Proceedings of the 2018 Winter Simulation Conference
M. Rabe, A.A. Juan, N. Mustafee, A. Skoogh, S. Jain, and B. Johansson, eds.

VIRTUAL METROLOGY MODELING BASED ON GAUSSIAN BAYESIAN NETWORK

Wei-Ting Yang
Jakey Blue
Agnès Roussy

Department of Manufacturing Science & Logistics
Center of Microelectronics in Provence
Ecole des Mines de Saint-Etienne
880 route de Mimet
Gardanne, 13541, FRANCE

Marco S. Reis

CIEPQPF - Department of Chemical Engineering
University of Coimbra
Polo II
Rua Sílvio Lima
Coimbra, 3030-790, PORTUGAL

Jacques Pinaton

Department of Process Control
STMicroelectronics Rousset
190 Avenue Coq Monnet,
Rousset, 13106, FRANCE

ABSTRACT

For the past decades, Virtual Metrology (VM) has been widely studied and covered in the literature for semiconductor industries where cycle time is a critical aspect and the elimination of non-productive metrology measurements is expected to significantly contribute to its reduction. A wide variety of approaches has been proposed but not effectively implemented. An ideal VM model should be able to provide accurate predictions and to reveal the hidden relationship among production/process factors. For this aim, we employ the Gaussian Bayesian Network (GBN) to investigate the implicit relationship not only between the metrology and the control factors but also among the production/process parameters. Instead of working purely as a black-box data-driven methodology, GBN enables the flexibility to integrate domain knowledge through the corresponding connected graph. The effectiveness of proposed approach is validated using real industrial data from a Chemical-Mechanical Polishing (CMP) process.

1 INTRODUCTION

In semiconductor manufacturing, the stability of the process and the corresponding quality of wafer are always of primary concern. However, considering the limited metrology tool capacity and restricted wafer life cycle, only a few sampled wafers are actually sent for inspection. In order to circumvent these limitations of metrology and expand the feasibility of process quality monitoring, virtual metrology (VM) approaches have been developed (Chen et al. 2005). Virtual metrology models aim to predict metrology measurements of the wafers based on Fault Detection and Classification (FDC) data comprising abundant information collected from process machines by embedded sensors, including both physical and chemical parameters which reflect the actual state during the process for all wafers. With such process information, VM firstly align data with the sampled wafers and identify the relationship between FDC information and metrology data by machine learning models. Considering different hidden data properties and model characteristics, various modeling methods have been studied and evaluated, such as linear regression, partial least squares, neural networks, k-nearest neighbors, regression trees, etc. (Chang et al. 2005; Khan et al. 2007; Kang et al. 2009; Zeng and Spanos 2009). With the VM system in the production line, not only the monitoring mechanism can be improved, but also the overall control can be improved, providing furthermore valuable information for abnormal wafer detection, run-to-run control, and predictive maintenance (Khan et al. 2008; Kang et al. 2011).

The selection of important factors from the numerous FDC parameters is a crucial task, and therefore most studies consider algorithms able to perform feature selection as part of their learning procedure (Chang et al. 2005; Khan et al. 2007; Kang et al. 2009; Zeng and Spanos 2009), while other studies consider penalty-based approaches like Lasso and the Elastic Net (Pampuri et al. 2011; Susto et al. 2013). In this paper, we propose to use of Gaussian Bayesian Networks (GBN) as a learning algorithm for VM modeling. This methodology present the ability to select features while considering their conditional dependency. In other words, GBNs are able to explicitly express the relationship among the variables in a form of the connected graph (Heckerman and Geiger 1995; Cowell, R.G. et al. 1999), making the interpretation easier. In recent years, GBNs which have been applied in various domains, e.g., in genetics, where GBNs were used to explore the conditional dependency among the huge amount of genes (Bühlmann et al. 2010). Quite surprisingly, to the best of our knowledge, GBNs have not been used yet to derive VM models for semiconductor manufacturing. Therefore, the purpose of this paper is to demonstrate the potential of application of this approach in VM applications on semiconductor industry, through a real case study.

The remainder of the paper is organized with the following structure. In Section 2, the concept of GBN and its main properties are presented. In Section 3, we introduce the procedure for VM using GBN. The case study is presented in Section 4, illustrating the application of the proposed approach. Finally, the conclusions and future work are summarized in Section 5.

2 GAUSSIAN BAYESIAN NETWORKS

2.1 Foundations of Gaussian Bayesian Networks

A Bayesian Network is a probabilistic model expressing the conditional dependencies of a set of variables by a directed acyclic graph (DAG) (Spirtes et al 1993). Bayesian Networks can help to investigate the cause-effect relationship since the relationship between two variables could be asymmetric, e.g., X causes Y while Y cannot cause X; a directed graph can better describe this relationship by considering X as a parent node and Y as a child node (Figure 1).

Figure 1: An example of directed graph.

Gaussian Bayesian Networks (GBN) are a special case of Bayesian Networks, in which all the variables are continuous and the joint density function is a multivariate Gaussian distribution. Let us first introduce the conditional distribution and then consider the graphic model that schematically describes such probability distribution. Let $\mathrm{D} = [X_1, X_2, \ldots, X_q]^T$ be a n by q matrix with n samples and q random variables, that follows a multivariate Gaussian distribution, $\mathcal{N}(\mu, \Sigma)$, with mean vector $\mu \in \mathbb{R}^q$ and covariance $\Sigma \in \mathbb{R}^{q \times q}$.

The graph $\mathcal{G} = (\mathrm{V}, \mathrm{E})$, is composed by the set $\mathrm{V} = \{V_1, V_2, \ldots, V_q\}$ of nodes (vertices): each $V_k \in \mathrm{V}$, for $k = 1, \ldots, q$, indicates one of the random variables in D. Here, a random variable X_k is also referred to the node V_k in graph $\mathcal{G}$. E is the set of directed edges and each $e \in E$ indicates the link between two nodes. Absence of an edge represents the existence of conditional independence between the corresponding variables. The global distribution of $\mathcal{G}$ can be decomposed into the local distribution of individual variables X_k as $p(X_1, X_2, \ldots, X_q) = \prod_{k=1}^{q} p(X_k | X_{pa(k)})$, where $p(X_k | X_{pa(k)})$ is the conditional probability of X_k given its parent nodes $X_{pa(k)}$.

GBNs present the Markov property. Let $mb(X_k)$ denotes the Markov blanket of node X_k which is the set that includes its parent nodes, its child nodes and the parent nodes of its child nodes (Cowell, R.G. et al. 1999). The Markov blanket of X_k contains all variables that shield X_k from the rest of nodes, which means that $mb(X_k)$ is the only knowledge needed to predict the behavior of X_k.

2.2 Structure Learning

In practice, the structure of a Bayesian network can either be provided by a domain expert or be extracted from data. As our objective in this study is to reveal the hidden structure of variables, a data-driven structure learning approach will be adopted.

Learning the structure of DAGs can be difficult and computational intensive, because the set of possible DAGs is enormous. There are two main categories of approaches to learn the graphical structure from data: constraint-based and score-based (Scutari 2010). The constraint-based algorithms identify the conditional independencies of all variables with statistical tests to determine if each edge exists. Note that the outcome of this method is affected by the order of testing.

The score-based algorithms firstly score each possible graphical structure based on how well it describes the observed data, and then search for the graphical structure with the highest score. Considering a set of models $\mathcal{G}_s$, each model $\mathcal{G} \in \mathcal{G}_s$ indicates a possible network structure. The probability of model $\mathcal{G}$ given data $\mathcal{D}$ can be written as

$$p(\mathcal{G}|\mathcal{D}) = \frac{p(\mathcal{D}|\mathcal{G})p(\mathcal{G})}{p(\mathcal{D})}, \tag{1}$$

since $p(\mathcal{D})$ is independent of model $\mathcal{G}$, we can say $p(\mathcal{G}|\mathcal{D}) \propto p(\mathcal{D}|\mathcal{G})p(\mathcal{G})$.

A Bayesian score is defined as $Score_B(\mathcal{G},\mathcal{D}) = \log p(\mathcal{D}|\mathcal{G}) + \log p(\mathcal{G})$. Generally, we assume $p(\mathcal{G})$ is a uniform distribution and it can be ignored in this expression. From a Bayesian learning point of view, the parameters are random variables denoted as $\theta_{\mathcal{G}}$, therefore, the marginal probability in Bayesian scores can be rewritten as $p(\mathcal{D}|\mathcal{G}) = \int p(\mathcal{D}|\mathcal{G},\theta_{\mathcal{G}})p_{\mathcal{G}}(\theta_{\mathcal{G}})d\theta_{\mathcal{G}}$, where $p_{\mathcal{G}}(\theta_{\mathcal{G}})$ is prior distribution for its parameter $\theta_{\mathcal{G}}$. Some suggestions of prior distributions are given in literature (Cowell, R.G. et al. 1999).

The computation of the marginal likelihood is not simple because there is no closed form expression for it. The Bayesian information criterion (BIC) can be considered as an alternative which approximates the logarithm of $p(\mathcal{D}|\mathcal{G})$, by $Score_{BIC}(\mathcal{G},\mathcal{D}) = -\log(\hat{L}) - \frac{1}{2}d_{\mathcal{G}}\log n$, where $\hat{L} = p(\mathcal{D}|\mathcal{G},\hat{\theta}_{\mathcal{G}})$ is the maximum value of likelihood function, $\hat{\theta}_{\mathcal{G}}$ is the maximum likelihood estimate, $d_{\mathcal{G}}$ is the model complexity and n is sample size (Cowell, R.G. et al. 1999).

In this study, the most fundamental score-based search algorithm, Hill-Climbing, is employed to provide a greedy search on the feasible space of the directed graphs with BIC as a scoring function. Generally, Hill-Climbing starts from either an empty, full or random network, and then considers every possible movement of the current network, including to add an edge, to remove an edge, or to alter the direction of an edge; the movement with the highest score is selected and the procedure is repeated. The learning procedure stops when no improvement can be achieved by changing any single edge. In addition to identifying the direction of edges from data, the algorithm also provides the flexibility to integrate pre-defined directions for some specific edges; thus, we will include some rules based on the domain knowledge so that the model can fit properly.

3 APPLYING GBN TO VIRTUAL METROLOGY

3.1 Data Preprocessing

There are two types of datasets under analysis in this work: FDC and Metrology. Since these datasets present different levels of granularity, data preprocessing is surely needed and should be the very first step to begin with.

Multiple sensors embedded in the process equipment can collect real-time signals with high resolution. In order to transform temporal FDC data into wafer-based data, the conventional approach is to summarize the observations in each step by descriptive statistics. In this study, sample averages and variances were selected to be the summarizing indicators of each FDC parameter. Therefore, n wafers are processed and r parameters, for each one of the s steps, are collected by equipment sensors, leading to $p = s \times r \times 2$ indicators to be generated and denoted by x_{ij}, where $i = 1, .., n$ and $j = 1, \dots, p$.

To maintain production efficiency, only a few wafers will be sampled to take the measurements, such as thickness and depth, so that the process quality can be evaluated and monitored. We denote $y_j$as the measurement variable for wafer $i = 1, \dots, n$.

The two types of data are consolidated at the wafer level. All *p* FDC variables are put together with the metrology in a $(p + 1)$ by n matrix $D = [X_1, X_2, \dots, X_p, Y]$, which follows a multivariate Gaussian distribution $\mathcal{N}(\mu, \Sigma)$, where n is the number of wafers.

3.2 GBN for Dependency Exposure and Prediction

The first objective of the proposed approach is to build a GBN to better express the relationship between product quality (metrology) and equipment conditions (FDC). Here, instead of purely data-driven modeling, we would like to integrate some domain knowledge so that this model can describe the real phenomenon better. Considering the direction of edges in GBN indicate the causal effects between the variables, we should carefully exclude some edges which impossibly exist. As the measurement data can be considered as the output of the process, which means that those FDC parameters might be the cause of metrology, while metrology cannot be the cause of FDC parameters. Those can be done by setting up some rules in blacklist to specify the block directed edges. In other words, all directed edges from metrology to FDC parameters would be put in blacklist.

With the dataset and a pre-defined blacklist, we obtained a GBN $\mathcal{G}(\mathcal{D})$ by a heuristic learning algorithm, which expresses all pairwise conditional dependency for each node. With an example in Figure 2, we can clarify how those FDC parameters correlated to each other and identify the direct and indirect impact to metrology.

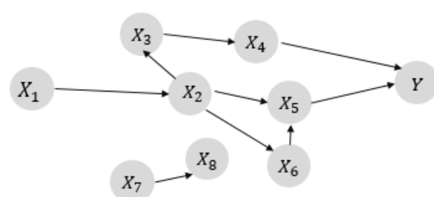

Figure 2: An example of Gaussian Bayesian Network.

Then, we choose the metrology variable y as a starting point and extract its corresponding Markov blanket, denoted as $mb(Y)$. As variables belong to $mb(Y)$ include the knowledge needed to predict Y, we thus have, $\hat{y} = f_Y(x_m)$, where $X_m \in mb(Y)$ (Figure 3).

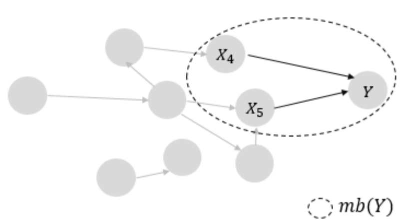

Figure 3: Markov blanket of variable Y.

4 CASE STUDY

To validate the proposed approach, a case study conducted on real data was carried out. The dataset includes FDC data from the Chemical–Mechanical Polishing (CMP) process and the metrology measurements after the process.

The typical CMP tools include a rotating platen with a pad, a carrier that holds wafer upside-down, a head to press the wafer against the pad, and a pad conditioner (Figure 4). Through the chemical slurry and mechanical force, unwanted materials on the wafer can be removed and the surface of the wafer will be smooth.

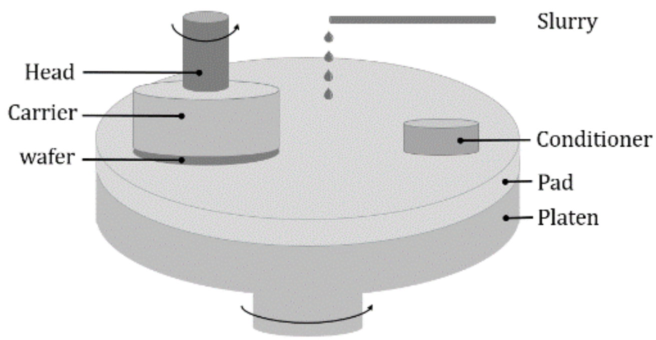

Figure 4: A typical CMP tool.

4.1 Data Description and Preprocessing

The FDC parameters in the dataset cover the platen usage, the head speed, the conditioning pressure, etc. A deterministic sampling policy was put in practice for the CMP progress. Two wafers per lot are sampled to take the metrology measurements, which are the thicknesses of post-polishing oxidation in association with the target T. Furthermore, it is known that the performance of the processes before CMP will inevitably influence the output of CMP. Two pre-CMP measurements, the post-etching depth and post-CVD thickness, denoted as Z_1 and Z_2, are also included in the variable set.

Before proceeding to model construction, a stationary check of metrology data indicated that this series is autocorrelated. Thus, a pseudo-variable that is lagged by one period of Y will be included in the variable set.

After pre-processing and consolidation of data granularity, 545 wafers with 149 variables are ready for analyzing. The data matrix can be written as $D = [X_{1(t)}, \dots, X_{145(t)}, Y_{(t)}, Y_{(t-1)}, Z_{1(t)}, Z_{2(t)}]$ where $Y_{(t-1)}$ are denoted as the metrology from the previous wafer. As discussed in the previous section, we should combine domain knowledge to set up the rules to prevent the unreasonable result. Firstly, metrology at time t cannot be the cause of other variable at time $t-1$; Similarly, pre-CMP measurement and FDC variables at time t cannot be the cause of metrology at time $t-1$. Those rules were listed in a blacklist as shown in Table 1.

Table 1: Blacklist of edge for GBN.

From	To	#edges
$Y_{(t)}$	$Y_{(t-1)}, X_{(t)}, Z_{1(t)}, Z_{2(t)}$	148
$\{X_{(t)}, Z_{1(t)}, Z_{2(t)}\}$	$Y_{(t-1)}$	147
$X_{(t)}$	$Z_{1(t)}, Z_{2(t)}$	290

4.2 Model Evaluation

The dataset has been ordered by process time, the first 80% wafers are labeled as training set and the remaining 20% are labeled as testing set. Then based on the training set and the blacklist setting, we obtain the model $\mathcal{G}(\mathcal{D})$. Following we extract the Markov blanket $mb(Y)$ as shown in Figure 5. This graph clearly depicts the relationship between the process variables and the metrology (Y). Both of the pre-CMP measurements, i.e., Z_1 and Z_2, have the direct impact on the metrology. Other nine variables also contribute to explain the variation within the metrology. The comparison of observed metrology and predicted metrology for both training set and testing set is presented in Figure 6.

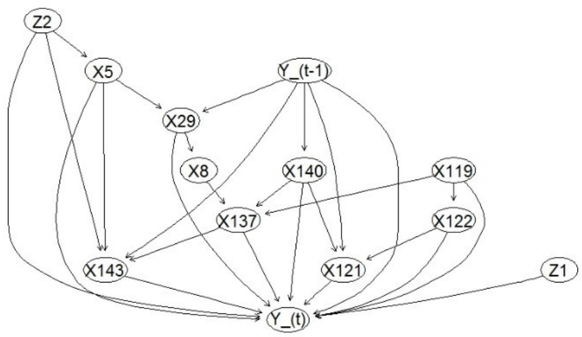

Figure 5: Markov blanket of metrology $\boldsymbol{Y}_{(\boldsymbol{t})}$ in the fitted GBN.

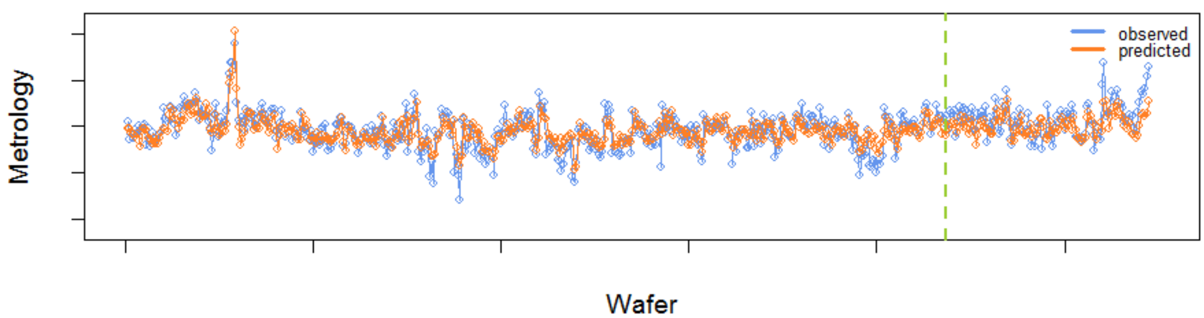

Figure 6: Compare the observed values and predicted values of the GBN model. Training set and testing set are separated by dash line.

Here, we evaluate the performance of the model with the index – Mean Absolute Percentage Error (MAPE). The MAPE of the training set and testing set are 0.467% and 0.562%, respectively.

$$MAPE\% = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{\widehat{r_i} - r_i}{r_i}\right| \tag{2}$$

We compare the result with other algorithms as Table 2, the performance of GBN, stepwise regression and Lasso regression are similar, but if we consider the number of selected variables, it seems like that GBN more capable of targeting efficiently crucial variables while maintaining the competitive predictive ability. Moreover, explicit cause-effect among variables enables process engineers to evaluate this VM model from a physical point of view.

Table 2: The performance evaluation of different models.

MAPE%	GBN	Stepwise Regression	Lasso Regression	Regression Trees	Random Forest
Training Set	0.467%	0.460%	0.506%	0.410%	0.379%
Testing Set	0.562%	0.558%	0.582%	0.635%	0.663%
#variables	12	21	16	70	--

5 CONCLUSION

In this paper, we introduce the Gaussian Bayesian Network (GBN) as a learning algorithm for VM modeling. GBN not only investigates the relationship between the metrology measurements and the control factors, but also provides an overall connection among the production/process parameters. Through a practical case study, we demonstrated that the model is capable of providing accurate predictions and, at the same time, clarifying the causal-effect among all the variables.

There are still remained works to be studied in the future. For example, employing a mixed graphical model which is able to deal with both discrete and continuous variables so that information, such as product type, recipe etc., can be taken into account. Additionally, comparing various learning algorithms and selecting the appropriate one for the purpose of virtual metrology should be well evaluated.

As Virtual Metrology is an important subject for process control, we firstly consider to apply GBN to this application as a starting point. However, there are various applications related to process control and should be considered simultaneously, such as run-to-run control, predictive maintenance. Instead of individual models, GBN shall be able to combine multiple applications in single learning model. Therefore, we will focus on applying GBN to global process control for the further study.

REFERENCES

Bühlmann, P., M. Kalisch, and M. H. Maathuis. 2010. "Variable Selection in High-dimensional Linear Models: Partially Faithful Distributions and the PC-simple Algorithm." *Biometrika* 97(2):261–278.

Chen, P., S. Wu, J. Lin, F. Ko, H. Lo, J. Wang, C. H. Yu, and M. S. Liang. 2005. "Virtual Metrology: A Solution for Wafer to Wafer Advanced Process Control." In *Proceedings of the IEEE International Symposium on Semiconductor Manufacturing 2005*, September 13th-15th, San Jose, USA, 155–157.

Cowell, R.G., P. Dawid, S. L. Lauritzen, and D. J. Spiegelhalter. 1999. *Probabilistic Networks and Expert Systems: Exact Computational Methods for Bayesian Networks*. New York: Springer-Verlag.

Heckerman, D. and D. Geiger. 1995. "Learning Bayesian Networks: A Unification for Discrete and Gaussian Domains." In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, edited by P. Besnard and S. Hanks, 274–284. San Francisco, California: Morgan Kaufmann Publishers Inc.

Kang, P., H. Lee, S. Cho, D. Kim, J. Park, C.-K. Park, and S. Doh. 2009. "A Virtual Metrology System for Semiconductor Manufacturing." *Expert Systems with Applications* 36(10):12554–12561.

Kang, P., D. Kim, H. Lee, S. Doh, and S. Cho. 2011. "Virtual Metrology for Run-to-Run Control in Semiconductor Manufacturing." *Expert Systems with Applications* 38(3):2508–2522.

Khan, A.A., J.R. Moyne, and D.M. Tilbury. 2007. "An Approach for Factory-Wide Control Utilizing Virtual Metrology." *IEEE Transactions on Semiconductor Manufacturing* 20(4):364–375.

Khan, A.A., J.R. Moyne, and D.M. Tilbury. 2008. "Virtual Metrology and Feedback Control for Semiconductor Manufacturing Processes Using Recursive Partial Least Squares." *Journal of Process Control* 18(10):961–974.

Pampuri, S., A. Schirru, G. Fazio, and G.D. Nicolao. 2011. "Multilevel Lasso Applied to Virtual Metrology in Semiconductor Manufacturing." In *Proceedings of the 2011 IEEE International Conference on Automation Science and Engineering*, August 24th-27th, Trieste, Italy, 244–249.

Spirtes, P., C. Glymour, and R. Scheines. 1993. *Causation, Prediction, and Search*. NY: Springer-Verlag.

Scutari, M. 2010. "Learning Bayesian Networks with the bnlearn R Package." *Journal of Statistical Software*, 35(3):1–22.

Susto, G.A., A.B. Johnston, P.G. O'Hara, and S. McLoone. 2013. "Virtual Metrology Enabled Early Stage Prediction for Enhanced Control of Multi-stage Fabrication Processes." In *Proceedings of the 2013 IEEE International Conference on Automation Science and Engineering (CASE)*, August 17th-20th, Madison, USA, 201–206.

Wan, J., S. Pampuri, P.G. O'Hara, A.B. Johnston, and S. McLoone. 2014. "On Regression Methods for Virtual Metrology in Semiconductor Manufacturing." In *Proceedings of the 25th IET Irish Signals Systems Conference 2014 and 2014 China-Ireland International Conference on Information and Communications Technologies (ISSC 2014/CIICT 2014)*, June 26th-27th, Limerick, Ireland, 380–385.

Yung-Cheng, J.C., and F.-T. Cheng. 2005. "Application Development of Virtual Metrology in Semiconductor Industry." In *Proceedings of the 31st Annual Conference of IEEE Industrial Electronics Society (IECON 2005)*, November 6th-10th, Raleigh, North Carolina, 124–129.

Zeng, D., and C.J. Spanos. 2009. "Virtual Metrology Modeling for Plasma Etch Operations." *IEEE Transactions on Semiconductor Manufacturing* 22(4):419–431.

AUTHOR BIOGRAPHIES

WEI-TING YANG received the B.S. degree in Statistics from National Cheng Kung University and the M.S. degree in Industrial Engineering from National Taiwan University,Taiwan, in 2007 and 2009, respectively, where she is currently pursuing the Ph.D. degree in Industrial Engineering. Her current research interests are statistical data mining and process control in semiconductor manufacturing. Her e-mail address is wei-ting.yang@emse.fr.

JAKEY BLUE is Associate Professor of Department of Manufacturing Sciences and Logistics at École des Mines de Saint-Étienne, France. He received his Ph.D. in Mechanical Engineering from National Taiwan University, Taiwan in 2010. His current research interests include statistical data mining, demand planning and forecasting, object-oriented modeling and implementation, and, in particular, the applications to semiconductor engineering data analysis and advanced process control. As the corresponding author, his email address is lan@emse.fr.

AGNÈS ROUSSY received her B.S. degree in physics from University Montpellier II, France in 1997, the M.S. degree in semiconductor physics from University Montpellier II, France in 1999, and then a Ph.D. degree in Optics Optoelectronics and Microwave from the Institute National Polytechnics Grenoble, France in 2002. She is currently Associate Professor with HDR of Department of Manufacturing Sciences and Logistics at École Nationale Supérieure des Mines de Saint-Étienne, France. Her email address is roussy@emse.fr.

MARCO S. REIS received the Ph.D. degree from the University of Coimbra, in 2006. He is a Professor in the Department of Chemical Engineering, University of Coimbra, Coimbra, Portugal. He is currently the President of the European Network for Business and Industrial Statistics and responsible for the Process Systems Engineering Research Group in the Department of Chemical Engineering, University of Coimbra. His current research interests include process systems engineering (system identification, monitoring, and control), chemometrics, industrial statistics, multivariate image analysis, systems biology, and process improvement through initiatives such as Six Sigma and lean manufacturing. His email address is marco@eq.uc.pt.

JACQUES PINATON joined STMicroelectronics Rousset in 1984. After five years with the Process Engineering Group, he joined the Device Department to implement SPC and process control methodology and tools. He participated in the start-up of three fab generations. He is currently the Manager of the Process Control System Group with STMicroelectronics Rousset, Rousset, France. He is an Engineer in metallurgy with the Conservatoire National des Arts et métiers, Aix-en-Provence, France. He is also leading various research and development programs on manufacturing science in ST Rousset, including programs on automation, APC, and diagnostics. His email address is jacques.pinaton@st.com.