International Journal of Production Research, 2014
Vol. 52, No. 15, 4469–4487, http://dx.doi.org/10.1080/00207543.2013.865854

Taylor & Francis
Taylor & Francis Group

A run-to-run controller for product surface quality improvement

Lulu Bao, Kaibo Wang* and Tianying Wu

Department of Industrial Engineering, Tsinghua University, Beijing, China

(*Received 25 June 2013; accepted 5 November 2013*)

In semiconductor manufacturing, the surface quality of silicon wafers has a significant impact on the subsequent processes that produce devices using the wafers as a component. The surface quality of a wafer is characterised by a two-dimensional (2-D) data structure: the geometric requirement for the wafer surface is smooth and flat and the thickness should fall within certain specification limits. Therefore, both low deviation and high uniformity are desirable for control over the wafer quality. In this work, we develop a run-to-run control algorithm for improving wafer quality. Considering the unique 2-D data structure, we first construct a model that encompasses the spatial correlation of the observations on the wafer surface to link the wafer quality with the process variables, and subsequently develop a recursive algorithm to generate optimal set points for the controllable factors. More specifically, a Gaussian-Kriging model is used to characterise the spatial dependence of the thickness measures of the wafer and a recursive least square method is employed to update the estimates of the model parameters. The performance of the new controller is studied via simulation and compared with existing controllers, which demonstrates that the newly proposed controller can effectively reduce the surface variations of the silicon wafers.

Keywords: Kriging; process modelling; run-to-run control; silicon wafer

1. Introduction

In recent years, semiconductors have become increasingly important in the high technology industry, and the wafer, which is a primary semiconductor element, is evolving as an indispensable ingredient in diverse industries. At the same time, because the quality of the wafers has a great impact on the subsequent products, the requirements for wafer quality are becoming more demanding. Among other factors, control of manufacturing equipment with optimised protocols is one major approach to stabilising process output, improving quality and yield and reducing cost.

In industry, the quality requirement for a wafer is twofold. In this work, we use wafer thickness as an example. First, the average thickness should meet a specific target value. Second, the variation in the thickness across the entire surface of the wafer should be sufficiently small to produce low defect rates in the chips fabricated on the surface during downstream processes. In other words, the quality standard in this context applies to both the absolute thickness value and the uniformity of the thickness.

In the wafer preparation process, lapping is one critical step for improving the geometric quality (Lin and Wang 2012; Wang and Lin 2013). The thickness heat map of a sample wafer after the lapping process is shown in Figure 1. It is evident from the sample that the thickness variation shows certain patterns: the thicknesses at different points on the wafer vary, and the changes are smooth from one side to the other. In the semiconductor industry, although thousands of measurement points are recorded by advanced metrology equipment, practitioners use only a few aggregated indicators for quality evaluation. For example, the centre thickness, which is easily obtained at the centre location of the wafer, is usually reported to reflect the thickness of a wafer, and the total thickness variation (TTV), which is defined as the difference between the largest and smallest thickness value of the wafer, is calculated from all measurement points to reflect the overall flatness of the wafer.

In semiconductor manufacturing, wafers are processed in batches. Due to the changes in the raw material of the batches, sudden component or environmental shifts and slow drifts due to equipment age effects are frequently observed in the process. In current industrial practice, advanced process control (APC) technologies gain extensive attention of many semiconductor manufacturers, like AMD, Intel, Motorola and TI, and have been applied to their production lines (Qin et al. 2006). Statistical process control (SPC) is a sort of APC technologies, in which control charts have been implemented to monitor the process based on the aggregated quality metrics, and equipment maintenance follows

*Corresponding author. Email: kbwang@tsinghua.edu.cn

© 2013 Taylor & Francis

abnormal quality changes. Because SPC cannot generate prescriptions for the control actions, another technique, run-to-run (R2R) control arises to remedy the limitation by combining SPC and feedback control (Sachs, Hu, and Ingolfsson 1995).

R2R control is a popular technique in semiconductor manufacturing for adjusting batch-based processes and has been proven effective in removing or reducing the influence of drift or shift disturbances on product quality (Del Castillo and Hurwitz 1997). For the wafer manufacturing process, most traditional R2R controllers were developed for single or multiple quality characteristics, a situation obviously different from the rich information available in Figure 1. Some researchers adopted a single quality metric to develop the R2R controller, like the centre thickness of the wafer and the removal rate in the chemical mechanical polishing process (Chen and Guo 2001). Some researchers observed that controlling only the response at the centre of the wafer is not sufficient to maintain the uniformity of the wafer (Butler and Stefani 1994). Therefore, the response surface modelling technique was used to relate the thickness at other sites with the centre thickness, and the difference of the site thicknesses was subsequently used as a constraint in generating the optimal control actions. To address the problem on the uniformity, certain works tried to find a better quality metric to incorporate the uniformity into the process output. For example, Lin and Spanos (1990) and Guo and Sachs (1993) measured the uniformity with the standard deviation divided by the mean and used it as the response in modelling of the uniformity of the wafer thickness in a poly-silicon LPCVD deposition process. Some remaining authors used both the thickness and uniformity as the objective to solve the problem. For instance, Boning et al. (1996) used the average value and the standard deviation of the removal rate calculated via the thickness at the measurement points as the control output to meet the target removal rate and achieve uniformity. However, the simulation study in this work also showed that no clear difference exists for the uniformity between control and no control. Seeing from the results of these controllers, their performance with respect to the uniformity was not good (only acceptable) although the drift in the removal rate can be compensated using R2R control. Hughes-Oliver et al. (1998) pointed out the crucial reason for the unsatisfactory performance of the controllers is that the uniformity is complex and cannot be measured simply by an aggregated metric without loss of information on the variation pattern across the wafer. It is clear that only limited information is gathered from the aggregated quality indicators. Therefore, the metrics used for monitoring and control are insensitive and they may delay correction of the quality deterioration process. Another drawback of these controllers is that it is difficult to build and interpret the model for aggregated metrics (Hughes-Oliver et al. 1998). Thus, a better R2R controller that encompassing the rich spatial information on the wafer is in need.

In this work, we develop a novel R2R controller and use the rich spatial information shown in Figure 1 to adjust the process. We first build a Gaussian-Kriging model, which uses the raw thickness information and the spatial correlation of the observations on the wafer to link the quality output with the process variables. Next, a protocol generation algorithm is developed based on the Gaussian-Kriging model to generate the optimal set points for the controllable factors.

The remainder of this article is organised as follows. The second section reviews the existing works on R2R control. In the third section, the control objective is first defined by considering the twofold requirement for surface quality control, and a Gaussian-Kriging model is subsequently developed by considering the spatial correlations among the measurement sites. In the fourth section, the optimal control action for the process is acquired, and a modified recursive least squares estimation procedure is also presented for online updating of the model parameters. In the fifth section, the

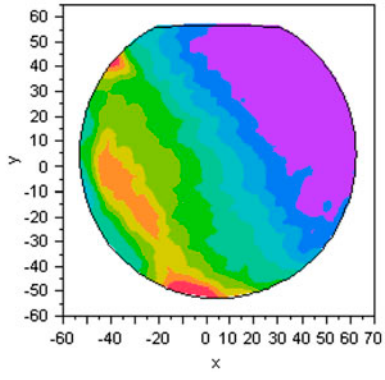


Figure 1. Wafer thickness heat map.

performance of the new controller is studied and compared with that of the existing controllers via numerical simulation. Section 6 concludes the paper with suggestions for future research.

2. Literature review of R2R control

The R2R control methodology has been widely used in semiconductor manufacturing for generating optimal recipes and stabilising process outputs (Wang and Tsung 2007). Based on the observed deviations from the target value in the historical batches, a new recipe is generated to minimise the expected deviation in the new run.

Certain commonly used controllers, such as the EWMA controller (Ingolfsson and Sachs 1993) and the double EWMA (dEWMA) controller (Butler and Stefani 1994), are designed for cases with one or multiple inputs and one output variable. In the wafer fabrication process, it is possible to apply these controllers if the aggregated indicator, like the centre thickness, or an aggregated metric of uniformity is used as a single output variable. However, as mentioned, due to the various variation patterns on the wafer surface, a single aggregate quality metric is not sufficiently representative of both the thickness and flatness. Even though multiple aggregated metrics are used as the outputs of the controller, the performance is not favourable due to the information loss by using summary indicators. Therefore, a process control approach that uses the more complete 2-D spatial data is more desirable.

Therefore, it is naturally to put forward the idea that using the raw thickness data for model building instead of aggregated indicators to avoid loss of information. If we treat each measurement point on the wafer as an individual variable, the process could be viewed as a system with multiple inputs and multiple outputs (MIMO). In the literature, such controllers as the multivariate dEWMA controller (Del Castillo and Rajagopal 2002), fuzzy controller (Fan et al. 2007), the ridge-solution controller and the right-inverse controller (Rajagopal and Del Castillo 2003) have been developed for such processes. These controllers define the control objective as the simultaneous minimisation of the deviations of all output variables from their respective targets. However, later on, we will show that as the number of measurement points becomes much greater than the number of controllable variables such MIMO controllers cannot be applied directly.

Apart from the aforementioned EWMA-type controllers, R2R controllers also have been developed following different rules. For instance, Del Castillo and Yeh (1998) and Zhe et al. (1996) designed an optimising adaptive quality controller to address the non-linearity in the actual manufacturing process. Himmel and May (1993) reported a controller based on a neural network and Hankinson et al. (1997) designed an internal model-based R2R controller. Adivikolanu and Zafiriou (2000) proposed a knowledge-based interactive R2R controller.

In the industrial practice, many controllers have been implemented and employed in the manufacturing lines, which has brought in great benefit. Qin and Badgwell (2003) reviewed and summarised a series of model predictive controllers (MPC) that are available for commercial use, including the linear quadratic Gaussian controller, model predictive heuristic controller, dynamic matrix controller and so on. They also surveyed the MPC technology products, based on the data provided by vendors like Adersa, DMS and Honeywell. Moyne, Chaudhry and Telfeyan (1995) designed and implemented a multi-branch R2R controller based on fuzzy logic and database learning mechanisms in the plasma etching process. Boning et al. (1996) applied the EWMA controller to the CMP process of the wafers. A EWMA controller was applied to the metal sputter deposition process at Texas Instruments and improved the process capability by 10% (Smith et al. 1998). AMD (Toprac 1999) brought in a APC system to control the poly-gate critical dimension, which helped reduce the photolithography rework rates from 12 to 2%. Edgar, Campbell and Bode (1999) adopted a MPC approach in the R2R control of CMP and lithography and received a dramatic increase in the process throughput at AMD. Bode, Ko, and Edgar (2004) implemented a linear MPC to control the overlay in the lithography process of AMD and successfully reduced the process variance. However, these R2R controllers cannot be applied to our problem directly because the problem of using aggregated indicators and neglecting the rich spatial information still exists.

Therefore, in our work, the raw thickness data are used for model building instead of the aggregated indicators to avoid loss of information. Considering the structure of the rich data available from the metrology equipment, we develop a Gaussian-Kriging model to fit the spatial measurements and employ a recursive algorithm to update the model parameters online to compensate for process shifts or drifts. The controller is also applicable to other processes with similar data structures and requirements for quality variables.

3. Process modelling with 2-D surface response

In this work, we target the development of a new controller for a process with 2-D output quality. Due to the unique data structure available for R2R control, in the following, we first introduce the objective function used to control a 2-D-type surface quality and subsequently develop a process model that links the output with the controllable factors.

3.1 *The response and objective function of the controller*

As discussed above, the existing R2R controllers cannot adequately address the wafer control problem for the following reasons. First, the focus of this work is control of the quality of the wafer thickness, which is evaluated using a spatial 2-D data structure rather than a single or multiple variables. The special structure of the process output makes the problem different and challenging. Second, the requirement for wafer thickness is twofold: the average thickness should meet a target value and the uniformity of the thickness values, which reflects the flatness of the wafer, should be small.

Hughes-Oliver et al. (1998) noted that it is not practical to optimise the uniformity first and subsequently adjust the process to the target because the uniformity and thickness may be affected by the same variable. Therefore, the authors used a combined measure of the uniformity and target thickness to realise the simultaneous optimisation of the two goals.

Similarly, in this work, we define

$$\mathrm{MSE} = E\left[\sum_{k=1}^{S}(y_k-\tau)^2\right] = E\left[\sum_{k=1}^{S}((y_k-\bar{y})^2+(\bar{y}-\tau)^2)\right] \tag{1}$$

where τ is the target value, y_k is the thickness value at the kth measurement point, S is the total number of measurement points on the wafer and $\bar{y}$ is the average thickness of all the measurement points. The objective function is composed of two components. The first component of the objective function attempts to minimise the differences among all of the points, which corresponds to the within-unit variation of the wafer product, whereas the second component adjusts the overall average to the target, which corresponds to the unit-to-unit variation in the production environment.

This objective function is different from that used in traditional R2R controllers in the sense that all of the measurement points on the product surface are considered and not just a single measurement point or a summary metric. This comprehensive consideration naturally encompasses both the mean deviation and the flatness, and is therefore suitable for controlling the quality of a product surface. By minimising the objective function, the two goals related to the mean and flatness will be reached simultaneously.

To minimise the objective function defined in Equation (1), we must find a model to characterise the wafer thickness and link the output with the controllable process factors.

Different surface modelling models are available in the literature. The commonly used surface modelling methods include the polynomial response surface (Guo and Sachs 1993; Taam 1998), spline (Lee, Wolberg, and Shin 1997) and wavelet (Valette and Prost 2004). These methods commonly ignore the spatial correlation in the measurement points, but the spatial correlation is an important feature for the thickness measurements of the wafers in this study.

The 2-D data structure we addressed in the wafer manufacturing process is similar to the spatial data in geostatistics. In geo-statistical modelling, the spatial dependence is taken into account and thus better performance is achieved with respect to inference, prediction and estimation for surface variability (Haran 2011). Modelling techniques such as Kriging (Cressie 1990), Markov random field (Hansen, Nair, and Friedman 1997) and Gaussian Markov random field (Rue and Tjelmeland 2002; Rue and Held 2005) are often used to model the 2-D spatial data. Among others, Kriging is the most widely used technique. The estimation in the Kriging modelling uses the best linear unbiased estimator (BLUE) if the assumption holds that the expectation of the response variable is zero and the covariance is known. In addition, the Kriging technique contains the minimal estimation variance (Myers 1994) and has also been proven to perform better than the traditional response surface model (Simpson et al. 1998) and cubic spline methods (Voltz and Webster 1990). Therefore, Kriging is chosen in this work to model the wafer thickness and cooperate with the R2R controller to improve the output quality.

3.2 *The Gaussian-Kriging model for wafer thickness*

To adapt the general Gaussian-Kriging model for characterising the wafer thickness and R2R process control, we use $y(t, k)$ to denote the thickness value at the k-th measurement point measured on a wafer produced in batch t. Based on the engineering understanding of the process, we specify the model as follows:

$$y(t,k) = \alpha + \boldsymbol{\gamma}'\mathbf{x}_k + \boldsymbol{\beta}'_{\mathbf{x}_k}\mathbf{u}_t + Z(\mathbf{x}_k) + \varepsilon_{t,k}. \tag{2}$$

The model in Equation (2) consists of two components: a general linear model and a stationary Gaussian process (Fang, Li, and Sudjianto 2010). In the general linear model, $\alpha + \boldsymbol{\gamma}'\mathbf{x}_k + \boldsymbol{\beta}'_{\mathbf{x}_k}\mathbf{u}_t + \varepsilon_{t,k}$, α is a constant, $\mathbf{x}_k = (x_{k1}, x_{k2})'$ is a (2×1) location vector denoting the coordinate position of the k-th point, $\boldsymbol{\gamma}$ is the coefficient vector of $\mathbf{x}_k$ with (2×1) order, the term $\boldsymbol{\gamma}'\mathbf{x}_k$ suggests that the site thickness is related to the site's coordinate, $\mathbf{u_t}$ is a d-dimensional vector denoting the set-points of the controllable factors at time t and $\boldsymbol{\beta}_{\mathbf{x}_k}$ is a d-dimensional coefficient vector of $\mathbf{u}_t$ that illustrates the effect

of the control actions on the process output. Because the responses at different sites may differ from each other even if they undergo the same control action, we further define

$$\boldsymbol{\beta}_{\mathbf{x}_k} = \boldsymbol{\beta}_0 + \mathbf{B}\mathbf{x}_k.$$

Thus, the output thickness is characterised by:

$$y(t,k) = \alpha + \boldsymbol{\gamma}'\mathbf{x}_k + \boldsymbol{\beta}_0'\mathbf{u}_t + \mathbf{x}_k'\mathbf{B}'\mathbf{u}_t + Z(\mathbf{x}_k) + \varepsilon_{t,k} \tag{3}$$

where $\boldsymbol{\beta}_0$ is a vector of length d, $\mathbf{B}$ is a matrix of order $(d \times 2)$ that defines the interaction effect between $\mathbf{x}_k$ and $\mathbf{u}_t$ and $\varepsilon_{t,k}$ is the white noise that denotes any measurement error that the thickness values may contain. Other random errors that cause independent variation in the sites are also accounted for by this term. We assume that $\varepsilon_{t,k}$ follows a normal distribution $N(0, \sigma_\varepsilon^2)$ and the random errors at different sites of different batches are all independent.

The stationary Gaussian process in the model $Z(\mathbf{x}_k)$ characterises the spatial correlation among the different points on the wafer. The $(Z(\mathbf{x}_1), Z(\mathbf{x}_2), \ldots, Z(\mathbf{x}_S))$ follows a multivariate Gaussian distribution with

$$\begin{aligned} E(Z(\mathbf{x}_k)) &= 0 \\ \mathrm{Cov}(Z(\mathbf{x}_i), Z(\mathbf{x}_j)) &= \sigma_Z^2 R(\boldsymbol{\theta}, \mathbf{x}_i - \mathbf{x}_j) \end{aligned} \tag{4}$$

where $R(\cdot)$ is a preset correlation function that controls the smoothing of $Z(\mathbf{x}_k)$ and σ_Z^2 is the variation of the Gaussian process. The following widely accepted Gaussian correlation function (Koehler and Owen 1996) is used to specify the correlation between different points on the wafer:

$$R(\boldsymbol{\theta}, \mathbf{x}_i - \mathbf{x}_j) = \exp\{-\theta \cdot ||\mathbf{x}_i - \mathbf{x}_j||^2\} \tag{5}$$

where θ is a constant that affects the strength of the correlation between measurements at points i and j. This correlation matrix shows that the correlation between two sites depends on the distance between them; a longer distance leads to a smaller correlation. The idea of using the distance as a symbol of correlation is widely adopted in spatial statistics.

Based on the above Gaussian-Kriging model, for any measurement point k on the wafer, the output random variable $y(t,k)$ has a Gaussian distribution with expectation $\alpha + \boldsymbol{\gamma}'\mathbf{x}_k + \boldsymbol{\beta}_0'\mathbf{u}_t + \mathbf{x}'_k\mathbf{B}'\mathbf{u}_t$ and variance $\sigma_Z^2 + \sigma_\varepsilon^2$.

3.3 *Model justification*

Because we intend to use the model in Equation (3) to characterise the real wafer thickness, before designing the R2R controller, we must verify the model using real sample data and engineering knowledge.

In Equation (3), $Z(\mathbf{x}_k)$ models the spatial variation of the wafer thickness and is assumed as a stationary Gaussian process that satisfies the condition given in Equation (4). The expectation of $Z(\mathbf{x}_k)$ is zero because the intercept contained in the model is able to capture the non-zero portion of the thickness. The assumption on the covariance implies that the correlation between the two sites is determined by the distance between them: the smaller the distance, the higher the correlation.

To verify the assumption on the covariance matrix and the correlation structure in Equation (5), we randomly select a wafer and calculate its semi-variogram, which describes the degree of spatial dependence by

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2N_{\mathbf{h}}} \sum_{i=1}^{N_{\mathbf{h}}} [y(\mathbf{x}_i + \mathbf{h}) - y(\mathbf{x}_i)]^2 \tag{6}$$

where $N_{\mathbf{h}}$ is the number of site pairs with pairwise distance $\mathbf{h}$ and $y(\mathbf{x}_i)$ is the thickness at the i-th site with coordinate vector $\mathbf{x}_i$ (Cressie 1993). The corresponding variogram for the Gaussian correlation function in Equation (5) is a Gaussian semi-variogram model defined by

$$\gamma(\mathbf{h}) = \begin{cases} 0, & \mathbf{h} = \mathbf{0}, \\ c_0 + c_e\{1 - \exp(-\theta \cdot ||\mathbf{h}||^2)\}, & \mathbf{h} \neq \mathbf{0}. \end{cases} \tag{7}$$

The estimated semi-variogram in Equation (6) can be compared with the theoretical semi-variogram in Equation (7) to check the validity of the Gaussian correlation structure.

We sample a total of 86 measurement points uniformly distributed on the wafer and draw the sample semi-variogram in Figure 2, which shows that the sample semi-variogram (dots) fits well with the Gaussian semi-variogram model (solid line).

Equation (3) also attempts to link the thickness values with the process parameters. In practice, the thicknesses of all sites are affected by the controllable factors of the process, i.e. the pressure, the velocity and the processing time, which

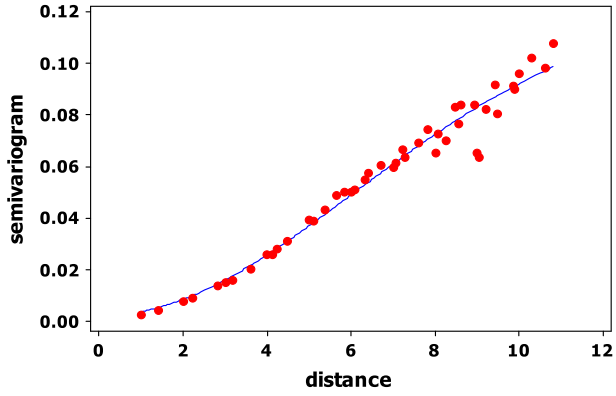


Figure 2. The sample and fitted semi-variogram.

are denoted by $\mathbf{u}_t$ in the model. The optimal settings for the controllable factors should be generated by the R2R controller and applied by the operators in practice.

One important observation from Figure 1 is that the thickness varies with the location of the corresponding sites. In addition, the wafer also shows a clear variation pattern: the site thicknesses on a wafer usually increase from one side to the other. These variation patterns can be explained by the manufacturing process used to produce the wafers. The wafer fabrication process begins with ingot pulling and slicing. A silicon ingot produced from the crystal pulling process is sliced into pieces using a wire saw. With the aid of a slurry, the cutting wires move into the ingot from one side and move out of the ingot from the other side, thus cutting the ingot into pieces of wafer (Wang and Han 2013). Due to the long processing time and uneven distribution of the cutting slurry, it is easy to generate wafers that are thin on one side and thick on the other side. The lapping operation after the slicing process cannot completely remove such unevenness. During the lapping operation, different sites on the wafer will experience different grinding paths. With the additional unevenness of the lapping pad and lapping slurry, a systematic thickness variation pattern may be introduced. Therefore, the location information $\mathbf{x}_k$ is considered a necessary variable in the model.

In addition, the process parameters and the location have certain interaction effects. In other words, the responses of different sites to the changes in process parameters are different. This phenomenon can be attributed to the uneven distribution of the lapping slurry in the lapping process. When the rotation speed and pressure increase, it becomes more difficult for the slurry to distribute evenly, thus affecting the variation pattern and thickness value at different locations. Therefore, the interaction between the location and the control action $\mathbf{x}'_k\mathbf{B}'\mathbf{u}_t$ is considered in the model.

4. The optimal surface control action

After developing the model of the output thickness and the control actions, in this section, we derive the optimal settings that minimise the loss function in Equation (1), which is also the objective of the R2R controller.

Knowing that

$$E(\varepsilon_{t,k}) = 0, \quad V(\varepsilon_{t,k}) = \sigma_\varepsilon^2$$
$$E[Z(\mathbf{x}_k)] = 0, \quad V(Z(\mathbf{x}_k)) = \sigma_Z^2 R(0) = \sigma_Z^2,$$

and denoting $c_1 = \alpha + \boldsymbol{\gamma}'\mathbf{x}_k + \boldsymbol{\beta}_0'\mathbf{u}_t + \mathbf{x}'_k\mathbf{B}'\mathbf{u}_t - \tau$, the loss function can be simplified as:

$$\begin{aligned} MSE &= E\left[\sum_{k=1}^{S}(c_1 + Z(\mathbf{x}_k) + \varepsilon_{t,k})^2\right] \\ &= \sum_{k=1}^{S}[c_1^2 + E(Z(\mathbf{x}_k)^2) + E(\varepsilon_{t,k}^2)] \\ &= \sum_{k=1}^{S}(c_1^2 + \sigma_Z^2 + \sigma_\varepsilon^2). \end{aligned}$$

Let

$$c_2 = \alpha + \boldsymbol{\gamma}'\mathbf{x}_k - \tau.$$

Because σ_Z^2 and σ_ε^2 are constants and are not affected by the control action, we can safely redefine the objective function by removing the terms that are not affected by $\mathbf{u}_t$ as follows

$$\begin{aligned}\mathrm{MSE}' &= \sum_{k=1}^{S} c_1^2 \\ &= \sum_{k=1}^{S} (c_2^2 + 2c_2\boldsymbol{\beta}_0'\mathbf{u}_t + 2c_2\mathbf{x}'_k\mathbf{B}'\mathbf{u}_t + 2\mathbf{u}_t'\boldsymbol{\beta}_0\mathbf{x}'_k\mathbf{B}'\mathbf{u}_t + \mathbf{u}_t'\boldsymbol{\beta}_0\boldsymbol{\beta}_0'\mathbf{u}_t + \mathbf{u}_t'\mathbf{B}\mathbf{x}_k\mathbf{x}'_k\mathbf{B}'\mathbf{u}_t).\end{aligned}$$

By differentiating the objective function with respect to $\mathbf{u}_t$ and forcing it to zero, we obtain

$$\left[\sum_{k=1}^{S} (\boldsymbol{\beta}_0 + \mathbf{B}\mathbf{x}_k)(\boldsymbol{\beta}_0 + \mathbf{B}\mathbf{x}_k)'\right]\mathbf{u}_t = -\sum_{k=1}^{S} (c_2\boldsymbol{\beta}_0 + c_2\mathbf{B}\mathbf{x}_k). \tag{8}$$

When the number of controllable factors is large, there exist multiple choices of $\mathbf{u}_t$ that can satisfy Equation (8). In practice, a minimum control effort is usually preferred. Therefore, following the common treatment in the literature (Tseng, Chou, and Lee 2002), we treat Equation (8) as a constraint and minimise the magnitude of changes brought on by the control action

$$\min(\mathbf{u}_t - \mathbf{u}_{t-1})'(\mathbf{u}_t - \mathbf{u}_{t-1}).$$

Solving the above function, the optimal control action for the next run can be obtained as

$$\mathbf{u}_t = \left[\mathbf{I} - \mathbf{A}'(\mathbf{A}\mathbf{A}')^{-1}\mathbf{A}\right]\mathbf{u}_{t-1} + \mathbf{A}'(\mathbf{A}\mathbf{A}')^{-1}\left[-\sum_{k=1}^{S} (c_2\boldsymbol{\beta}_0 + c_2\mathbf{B}\mathbf{x}_k)\right]$$

where

$$\mathbf{A} = \sum_{k=1}^{S} (\boldsymbol{\beta}_0 + \mathbf{B}\mathbf{x}_k)(\boldsymbol{\beta}_0 + \mathbf{B}\mathbf{x}_k)' \tag{9}$$

In practice, it is possible that $\mathbf{A}$ is not full rank under certain special circumstances, and thus $\mathbf{A}\mathbf{A}'$ may be singular. To avoid illegal inverse operations in calculating $\mathbf{u}_t$, we modify the matrix by adding a small diagonal matrix, which is a common technique used to handle similar issues. Therefore, the final optimal control action is given by

$$\mathbf{u}_t = [\mathbf{I} - \mathbf{A}'(\mathbf{A}\mathbf{A}' + \mu\mathbf{I})^{-1}\mathbf{A}]\mathbf{u}_{t-1} + \mathbf{A}'(\mathbf{A}\mathbf{A}' + \mu\mathbf{I})^{-1}\left[-\sum_{k=1}^{S} (c_2\boldsymbol{\beta}_0 + c_2\mathbf{B}\mathbf{x}_k)\right].$$

4.1 *Online update of model parameters*

In the previous section, we obtained the optimal settings for the process parameters. Due to process drifts and initial bias, the parameters in the process model may change over time. Therefore, in R2R control, the model parameters must be updated continuously in a batch-by-batch manner so that the model can better capture the real-time process dynamics.

The most widely used updating method is the EWMA algorithm. However, because the data used in this process are of high dimension and a unique spatial structure, the EWMA update equation cannot be applied directly. The usual method for estimating parameters for the simple Kriging model is the least squares method. In the R2R control scenario, observations are collected batch-by-batch, and it is inefficient or even unrealistic to maintain all historical data from a process and use all of this information for model fitting. Therefore, in this work, we use a modified online least squares method known as the recursive lease square (RLS) to update the model parameters.

In using RLS in model estimation, the parameters are updated each time a batch finishes and a new observation becomes available. In this way, the historical information is naturally stored in the parameters, and extensive calculations using all historical observations are thus avoided.

For simplicity of notation, we define a vector

$$\mathbf{h}(t,k) = [1, x_{k1}, x_{k2}, x_{k1}u_{t1}, \ldots, x_{k1}u_{td}, x_{k2}u_{t1}, \ldots, x_{k2}u_{td}, u_{t1}, u_{t2}, \ldots, u_{td}]'$$

that contains all of the linear components in Equation (3). Let $\boldsymbol{\beta}(t)$ be the corresponding coefficients with $n = 3d + 3$ elements. Equation (3) can then be expressed as

$$y(t,k) = \boldsymbol{\beta}'(t)\mathbf{h}(t,k) + Z(\mathbf{x}_k) + \varepsilon_{t,k}$$

Because the model contains $Z(\mathbf{x}_k)$ for illustrating the spatial correlation of the measurement points, the traditional RLS algorithm must modified.

Assume the parameters are estimated accurately and the prediction error of the model for point k in batch t would be given by

$$e(t,k) = y(t,k) - \hat{y}(t,k) = y(t,k) - \boldsymbol{\beta}'(t)\mathbf{h}(t,k) = Z(\mathbf{x}_k) + \varepsilon_{t,k}$$

Let $\mathbf{e}_t = [e(t,1), e(t,2), \ldots, e(t,S)]^T$ represent the prediction error vector of the t-th batch. The points on the same wafer are correlated such that the covariance matrix of the tth wafer is

$$\boldsymbol{\Sigma}_t = \sigma_Z^2 \mathbf{R}_{\boldsymbol{\theta}} + \sigma_\varepsilon^2 \mathbf{I}$$

with $\mathbf{R}_{\theta(i,j)} = R(\boldsymbol{\theta}, \mathbf{x}_i - \mathbf{x}_j)$, which is defined in Equation (5).

Because the measurement points from different batches are not correlated, for any two batches i and j,

$$\mathrm{Cov}(\mathbf{e}_i, \mathbf{e}_j) = \mathbf{0}_{S\times S}$$

After collecting all information from t batches, the covariance of the vector of the residuals $\mathbf{e} = [\mathbf{e}_1', \mathbf{e}_2', \ldots, \mathbf{e}_t']'$ is given by a block-diagonal matrix

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_1 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_2 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\Sigma}_{t-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\Sigma}_t \end{bmatrix}$$

Referring to the RLS method with exponential forgetting, we introduce a time factor into the loss function of the generalised least square method to differentiate the time value of the residuals. The parameter estimation subsequently aims to minimise the following prediction error

$$V(t) = \sum_{i=1}^{t} \lambda(t,i)\mathbf{e}_i'\boldsymbol{\Sigma}_i^{-1}\mathbf{e}_i \tag{10}$$

where $\lambda(t,i)$ is the time factor that denotes the importance ratio of the data at time i when the manufacturing process has run up to time t and $0 < \lambda(t,i) \leqslant 1$. In this work, we define the time factor as

$$\lambda(t,i) = \lambda^{t-i}$$

which decreases exponentially when the time difference between i and t increases. By substituting the simplified time factor, Equation (10) can be written as

$$V(t) = \sum_{i=1}^{t} \lambda^{t-i}\mathbf{e}_i'\boldsymbol{\Sigma}_i^{-1}\mathbf{e}_i. \tag{11}$$

Let $\underline{\mathbf{e}}_i = \sqrt{\lambda^{t-i}}\mathbf{e}_i$, and Equation (11) can be equivalently expressed as

$$V(t) = \sum_{i=1}^{t} \underline{\mathbf{e}}_i'\boldsymbol{\Sigma}_i^{-1}\mathbf{e}_i.$$

Based on the above definition, the estimates of the model parameters at time t is given by

$$\boldsymbol{\beta}(t) = (\mathbf{H}\boldsymbol{\Sigma}^{-1}\mathbf{H}')^{-1}(\mathbf{H}\boldsymbol{\Sigma}^{-1}\mathbf{y})$$

where

$$\mathbf{H} = [\mathbf{H}(1), \mathbf{H}(2), \ldots, \mathbf{H}(t)],$$

$$\mathbf{H}(i) = [\underline{\mathbf{h}}(i,1), \underline{\mathbf{h}}(i,2), \ldots, \underline{\mathbf{h}}(i,S)], \quad i = 1, \ldots, t,$$

$$\underline{\mathbf{h}}(i,k) = \sqrt{\lambda^{t-i}}\mathbf{h}(i,k), \quad k = 1,\ldots,S,$$

$$\mathbf{y} = [\mathbf{y}(1)', \mathbf{y}(2)', \ldots, \mathbf{y}(t)']',$$

$$\mathbf{y}(i) = [\underline{y}(i,1), \underline{y}(i,2), \ldots, \underline{y}(i,S)]',$$

$$\underline{y}(i,x) = \sqrt{\lambda^{t-i}}y(i,x).$$

Next, we can design the iterative algorithm using

$\mathbf{H}(t)$ to assure online and timely updating of the control model. First, we decompose the inverse covariance matrix as

$$\mathbf{\Sigma}_t^{-1} = \mathbf{G}_t'\mathbf{G}_t$$

The matrix $\mathbf{G}_t$ can be obtained via Cholesky decomposition. We further define

$$\bar{\mathbf{y}}(t) = \mathbf{G}_t\mathbf{y}(t)$$

$$\bar{\mathbf{H}}(t) = \mathbf{H}(t)\mathbf{G}_t'$$

Finally, we update the parameters using the recursive least square estimation with consideration of an exponential forgetting factor (Astrom and Wittenmark 1995) and a correction structure. The update equations are given as follows:

$$\mathbf{K}_t = \mathbf{P}_{t-1}\bar{\mathbf{H}}(t)[\lambda\mathbf{I} + \bar{\mathbf{H}}(t)'\mathbf{P}_{t-1}\bar{\mathbf{H}}(t)]^{-1}$$

$$\hat{\boldsymbol{\beta}}_t = \hat{\boldsymbol{\beta}}_{t-1} + \mathbf{K}_t(\bar{\mathbf{y}}_t - \bar{\mathbf{H}}(t)'\hat{\beta}_{t-1})$$

$$\mathbf{P}_t = [\mathbf{I} - \mathbf{K}_t\bar{\mathbf{H}}(t)']\mathbf{P}_{t-1} + \mathbf{I}_n[tr(\mathbf{K}_t\bar{\mathbf{H}}(t)'\mathbf{P}_{t-1})/n].$$

In the iterative procedures, $\mathbf{K}_t$ is a $n \times S$ matrix indicating the weight of the prediction error of batch $t-1$, $\mathbf{P}_t$ is a $n \times n$ matrix that is in proportion to the covariance matrix of the parameters and a correction structure $\mathbf{I}[tr(\mathbf{K}_t\bar{\mathbf{H}}(t)'\mathbf{P}_{t-1})/n]$ is used to maintain the trace of $\mathbf{P}_t$ as a constant to avoid the occurrence of estimation windup, which means that the matrix $\mathbf{P}_t$ increases exponentially because $\lambda \leqslant 1$ (Del Castillo and Hurwitz 1997). When a new batch finishes and a new observation becomes available, the parameter estimate is updated from $\hat{\boldsymbol{\beta}}_{t-1}$ to $\hat{\boldsymbol{\beta}}_t$. Because the update procedure works sequentially, it makes online estimation convenient.

Because the main feature of the new controller is the use of the Kriging technique in the model's spatial correlation process output, we refer to the proposed scheme as the Kriging R2R controller. In the following section, we conduct extensive studies to compare the performance of the newly proposed controller with that of existing controllers.

4.2 *A comparison with the EWMA controller*

Because the EWMA controller is one of the most popular controllers used in industry, in this section, we compare the proposed Kriging R2R controller with the EWMA controller.

From the model framework, it can be observed that the Gaussian-Kriging model takes the location parameters into account and attempts to characterise the response variables using spatially correlated data, whereas the original EWMA controller, which is developed from a simple or multiple linear regression model, does not use the location information and can only consider single or multiple response variables, which represents an essential difference between the two models.

If the locations of all sites are ignored or if all sites are assumed to share the same location $\mathbf{x}_0$, the process model in Equation (3) would simplify to

$$\begin{aligned} y(t) &= \alpha + \boldsymbol{\gamma}'\mathbf{x}_0 + \boldsymbol{\beta}_0'\mathbf{u}_t + \mathbf{x}_0'\mathbf{B}'\mathbf{u}_t + Z(\mathbf{x}_0) + \varepsilon_t \\ &= \kappa + \boldsymbol{\rho}'\mathbf{u}_t + \delta \end{aligned}$$

where

$$\kappa = \alpha + \boldsymbol{\gamma}'\mathbf{x_0}$$
$$\boldsymbol{\rho} = \boldsymbol{\beta}_0 + \mathbf{B}\mathbf{x_0}$$
$$\delta \sim N(0, \sigma_\varepsilon^2 + \sigma_Z^2)$$

The simplified process model is analogous to the one used for developing the EWMA controller. In such case, the optimal control action is obtained by solving the following equation

$$(\boldsymbol{\beta}_0 + \mathbf{B}\mathbf{x_0})(\boldsymbol{\beta}_0 + \mathbf{B}\mathbf{x_0})'\mathbf{u}_t = (\tau - \alpha - \boldsymbol{\gamma}'\mathbf{x}_0)[\boldsymbol{\beta}_0 + \mathbf{B}\mathbf{x_0}]$$

Or more simply

$$\boldsymbol{\rho}'\mathbf{u}_t = (\tau - \kappa)$$

Obviously, the control action of the new model is the same as that of the original EWMA model. However, when a spatial structure exists in the data, the new model is expected to capture the variation structure more accurately, thus leading to better control performance. This expectation will be studied further in the next section.

5. Performance study

To study and verify the effectiveness of the Kriging R2R controller, in this section, we use simulations to compare the Kriging R2R controller with other traditional controllers, including the dEWMA controller (Butler et al. 1994), the multivariate dEWMA controller (Rajagopal and Del Castillo 2003) and the self-tuning controller (Del Castillo and Hurwitz 1997; Jen, Jiang, and Fan 2004). The four controllers are tested under four real process models to check the robustness. The setup for the simulation is introduced first, followed by a presentation of the exact forms of the competing controllers, which are introduced later. Finally, the numerical results are presented and discussed for performance comparison.

5.1 *Setup for the simulation study*

For illustration purposes, we assume that eight data points, distributed as shown in Figure 3, are sampled from the wafer surface. The positions of the points are selected as:

$$\{(1,2),(2,5),(-1,-2),(-2,-5),(1,-2),(2,-5),(-1,2),(-2,5)\}$$

Note that the Kriging R2R controller could be used for wafers with more sampling points. We here choose eight data points for demonstration purposes to speed up the simulation.

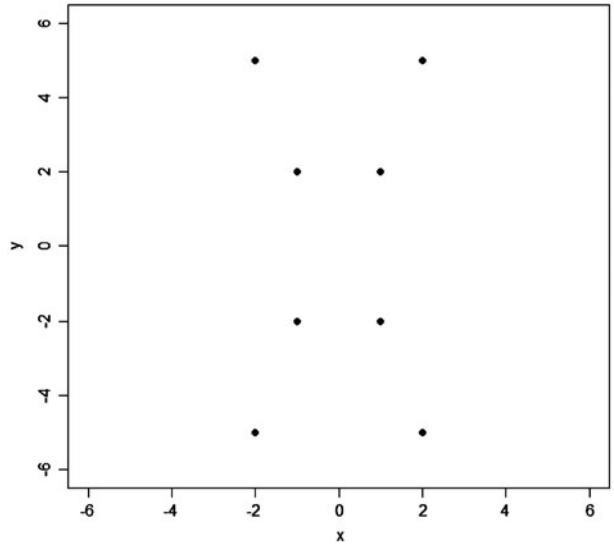


Figure 3. The selected points on the wafer.

We study the performance of the controller when being applied to simulated processes governed by the following four models:

The first process model is the one on which the proposed Kriging R2R controller is derived:

$$y(t,k) = \alpha + \boldsymbol{\gamma}'\mathbf{x}_k + \boldsymbol{\beta}_0'\mathbf{u}_t + \mathbf{x}_k'\mathbf{B}'\mathbf{u}_t + Z(\mathbf{x}_k) + \varepsilon_{t,k} \quad \text{(Process Model 1)}.$$

The second process model fits the usual assumption of the EWMA controller:

$$y(t,k) = \alpha + \boldsymbol{\beta}_0'\mathbf{u}_t + \varepsilon_{t,k} \quad \text{(Process Model 2)}.$$

The next process model fits that assumed in the self-tuning controllers:

$$y(t,k) = \alpha + \boldsymbol{\gamma}'\mathbf{x}_k + \boldsymbol{\beta}_0'\mathbf{u}_t + \mathbf{x}'_k\mathbf{B}'\mathbf{u}_t + \varepsilon_{t,k} \quad \text{(Process Model 3)}.$$

The last process model is:

$$y(t,k) = \alpha + \boldsymbol{\beta}_0'\mathbf{u}_t + Z(\mathbf{x}_k) + \varepsilon_{t,k} \quad \text{(Process Model 4)}.$$

Without loss of generality, we arbitrarily set the model parameters as follows

$$\alpha = 551, \quad \boldsymbol{\gamma} = \begin{bmatrix} 0.16 \\ 0.7 \end{bmatrix}, \quad \boldsymbol{\beta}_0 = \begin{bmatrix} 4 \\ -2 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 0.1 & -0.12 \\ -0.06 & 0.11 \end{bmatrix}$$

$$\sigma_Z = 1.15, \quad \sigma_\varepsilon = 0.87, \quad \theta = 0.15$$

The target value for the thickness of the wafers is set to $\tau = 550$. The process contains two controllable factors that can be adjusted to change the process output, meaning that the dimension of $\mathbf{u}_t$ is $d = 2$.

5.2 *Traditional controllers for comparison*

5.2.1 *The dEWMA controller*

The first contrast controller is a dEWMA controller with multiple inputs and a single output. The single response variable in this controller is the average thickness of eight measurement points on a wafer. The purpose is to compare this controller with the Kriging R2R controller and investigate the importance of using the raw thickness data instead of the summary indicators. Because the dEWMA controller cannot incorporate spatial information into the process model, the output thickness is assumed to follow

$$y(t) = \alpha + \boldsymbol{\beta}'\mathbf{u}_{t-1} + N_t, \quad N_t = \delta t + \varepsilon_t$$

According to the right inverse controller for the multiple-input/single-output situation (Nair, Taam, and Ye 2002), we describe the updating method for the controller as

$$a_t = \lambda_1(y(t) - \mathbf{b}'\mathbf{u}_{t-1}) + (1-\lambda_1)a_{t-1}$$

$$D_t = \lambda_2(y(t) - \mathbf{b}'\mathbf{u}_{t-1} - a_{t-1}) + (1-\lambda_2)D_{t-1}$$

where a_t is the estimate of α_t, $\mathbf{b}$ is the estimate of $\boldsymbol{\beta}$, and D_t is the estimate of δ_t. The value of $\mathbf{b}$ is estimated using the random thickness data generated in the initialisation of the model and is subsequently set as a constant. The updating parameters λ_1 and λ_2 are set to 0.1 in the simulation. The control action is given by

$$\mathbf{u}_t = \frac{T - a_t - D_t}{\mathbf{b}'\mathbf{b}}\mathbf{b} + \left(I - \frac{\mathbf{b}\mathbf{b}'}{\mathbf{b}'\mathbf{b}}\right)\mathbf{u}_{t-1}$$

5.2.2 *The multivariate dEWMA controller*

The second controller for comparison is a multivariate dEWMA (multi-dEWMA) controller with MIMO. The multivariate dEWMA controller in the MIMO situation was proposed by Rajagopal and Del Castillo (2003) and updates the intercept and the drift at the same time. The raw thickness data are used in this controller. The thickness vector of eight measurement points of the t-th batch is the response variable vector $\mathbf{y}(t)$, and there are two process parameters in the control variable vector $\mathbf{u_t}$.

The multi-dEWMA controller assumes the process output follows

$$\mathbf{y}(t) = \boldsymbol{\alpha} + \boldsymbol{\beta} u_{t-1} + \mathbf{N}_t, \quad \mathbf{N}_t = \boldsymbol{\delta} t + \boldsymbol{\varepsilon}_t$$

where

$$\mathbf{y}(t) = (y(t,1), y(t,2), \ldots, y(t,8))', \quad \mathbf{u}_{t-1} = (u_{t1}, u_{t2})'$$

are the output and input vectors, respectively. In addition,

$$\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_8)', \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \\ \vdots & \vdots \\ \beta_{81} & \beta_{82} \end{pmatrix}$$

are unknown constants. The term $\mathbf{N}_t = \boldsymbol{\delta} t + \boldsymbol{\varepsilon}_t$ denotes a noise model with a multivariate deterministic trend, where $\boldsymbol{\delta}$ is a (8×8) matrix in which element (i, i) is the average drift rate of the corresponding output variable. The parameters in the controller are updated by

$$\mathbf{A}_t = \boldsymbol{\Lambda}_1 (\mathbf{y}_t - \mathbf{B}\mathbf{u}_{t-1}) + (\mathbf{I} - \boldsymbol{\Lambda}_1)\mathbf{A}_{t-1}$$

$$\mathbf{D}_t = \boldsymbol{\Lambda}_2 (\mathbf{y}_t - \mathbf{B}\mathbf{u}_{t-1} - \mathbf{A}_{t-1}) + (\mathbf{I} - \boldsymbol{\Lambda}_2)\mathbf{D}_{t-1}$$

where $\mathbf{A}_t$ is the estimate of the intercept after batch t, $\mathbf{D}_t$ is the estimate of the total drift after batch t, $\mathbf{B}$ is the estimate of matrix $\boldsymbol{\beta}$ and is estimated through using the random thickness data generated in the initialisation of the model, and $\boldsymbol{\Lambda}_1$ and $\boldsymbol{\Lambda}_2$ are weight matrices with (8×8) order. In this model, the weight matrix is set as

$$\boldsymbol{\Lambda}_1 = 0.1 \times \mathbf{I}, \boldsymbol{\Lambda}_2 = 0.1 \times \mathbf{I}$$

The control action of the controller is guided by

$$\mathbf{u}_t = (\mathbf{I} - \mathbf{B}'(\mathbf{B}\mathbf{B}' + \mu\mathbf{I})^{-1}\mathbf{B})\mathbf{u}_{t-1} + \mathbf{B}'(\mathbf{B}\mathbf{B}' + \mu\mathbf{I})^{-1}(\boldsymbol{\tau} - \mathbf{A}_t - \mathbf{D}_t)$$

where $\boldsymbol{\tau}$ is the target vector with all elements equal to 550 and $\mu\mathbf{I}$ is a correction structure to avoid singularity.

It should be noted that when implementing the multi-dEWMA controller, the number of responses is larger than the number of process parameters, a situation that has been rarely studied in the past literature.

5.2.3 *The self-tuning controller*

A recursive estimation algorithm is the key to self-tuning controllers (Ramsay 1982) and is also the method modified for parameter estimation in our work. In this work, we adopt a self-tuning controller with a simple linear regression model.

$$y(t,k) = \alpha + \gamma' \mathbf{x}_k + \boldsymbol{\beta}_0' \mathbf{u}_\mathbf{t} + \mathbf{x}'_k \mathbf{B}' \mathbf{u}_t + \varepsilon_{t,k}$$

Compared with the proposed control model, the difference between the two controllers is that the self-tuning controller does not apply the Gaussian-Kriging technique for modelling the spatial correlation. This difference is explicitly presented in the covariance matrices for the two controllers. The covariance matrix for the proposed controller is $\boldsymbol{\Sigma}_t = \sigma_Z^2 \mathbf{R}_\boldsymbol{\theta} + \sigma_\varepsilon^2 \mathbf{I}$, whereas that of the self-tuning controller is $\boldsymbol{\Sigma}_t = \sigma_\varepsilon^2 \mathbf{I}$. The objective function, the control action and the updating method are nearly the same as those of our controller.

5.3 *Model initialisation*

In industrial applications, we should first control the production process based on experience to obtain the selected historical data necessary to set the parameters of the control models. In this simulation, we assume that production data from 100 wafers are available to initialise each control model.

For the two types of EWMA model, we use the least squares method to fit the control model. For the self-tuning controller, we use the generalised least squares approach to fit the model. For the Kriging R2R controller, we first assume that the estimation of variances σ_z^2 and σ_ε^2 are accurate because of the complexity of the initialisation method.

5.4 *Simulation results*

Using the four controllers, we run the simulation for 100 repetitions with 1000 steps in each simulation with each step representing a wafer produced by the process. For all controllers, the following performance metrics are calculated from the simulation results:

(1) SSE

where SSE is the summation of squared deviations from the target thickness for a wafer.

$$\sum_{k=1}^{S} (y(t,k) - \tau)^2$$

(2) TTV

where TTV is the difference between the maximum and the minimum thickness.

$$\max_{1 \leqslant k \leqslant S} (y(t,k)) - \min_{1 \leqslant k \leqslant S} (y(t,k))$$

(3) SD

where SD is the standard deviation of the thickness of $S = 8$ points on a wafer.

$$\sqrt{\frac{\sum_{k=1}^{S} (y(t,k) - \bar{y})^2}{S-1}}$$

(4) SSD

where SSD is the summation of all of the pairwise squared differences of the points on a wafer.

$$\sum_{1 \leq i,j \leq S} (y(t,i) - y(t,j))^2$$

(5) MD

where MD is the difference between the mean thickness and the target thickness of a wafer.

$$(\bar{y} - \tau)^2$$

Note that in this simulated process,

$$\sum_{k=1}^{S} (\bar{y} - \tau)^2 = \sum_{k=1}^{S} (y(t,k) - \tau)^2 - \sum_{k=1}^{S} (y(t,k) - \bar{y})^2 = SSE - 7SD^2$$

(6) #OS points

#OS points is shortened from the average number of points out of specifications per 1000 wafers. It is used to measure the defective level of the wafer surface. The upper specification limit is set as 555, while the lower specification limit as 545.

(7) #OS wafers

Table 1. Performance of the controllers under process Model 1.

	dEWMA	Multi-dEWMA	Self-tuning	Kriging R2R
Average SSE	79.8790	6.2100E+13	19.7837	19.4541
Average TTV	9.0888	2.5500E+05	4.2535	4.2509
Average SD	3.2600	96,600.0000	1.4400	1.4400
Average SSD	612.1600	4.8800E+14	125.5900	125.4700
Average MD	0.4200	1.4500E+11	0.5100	0.4700
#OS points	772.38	93.49	15.41	11.89
#OS wafers	569.63	25.77	13.09	11.33

#OS wafers is the abbreviation of the average number of wafers out of specifications per 1000 wafers directly linked to the product yield. Same specifications are set as those of the #OS points. A wafer is regarded defective as long as some point on the wafer exceeds the specifications.

The first metric, SSE, is similar to our objective function and is used to measure the performance of the controller in achieving the two goals of target thickness and uniformity. The following three metrics are used to measure the uniformity of the surface. The MD metric is used to measure the ability of the controller to achieve the target. The last two metrics is to measure the defective rate level of the wafers.

The overall performance indicators are calculated as the average of all wafers in the 100 simulated runs, as shown in Tables 1–4. In the following, we first investigate the results under Model 1 in details, then briefly study the results under other models.

Table 1 shows the performance of the four competing controllers under Model 1. It is found that:

(a) Among the four controllers, the multi-dEWMA controller performs the worst and is incapable of controlling this process. Further investigation reveals that this controller has a larger number of output variables than the number of input variables. In the model for the multi-dEWMA controller, eight points are treated as eight output variables, whereas the process contains only two controllable factors. This situation leads to an unstable design of the controller. Therefore, unless we further modify the controller to restrict the output space or use other types of regression methods, controlling a surface is not equivalent to controlling a process with multivariate outputs; thus the multi-dEWMA controller cannot be applied directly.

Table 2. Performance of the controllers under process Model 2.

	dEWMA	Multi-dEWMA	Self-tuning	Kriging R2R
Average SSE	6.1520	6.1544	6.3628	6.1828
Average TTV	2.4769	2.4769	2.4769	2.4769
Average SD	0.8395	0.8395	0.8395	0.8395
Average SSD	42.3838	42.3835	42.3835	42.3838
Average MD	0.1068	0.1071	0.1331	0.1106
#OS points	0	0	0.16	0
#OS wafers	0	0	0.11	0

Table 3. Performance of the controllers under process Model 3.

	dEWMA	Multi-dEWMA	Self-tuning	Kriging R2R
Average SSE	69.0861	7.9240	8.4389	8.1390
Average TTV	8.4270	2.8595	2.8603	2.8567
Average SD	3.1052	0.9705	0.9706	0.9696
Average SSD	545.8560	56.4883	56.4772	56.3782
Average MD	0.1068	0.1079	0.1724	0.1365
#OS points	295.47	0	2.67	0.53
#OS wafers	267.93	0	1.54	0.48

Table 4. performance of the controllers under process Model 4.

	dEWMA	Multi-dEWMA	Self-tuning	Kriging R2R
Average SSE	17.0102	17.0311	17.2206	16.9216
Average TTV	3.9620	3.9620	3.9620	3.9620
Average SD	1.3452	1.3452	1.3452	1.3452
Average SSD	109.2122	109.2122	109.2124	109.2115
Average MD	0.4199	0.4225	0.4462	0.4088
#OS points	4.86	4.87	8.82	5.27
#OS wafers	4.79	4.79	6.46	5.03

(b) The dEWMA controller gives the lowest average MD value among all controllers but performs rather poorly in all other aspects. This result is not surprising since the dEWMA controller uses the average thickness as its output, and the control actions are generated to minimise the mean deviation only. Because the controller ignores the dispersion of the eight points, other metrics that measure the uniformity of the outputs are not satisfactory. This situation also demonstrates the necessity of incorporating the spatial information of the wafer into the process modelling and objective function design.

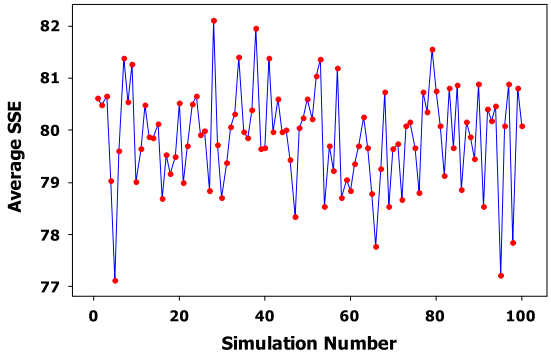
(c) The self-tuning controller performs slightly worse than the Kriging R2R controller in terms of average SSE, SSD and MD and equally well in terms of TTV and SD. Although both controllers use all eight measurement points, the self-tuning controller ignores the spatial correlations among the points, whereas the Kriging R2R controller considers the correlations. Therefore, the Kriging R2R controller produces more accurate parameter estimates and thus performs better.

Overall, the Kriging R2R controller shows the best overall performance because it yields the lowest SSE, the best uniformity and a satisfactory mean deviation.

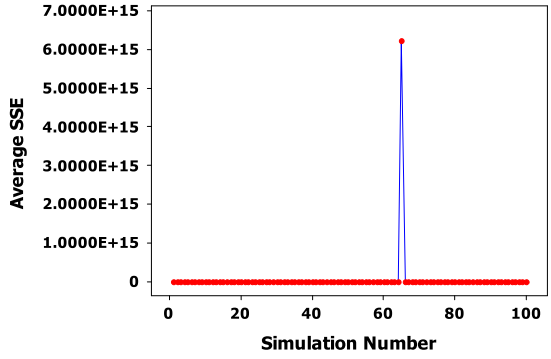
Models 2–4 simulate the cases in which the process output is not affected by either spatial correlations or position information. Therefore, we apply the four competing controller to these models to study the robustness of the Kriging controller.

Under Model 2, as Table 2 shows, all controllers perform quite closely. The dEWMA controller, which is originally developed for Model 2, is slighter better in terms of average SSE and MD, while the proposed Kriging R2R controller is also comparative to others, although Model 2 considers neither the location information nor the spatial correlation.

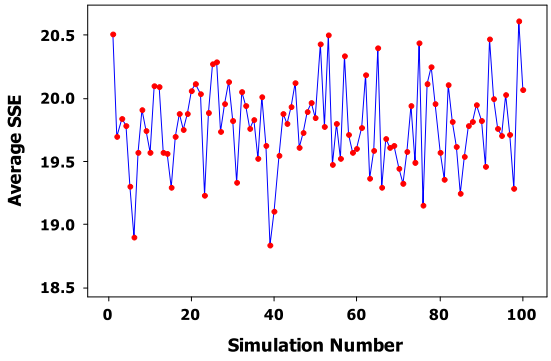
Under Model 3, the multi-dEWMA controller seems to have the best performance among the four controllers. However, further studies show that if the variance σ_ε in the model becomes larger, the multi-dEWMA control becomes unstable (with an extremely large SSE). This happens since the multi-dEWMA controller has a larger number of output variables than the number of input variables. Among the other three controllers, the dEWMA controller performs poorly as it fails to consider the location information in the process model, the Kriging R2R controller has the best overall performance, followed by the self-tuning controller.



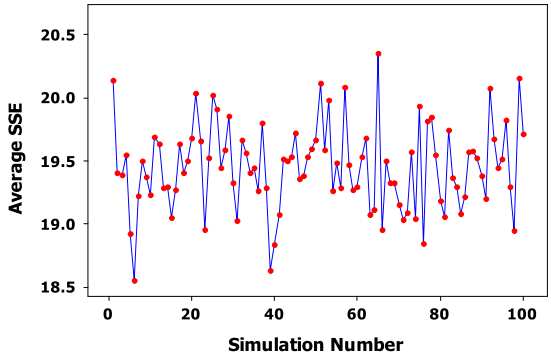
(a) Average SSE of the dEWMA controller



(b) Average SSE of the multi-dEWMA controller

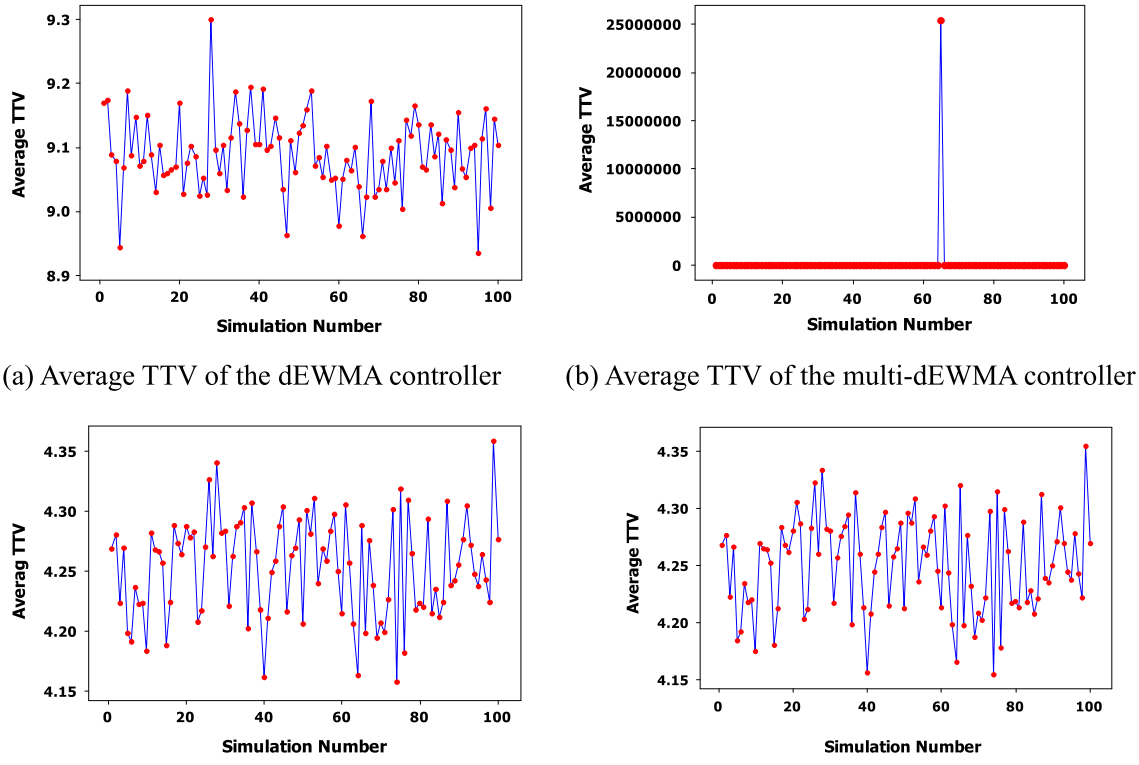


(c) Average SSE of the self-tuning controller



(d) Average SSE of the Kriging R2R controller

Figure 4. Average SSE values of different controllers in process Model 1.



(a) Average TTV of the dEWMA controller

(b) Average TTV of the multi-dEWMA controller

(c) Average TTV of the self-tuning controller

(d) Average TTV of the Kriging R2R controller

Figure 5. Average TTV values of different controllers in process Model 1.

Under Model 4, the performances of the four controllers stay quite close. A careful investigation reveals that proposed Kriging R2R controller is slightly better than the dEWMA controller and the multi-dEWMA controller in terms of average SSE, SSD and MD, but is slightly worse in terms of #OS points and #OS wafers. The self-tuning controller is the worst among all.

To assess the stability of the controller performance, we also calculate the average SSE and average TTV of 1000 wafers in each simulation and display these values in Figures 4 and 5, respectively (using Model 1 as an example). The figures suggest that under Model 1, except for the multi-dEWMA controller, the performances of the other three controllers are stable across the 100 replicates. As explained previously, the poor stability of the multi-dEWMA controller is caused by the large number output variables.

Through the above analysis, we conclude that:

(a) When the location information is explicitly included in the real process model (Models 1 and 3), the Kriging R2R controller and the self-tuning controller are more favourable than the dEWMA or the multi-dEWMA controller. The multi-dEWMA controller is not stable, especially when the variance of the model is large. When the location information is not considered by the process model (Models 2 and 4), the performance of the Kriging R2R controller is still acceptable, and is close to the best one, which shows that the Kriging R2R controller is robust to model misspecification.

(b) When the spatial correlation is taken into account by the process model (Models 1 and 4), the proposed Kriging R2R controller usually shows superior performance. On the contrary, even if the spatial correlation is not included in the actual process model (Models 2 and 3), the performance of the Kriging R2R controller is still close to the best one.

(c) The performances of the self-tuning controller and the proposed Kriging R2R controller are similar in most cases. However, the Kriging R2R controller is superior to self-tuning controller under all the four process models.

Therefore, we conclude that the proposed Kriging R2R controller is more favourable if the process model is correctly specified, and is also robust even if the model's location information or spatial correlation is missing from the process (the true model).

6. Conclusion

In this work, we proposed a new Kriging R2R controller intended to improve wafer quality in the wafer fabrication process. The aim of the controller design is to simultaneously meet the thickness target and improve the uniformity of the wafers.

Compared with conventional controllers, the features of the newly proposed Kriging R2R controller are summarised as follows: (a) The Kriging R2R controller makes use of the raw site thickness information and controls the thickness of multiple measurement points on the wafer rather than using a single summary statistic. This formulation avoids the loss of useful information and is important to control the quality of a surface. (b) The new process model considers the spatial correlation among the different measurement points, and thus can better capture the dynamics of the surface. (c) A modified recursive least squares method is used to estimate the model parameters. The modified algorithm operates in an online manner and avoids the need to save a massive amount of historical observations for parameter estimation.

The performance of the Kriging R2R controller is compared with other controllers via extensive simulation studies. When the real process output is guided by a process model with location information and spatial correlation, it is found that the proposed controller is capable of driving the overall thickness on target while maintaining high uniformity. When the real process output is not affected by location information or has no spatial correlation, the performance of the Kriging controller is stable and close to the one with the best performance. That is, although the Kriging R2R controller is developed to capture spatial correlation of a 2-D surface-type product, it is still robust even if the product shows weak or even no spatial correlations.

In this work, the major challenge stems from the 2-D quality measures. Compared with the traditional univariate or multivariate quality indices, the 2-D surface represents an important data structure for quality control in advanced manufacturing processes. Therefore, other quality control initiatives, i.e. SPC and experimental design targeted to such 2-D surface data, are interesting topics for future research.

Acknowledgement

We greatly thank the editor and the anonymous referees for their valuable comments, which have helped improve this work greatly.

Funding

This work was supported by the National Natural Science Foundation of China [grant number 71072012] and Tsinghua University Initiative Scientific Research Program.

References

Adivikolanu, S., and E. Zafiriou. 2000. “Extensions and Performance/Robustness Tradeoffs of the EWMA Run-to-run Controller by Using the Internal Model Control Structure.” *IEEE Transactions on Electronics Packaging Manufacturing* 23 (1): 56–68.

Astrom, K. J., and B. Wittenmark. 1995. *Adaptive Control*. 2nd ed. Boston, MA: Addison-Wesley.

Bode, C., B. Ko, and T. Edgar. 2004. “Run-to-run Control and Performance Monitoring of Overlay in Semiconductor Manufacturing.” *Control Engineering Practice* 12 (7): 893–900.

Boning, D. S., W. P. Moyne, T. H. Smith, J. Moyne, R. Telfeyan, A. Hurwitz, S. Shellman, and J. Tayor. 1996. “Run by Run Control of Chemical–mechanical Polishing.” *IEEE Transactions on Components, Packaging, and Manufacturing Technology, Part C* 19 (4): 307–314.

Butler, S. W., and J. A. Stefani. 1994. “Supervisory Run-to-run Control of Polysilicon Gate Etch Using in-situ Ellipsometry.” *IEEE Transactions on Semiconductor Manufacturing* 7 (2): 193–201.

Butler, S. W., J. Stefani, M. Sullivan, S. Maung, G. Barna, and S. Henck. 1994. “Intelligent Model-based Control-system Employing in-situ Ellipsometry.” *Journal of Vacuum Science & Technology a-Vacuum Surfaces and Films* 12 (4): 1984–1991.

Chen, A., and R.-S. Guo. 2001. “Age-based Double EWMA Controller and its Application to CMP Processes.” *Semiconductor Manufacturing, IEEE Transactions on* 14 (1): 11–19.

Cressie, N. 1990. “The Origins of Kriging.” *Mathematical Geology* 22 (3): 239–252.

Cressie, N. A. 1993. *Statistics for Spatial Data*, 69–83. Rev ed. New York: Wiley.

Del Castillo, E., and A. M. Hurwitz. 1997. “Run-to-run Process Control: Literature Review and Extensions.” *Journal of Quality Technology* 29 (2): 184–196.

Del Castillo, E., and R. Rajagopal. 2002. “A Multivariate Double EWMA Process Adjustment Scheme for Drifting Processes.” *IIE Transactions* 34 (12): 1055–1068.

Del Castillo, E., and J. Y. Yeh. 1998. “An Adaptive Run-to-run Optimizing Controller for Linear and Nonlinear Semiconductor Processes.” *IEEE Transactions on Semiconductor Manufacturing* 11 (2): 285–295.

Edgar, T. F., W. Campbell, and C. Bode "Model-based Control in Microelectronics Manufacturing." *Decision and Control, 1999. Proceedings of the 38th IEEE Conference on*, 4185–4191.

Fan, S.-K., C. Fan, P. Kung, and C.-Y. Wang. 2007. "Development of Run-to-run (R2R) Controller for the Multiple-input Multiple-output (MIMO) System Using Fuzzy Control Theories." *International Journal of Production Research* 45 (14): 3215–3243.

Fang, K.-T., R. Li, and A. Sudjianto. 2010. *Design and Modeling for Computer Experiments*, 145–153. Boca Raton, FL: Chapman and Hall/CRC.

Guo, R.-S., and E. Sachs. 1993. "Modeling, Optimization and Control of Spatial Uniformity in Manufacturing Processes." *IEEE Transactions on Semiconductor Manufacturing* 6 (1): 41–57.

Hankinson, M., T. Vincent, K. B. Irani, and P. P. Khargonekar. 1997. "Integrated Real-time and Run-to-run Control of Etch Depth in Reactive Ion Etching." *IEEE Transactions on Semiconductor Manufacturing* 10 (1): 121–130.

Hansen, M. H., V. N. Nair, and D. J. Friedman. 1997. "Monitoring Wafer Map Data from Integrated Circuit Fabrication Processes for Spatially Clustered Defects." *Technometrics* 39 (3): 241–253.

Haran, M. 2011. "Gaussian Random Field Models for Spatial Data." In *Handbook of Markov Chain Monte Carlo*, edited by S. Brooks, A. Gelman, G. Jones, and X.-L. Meng, 449–478. Boca Raton, FL: Chapman and Hall/CRC.

Himmel, C. D., and G. S. May. 1993. "Advantages of Plasma Etch Modeling Using Neural Networks over Statistical Techniques." *IEEE Transactions on Semiconductor Manufacturing* 6 (2): 103–111.

Hughes-Oliver, J. M., J.-C. Lu, J. C. Davis, and R. S. Gyurcsik. 1998. "Achieving Uniformity in a Semiconductor Fabrication Process Using Spatial Modeling." *Journal of the American Statistical Association* 93 (441): 36–45.

Ingolfsson, A., and E. Sachs. 1993. "Stability and Sensitivity of an EWMA Controller." *Journal of Quality Technology* 25 (4): 271–287.

Jen, C. H., B. C. Jiang, and S. K. S. Fan. 2004. "General Run-to-run (R2R) Control Framework Using Self-tuning Control for Multiple-input Multiple-output (MIMO) Processes." *International Journal of Production Research* 42 (20): 4249–4270.

Koehler, J., and A. Owen. 1996. "Computer Experiments." *Handbook of Statistics* 13 (13): 261–308.

Lee, S., G. Wolberg, and S. Y. Shin. 1997. "Scattered Data Interpolation with Multilevel B-splines." *IEEE Transactions on Visualization and Computer Graphics* 3 (3): 228–244.

Lin, K.-K., and C. J. Spanos. 1990. "Statistical Equipment Modeling for VLSI Manufacturing: An Application for LPCVD." *IEEE Transactions on Semiconductor Manufacturing* 3 (4): 216–229.

Lin, J., and K. Wang. 2012. "A Bayesian Framework for Online Parameter Estimation and Process Adjustment Using Categorical Observations." *IIE Transactions* 44 (4): 291–300.

Moyne, J. R., N. Chaudhry, and R. Telfeyan. 1995. "Adaptive Extensions to a Multibranch Run-to-run Controller for Plasma Etching." *Journal of Vacuum Science & Technology A* 13 (3): 1787–1791.

Myers, D. E. 1994. "Spatial Interpolation: An Overview." *Geoderma* 62 (1): 17–28.

Nair, V. N., W. Taam, and K. Q. Ye. 2002. "Analysis of Functional Responses from Robust Design Studies." *Journal of Quality Technology* 34 (4): 355–370.

Qin, S. J., and T. A. Badgwell. 2003. "A Survey of Industrial Model Predictive Control Technology." *Control Engineering Practice* 11 (7): 733–764.

Qin, S. J., G. Cherry, R. Good, J. Wang, and C. A. Harrison. 2006. "Semiconductor Manufacturing Process Control and Monitoring: a Fab-Wide Framework." *Journal of Process Control* 16 (3): 179–191.

Rajagopal, R., and E. Del Castillo. 2003. "An Analysis and MIMO Extension of a Double EWMA Run-to-run Controller for Non-squared Systems." *International Journal of Reliability, Quality and Safety Engineering* 10 (04): 417–428.

Ramsay, J. 1982. "When the Data are Functions." *Psychometrika* 47 (4): 379–396.

Rue, H., and L. Held. 2005. *Gaussian Markov Random Fields: Theory and Applications*. Boca Raton, FL: Chapman and Hall/CRC.

Rue, H., and H. Tjelmeland. 2002. "Fitting Gaussian Markov Random Fields to Gaussian Fields." *Scandinavian Journal of Statistics* 29 (1): 31–49.

Sachs, E., A. Hu, and A. Ingolfsson. 1995. "Run by Run Process Control: Combining SPC and Feedback Control." *Semiconductor Manufacturing, IEEE Transactions on* 8 (1): 26–43.

Simpson, T. W., T. M. Mauery, J. J. Korte, and F. Mistree. 1998. "Comparison of Response Surface and Kriging Models for Multidisciplinary Design Optimization." 7th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization, 381–391, St. Louis, MO, AIAA-98, September 2–4.

Smith, T. H., D. S. Boning, J. Stefani, and S. W. Butler. 1998. "Run by Run Advanced Process Control of Metal Sputter Deposition." *Semiconductor Manufacturing, IEEE Transactions on* 11 (2): 276–284.

Taam, W. 1998. "A Case Study on Process Monitoring for Surface Features." *Quality and Reliability Engineering International* 14 (4): 219–226.

Toprac, A. J. 1999. "AMD's Advanced Process Control of Poly-gate Critical Dimension." *Proceedings of SPIE* 3882: 62–65.

Tseng, S.-T., R.-J. Chou, and S.-P. Lee. 2002. "A Study on a Multivariate EWMA Controller." *IIE Transactions* 34 (6): 541–549.

Valette, S., and P. Prost. 2004. "Wavelet-based Multiresolution Analysis of Irregular Surface Meshes." *IEEE Transactions on Visualization and Computer Graphics* 10 (2): 113–122.

Voltz, M., and R. Webster. 1990. "A Comparison of Kriging, Cubic Splines and Classification for Predicting Soil Properties from Sample Information." *Journal of Soil Science* 41 (3): 473–490.

Wang, K., and K. Han. 2013. "A Batch-based Run-to-run Process Control Scheme for Semiconductor Manufacturing." *IIE Transactions* 45: 658–669.

Wang, K., and J. Lin. 2013. "A Run-to-run Control Algorithm Based on Timely and Delayed Mixed-resolution Information." *International Journal of Production Research* 51 (15): 4704–4717.

Wang, K., and F. Tsung. 2007. "Run-to-run Process Adjustment Using Categorical Observations." *Journal of Quality Technology* 39 (4): 312–325.

Zhe, N., J. R. Moyne, T. Smith, D. Boning, E. Del Castillo, Y. Jinn-Yi, and A. Hurwitz. 1996. "A Comparative Analysis of Run-to-run Control Algorithms in the Semiconductor Manufacturing Industry." *Advanced Semiconductor Manufacturing Conference and Workshop, 1996. ASMC 96 Proceedings. IEEE/SEMI* 1996: 375–381.