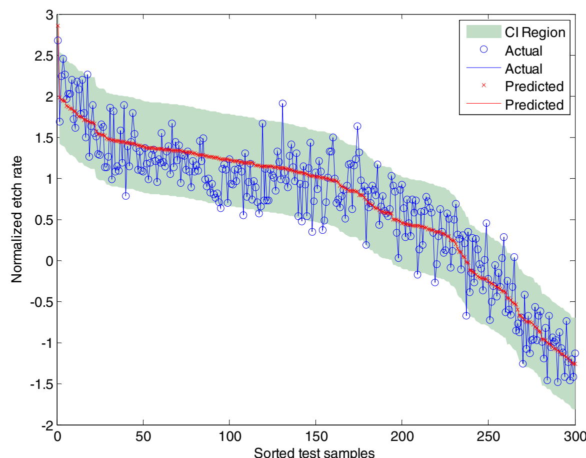

Figure 3: GPR predictions and confidence intervals for the optimum correlation-GPR VM model.

VM models were developed and optimised for each of the variable selection methods in accordance with Table 1-3. These models where then used to estimate the etch rate and corresponding 95% CI (α=0.05) for 300 previously unseen wafers and their performance evaluated in terms of the, NMSE, the CIPI and the percentage of samples violating the CI. The results obtained are summarized in Table 4. By way of example, the confidence intervals for the optimum FSR-ALR and Correlation-GPR models are plotted in Fig. 2 and 3, respectively. These plots show the CI and predicted and actual etch rate (normalised) for the test wafers, which have been sorted with respect to predicted etch rate to facilitate visualisation and assessment of the quality of the CI.

The results show that the ALR models perform much better in terms of estimating and building confidence intervals than the GPR method for the etch process data.

VI. CONCLUSIONS

This paper presents Aggregative Linear Regression (ALR) based VM models for predicting etch rate with confidence intervals. As many manufacturing processes involve large numbers of process variables and restricted access to samples the impact of variable selection processes on model prediction estimates and confidence intervals has also been considered in this work. Three methods are considered namely, correlation, forward step-wise regression (FSR), and decision tree based variable selection. Results demonstrate that the VM model developed using FSR based variables in ALR outperforms the VM models build using input variables selected based on correlation or decision tree methods.

The paper also presents a systematic approach to evaluating the confidence intervals generated by models. Etch rate predictions with confidence intervals are presented for three different VM models; Multiple Linear Regression (MLR), ALR, and Gaussian process regression (GPR). The results show that the ALR model with variables selected using FSR performs best in terms of percentage of CI violations, however this is at the cost of overly conservative CI estimates. GPR produces much tighter CI intervals around the estimates in general but has a much higher percentage of points falling outside the confidence bounds, but this may also be linked to the overall poor performance of GPR for this dataset. The superior performance of ALR and MLR compared to GPR is likely a reflection of the strength of the linear relationship between the OES measurements and etch rate**.**

ACKNOWLEDGMENTS

The authors gratefully acknowledge the financial support of the Irish Centre for Manufacturing Research (www.icmr.ie/) and Enterprise Ireland (grant CC/2010/1001).

REFERENCES

[1] W.G.M. van den Hoek and T. Mountsier, “ A New High Density Plasma Source for Void Free Dielectric Gap Fill, Technical Proceedings of the 1994 SEMI Technology Symposium (chiba, Japan), 1994, pp.195-200.

[2] Splichal M., Anderson H., 1987, "Application of Chemometrics to Optical Emission Spectroscopy for Plasma Monitoring," Proc. SPIE, vol. 2, pp. 189-203.

[3] Sugawara M., and Stansfield B., 2002, "Plasma etching: fundamentals and applications," Oxford Science Publication.

[4] Y. J. Chang, Y. Kang, C.-L. Hsu, C.-T. Chang and T. Y. Chan, 2006, "Virtual Metrology Technique for Semiconductor Manufacturing," Neural Networks, 2006, IJCNN' 06, International Joint Conferance on 2006, pp. 5289-5293.

[5] PingHsu C., Wu S., Junshien L., Ko F., Wang J., Yu C. H., and Liang M.S., 2005, " Virtual metrology: a solution for wafer to wafer advanced process control," IEEE International Symposium on Semiconductor Manufacturing, (ISSM 2005), 13th-15th Septemeber 2005, pp. 155 - 157.

[6] Lynn S., Ringwood, J., Ragnoli E., McLoone S., and MacGearailt N., 2005, "Virtual Metrology for Plasma Etch using Tool Variables," Advanced Semiconductor Manufacturing Conference, 2009, 10-12 May 2009, pp. 143 - 148.

[7] Chen R., Huang H., Spanos J., and Gatto M., 1996, "Plasma etch modelling using optical emission spectroscopy," J. Vac. Sci. Tech. A, Vac. Surf. Films, Vol. 14(3), pp. 1901-1906.

[8] Khan A., Moyne J., Tilbury D., 2007, "An Approch for Factory-Wide Control Utilizing Virtual Metrology," IEEE Tran. on Semi. Man. Vol. 20(4), pp. 364-375.

[9] Ringwood John V., Lynn S., Bacelli G., Ma B., Ragnoli E., and Mcloone Sean, 2010, "Estimation and Control in Semiconductor Etch: Practice and Possibilities," IEEE Trans. on Semiconductor Manufacturing, vol. 23(1), pp. 87-98.

[10] Barrett, B. E. and Gray, J. B. (1994), "A Computational Framework for Variable Selection in Multivariate Regression", Statistics and Computing, 4, 203-212.

[11] Neter, J., Kutner, M., Nachtsheim, C., and Wasserman, W. (1996), "Applied Linear Statistical Models," McGraw-Hill Companies, Inc., NY.

[12] Rencher, A. C. (1995), "Methods of Multivariate Analysis," John Wiley & Sons Inc., New York, New York.

[13] W.-Y. Loh. "Regression trees with unbiased variable selection and interaction detection," Statistica Sinica, 12:361–386, 2002.

[14] Bjorck A (1996) Numerical Methods for Least Squares Problems. SIAM, Philadelphia, PA.

[15] Breiman, L., “Bagging Predictors”, Machine Learning, Vol. 24, No. 2, pp. 123–140, 1996.

[16] Freund,Y. and R. E. Schapire, “A Decision-Theoretic Generalization of On-line Learning and an Application to Boosting”, European Conf. on Computational Learning Theory, pp. 23–37, 1995.

[17] C. K. I. Williams and C. E. Rasmussen. Gaussian Processes for regression, Chapter 8, pages: 514-520. MIT Press, 1996.