IIE Transactions (2012) **44**, 291–300
Copyright © "IIE"
ISSN: 0740-817X print / 1545-8830 online
DOI: 10.1080/0740817X.2011.568039

A Bayesian framework for online parameter estimation and process adjustment using categorical observations

JING LIN[1] and KAIBO WANG[2,*]

[1] *Department of E-commerical Product Marketing, Baidu Corp., Beijing 100085, China*
E-mail: lin_jing@baidu.com
[2] *Department of Industrial Engineering, Tsinghua University, Beijing 100084, China*
E-mail: kbwang@tsinghua.edu.cn

Received May 2010 and accepted February 2011

In certain manufacturing processes, accurate numerical readings are difficult to collect due to time or resource constraints. Alternatively, low-resolution categorical observations can be obtained that can act as feasible and low-cost surrogates. Under such situations, all classic statistical quality control activities, such as model building, parameter estimation, and feedback adjustment, have to be done on the basis of these categorical observations. However, most existing statistical quality control methods are developed based on numerical observations and cannot be directly applied if only categorical observations are available. In this research, a new online approach for parameter estimation and run-to-run process control using categorical observations is developed. The new approach is built in the Bayesian framework; it provides a convenient way to update parameter estimates when categorical observations arrive gradually in a real production scenario. Studies of performance reveal that the new method can provide stable estimates of unknown parameters and achieve effective control performance for maintaining quality.

Keywords: Categorical observations, parameter estimation, Bayesian method, Gibbs sampling, statistical process adjustment, statistical process control

1. Introduction

Statistical Process Adjustment (SPA) has been shown to be an efficient approach for ensuring output quality, improving production efficiency, and diminishing defects for Run-to-Run (R2R) processes in semiconductor manufacturing (Del Castillo, 2006). In a R2R process, a typical SPA algorithm utilizes process output and other environmental information to generate recipes for process adjustment. As noted by Del Castillo (2006), information used by the SPA framework can be versatile; the decision process to generate recipes can be viewed as a high-level control loop that may involve human factors and advanced statistical techniques. Extensive research on R2R process control has been reported in the literature, including the creation of the Exponentially Weighted Moving Average (EWMA) controller and its extensions (Ingolfsson and Sachs, 1993; Tseng *et al.*, 2007; Jin and Tsung, 2009), double EWMA controller (Chen and Guo, 2001), and the self-tuning controller (Del Castillo and Hurwitz, 1997). It should be noted that all of these commonly used controllers work on one condition; that is, quality readings are measured on a numerical scale.

Nevertheless, low-resolution categorical data are frequently seen in certain processes in semiconductor manufacturing nowadays when practical constraints limit the availability of numerical values. For instant, time-consuming measurement operations must be avoided in a high-yield process to guarantee production efficiency. Some high-precision equipment may be too expensive and not all processes have the capability to provide accurate measurements. Under such circumstances, conventional R2R controllers that rely on numerical values cannot be applied directly; some novel process control strategies using categorical observations have been designed to handle this new challenge (Spanos and Chen, 1997; Wang and Tsung, 2007; Shang *et al.*, 2008).

Spanos and Chen (1997) studied a dry develop process in which the quality characteristics were measured on a scale of "very rough," "rough," "smooth," and "very smooth." The authors built logistic regression models to characterize the relationship between output variables and controllable factors of the process. However, the models were built assuming that a historical offline dataset was available. Online observations were not used for model estimation or

[*] Corresponding author

update. Wang and Tsung (2007) proposed a feedback controller for categorical observations; however, the authors also assumed that an offline dataset was available for Phase I model building. Following the work of Wang and Tsung (2007), Shang *et al.* (2008) improved the controller by considering misclassifications. However, the authors did not consider the model estimation issue; the process model and all parameters were still assumed to be known in advance. Lin and Wang (2010) proposed a control scheme using an adjusted Maximum Likelihood (ML) function. However, the performance of the adjusted ML method heavily depended on the tuning parameter used by the method. An inappropriate parameter setting could lead certain estimates to infinite (Chipman and Hamada, 1996).

The successfull implementation of process adjustment strategies in an R2R process requires the developement of an appropriate model, obtaining estimates of unknown parameters, and providing feedback-control algorithms based on available information. All of these tasks, therefore, are dominated by the type, amount, and style of information that can be collected. In this research, we use the lapping process in semiconductor manufacturing as an example and propose a new Bayesian framework for statistical quality control. The new framework works on the basis of categorical observations and estimates and updates parameters using the Bayesian method. We expect that the Bayesian method can use categorical information more effectively, provide more accurate estimates of parameters, and generate better control performance, compared with the adjusted ML method in Lin and Wang (2010).

Lapping is a very important step in the wafer preparation process, which usually consists of slicing, lapping, chemical vapor deposition, and polishing. As the first mechanical treatment step on a wafer surface after slicing, lapping is critical to forming high-level quality characteristics for downstream fabrication. Among others, thickness is one geometric quality attribute that needs to be carefully controlled. However, accurate thickness values can only be obtained in a special inspection room using an expensive testing machine, which is both time-consuming and costly. In practice, wafers are only moved to the inspection room for quality inspection after the polishing operation has been completed. Therefore, no timely accurate data can be provided for implementing R2R control of the lapping process. Alternatively, right after the lapping operation, a less expensive but inaccurate machine is used to help operators group lapped wafers into five different categories, labeled as "very thin," "thin," "normal," "thick," and "very thick." Therefore, low-resolution categorical observations are available for R2R control of the lapping process.

The thickness of the lapped wafers is largely dominated by the lapping time and the thickness before the lapping step. The latter is the incoming wafer thickness generated by the slicing stage, which cannot be changed. Therefore, lapping time is the only controllable factor during production. To achieve an ideal thickness, the lapping time usually needs to be adjusted between runs, based on the other numerical inputs and categorical output. However, with only categorical observations available as the output of each run, the traditional EWMA controller is not applicable. Hence, a new controller that can generate control actions using categorical data is needed for this process.

This research intends to develop a quality control scheme using categorical observations following the Bayesian framework. The application of Bayesian methods for parameter estimation using categorical information has already been discussed in the existing literature. Chipman and Hamada (1996) proposed a Bayesian approach to estimate parameters in a Generalized Linear Model (GLM) with categorical variables using Gibbs sampling; their discussion used the assumption that the categorical observations were uncorrelated. Girard and Parent (2001) adjusted this Bayesian GLM and extended it to cases with autocorrelated observations. Lawrence *et al.* (2008) studied parameter estimation under multivariate categorical output. It should be noted that all of these methods work in an offline manner by assuming that historical observations have already been collected before model fitting. However, in an R2R process, products are produced batch by batch. With units sampled and measured for each batch (or run), data arrive gradually. Therefore, it is necessary and important to extend the Bayesian method to handle categorical data streams that grow step by step.

Jen and Jiang (2008) proposed a system that integrated R2R control, evolutionary operation, and response surface modeling. In this system, online experiment data were utilized to update estimated parameters, renew the process model, and obtain new recipes. Vanli and Del Castillo (2009) investigated an online Bayesian robust control problem and presented two new Bayesian approaches. To incorporate growing categorical data streams for SPA, Wang and Tsung (2010) constructed a Bayesian framework for recursive parameter estimation using Gibbs sampling. However, they only studied the estimation of parameters in the linear process model, while assuming the cutoff parameters to be known. Cutoff parameters, also called cut-points, are a vector against which samples are classified into mutually exclusive categories. A model that is critical to categorical observations generation will be presented in the following section. The cut-points can be seen as hidden rules in generating output against a categorical scale; these points cannot be measured directly and therefore need to be estimated.

In this research, we aim to propose a Bayesian framework for online parameter estimation and process adjustment based on categorical observations. The rest of this article is organized as follows. Section 2 introduces a general model for R2R processes with categorical observations. Based on this model, Section 3 presents the Bayesian method for online parameter estimation and process adjustment using categorical data; Section 4 studies the performance of the Bayesian framework. Finally, Section 5 concludes this article with potential topics for future research.

2. Process modeling

To characterize an R2R process, most existing literature (see, for example, Del Castillo and Hurwitz (1997), Wang and Tsung (2007, 2008, 2010), and Shang *et al.* (2008)) has chosen a linear model due to its simplicity. In fact, within a normal operational range, many R2R processes can be approximated by a linear model. Othman *et al.* (2006) showed that a linear model can be used to illustrate a lapping process. In this research, we also studied the lapping process via experimental design and found that the lapping process is adequately represented by the following equation:

$$y_t = a + bu_{t-1} + cx_t + \varepsilon_t, \tag{1}$$

where y_t is the process output at time t; u_{t-1} is the lapping time set at time $t-1$, which is the control factor; and x_t is the thickness of the incoming wafer from the slicing stage, which is observable but cannot be changed. Also, ε_t is process disturbance, and a, b, c are all coefficients. Without loss of generality, we assume $\varepsilon_t \sim N(0, \sigma^2)$.

It is worth noting that although developed from the lapping process, the model in Equation (1) is quite general and can be used to illustrate a wide range of processes with continuous input and output variables. For example, a similar form has been used by Ruegsegger *et al.* (1999) to illustrate a reactive ion etching process; Chen and Guo (2001) used a similar model (without considering the impact of incoming dimensions) to characterize a chemical–mechanical process in wafer fabrication.

The disturbance sequence in Equation (1) is assumed to be a normally distributed white noise series. That is, the process is assumed to be stable. This assumption generally holds for the lapping process since its material removal rate is very slow and the machine can be treated as a stable process. Theoretically speaking, a process already on target under white noise needs no further process adjustment. However, initial bias often exists in real processes, especially in short-run processes. The lapping process is such a short-run process. It has to be stoped each day between shifts; such stops cause the lapping machine to cool down during the break in production, thus leading to serious bias during the following warm-up period. In addition, different customer orders use different single-crystal silicon ingots, which cause the physical and mechanical properties of silicon wafers to change from order to order. Such changes often influence machine performance (e.g., removal speed). Therefore, the parameters in Equation (1) have to be updated continuously; R2R control should be implemented in this process to compensate for deviations.

In Equation (1), the numerical variable y_t cannot be obtained directly in the situations introduced in Section 1, including the studied lapping process. Therefore, we can treat y_t as a latent variable. An observable categorical variable, denoted by Y_t, is assumed to be linked with y_t by the following mapping function (see, for example, Chipman and Hamada (1996), Girard and Parent (2001), and Wang and Tsung (2007, 2010) for similar treatment of categorical variables):

$$\begin{cases} Y_t = 1 & \text{if} \quad y_t < \gamma_1, \\ Y_t = j & \text{if} \quad \gamma_{j-1} < y_t \le \gamma_j, \ j = 2, \ldots, c-1, \\ Y_t = c & \text{if} \quad y_t > \gamma_{c-1}, \end{cases} \tag{2}$$

where $\boldsymbol{\gamma} = [\gamma_1, \ldots, \gamma_{c-1}]^{\mathrm{T}}$ is the vector of the cutoff parameters that are used to classify samples into different categories.

In a real scenario, the intercept a in the linear model is a function of the temperature of the lapping pan and machine setup when a new order arrives, whereas the other coefficients b and c are relatively stable. Therefore, in this research b and c are assumed to be known from experience. The proposed method, however, can be extended to estimate and update these parameters without much difficulty. The wafer thickness before lapping, x_t, would be measured accurately after slicing; hence, it is known in the lapping stage. The cutoff points in Equation (2) cannot be measured directly; thus, they are unknown and need to be estimated.

3. Bayesian online estimation based on categorical observations

Let $\theta = \{a, \gamma\}$ be a set of unknown parameters to be estimated. In this section, we present a Bayesian framework for estimating and updating the unknown parameters θ online and generating control actions. This online method assumes that observations arrive smoothly and continuously. Whenever a run finishes and a new observation arrives, it will be utilized to update parameters and generate an optimal recipe for a new run. The integration of new observations of parameter estimation is conceptually shown in Fig. 1. When a new run finishes and a new observation

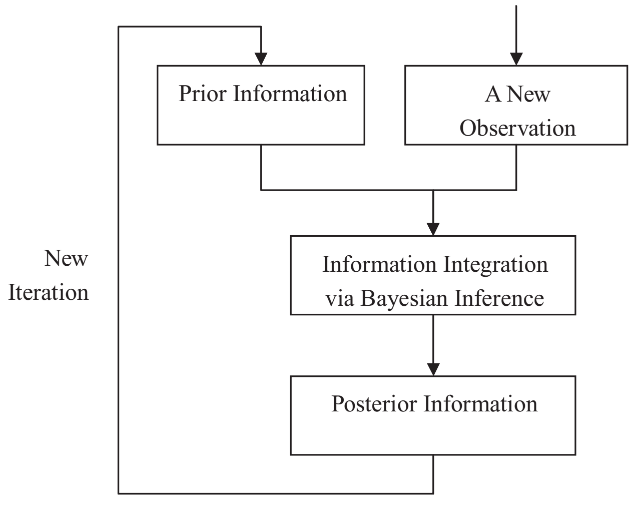

Fig. 1. A conceptual plot of the Bayesian framework.

is available, a prior estimate of the parameters is updated based on Bayesian inference equations; a posterior estimate is then obtained. This posterior serves the prior of the next run. Therefore, with new observations arriving continuously, estimates of parameters are expected to be increasingly accurate.

To construct a Bayesian framework for online parameter estimation, we need to go through two steps: first, choose an appropriate prior distributions; second, derive posterior distributions by integrating priors and samples. In Section 3.1 we introduce the prior and posterior distributions. However, due to the special type of categorical observations, it is difficult to obtain closed forms for all posterior distributions. Therefore, in Section 3.2, we derive the full conditional distribution of each parameter and then propose using Gibbs sampling to update its estimates. Gibbs sampling provides a convenient way for Bayesian estimation since it only requires full conditional distributions, which is usually much easier to obtain. In Section 3.3 we outline the whole procedure, and in Section 3.4 we present equations for R2R adjustment based on estimated parameters.

3.1. *Prior and posterior distributions for online estimation*

To use a Bayesian framework to update parameters, prior distributions for unknown parameters θ need to be chosen first. In a continuous updating scenario, it is reasonable to assume that a follows a normal distribution $N(a_0, \sigma_a^2)$, $\boldsymbol{\gamma} \sim N(\boldsymbol{\gamma}_0, \boldsymbol{\Sigma}_{\gamma,0})$, where I is the identity matrix, $\boldsymbol{\gamma}_0 = (\gamma_{1,0}, \ldots, \gamma_{c-1,0})^{\mathrm{T}}$, and $\boldsymbol{\Sigma}_{\gamma,0} = \sigma_{\gamma,0}^2 I$. When estimation evolves, the impact of such assumptions will diminish gradually.

During parameter estimation, the value of the cutoff parameter is required to satisfy that $\gamma_{\min} < \gamma_1 < \gamma_2 < \cdots < \gamma_{c-1} < \gamma_{\max}$. That is, the cut-points must form an ordered sequence and stay within a feasible range during the update. The joint posterior at step t is denoted by $f(\theta | Y_t, \ldots, Y_1)$, and when a new categorical observation Y_{t+1} is made, it is immediately utilized to update the current posterior based on the Bayes rule; that is,

$$\begin{aligned} f(\theta \,|\, Y_{t+1}\,, Y_t, \ldots, Y_1) &\propto f(Y_{t+1} \,|\, \theta) \times f(\theta \,|\, Y_t, \ldots, Y_1) \\ &= f(Y_{t+1} \,|\, a, \boldsymbol{\gamma}) \times f(\theta \,|\, Y_t, \ldots, Y_1). \end{aligned}$$

3.2. *Full conditional distribution for Gibbs sampling*

Due to the complex form of the parameter distributions, it is difficult to calculate the updated posterior $f(\theta|(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t), Y_{t+1})$ directly in this application. Therefore, we add a latent variable y_{t+1} and use Gibbs sampling to simulate all distributions. Gibbs sampling is a convenient way to derive marginal distributions in Bayesian analysis. It sequentially draws samples from each parameter's full conditional distribution (the distribution of one parameter given all other parameters); the sequence of samples can be used to estimate marginal distributions. Therefore, to estimate parameter values, it is sufficient to obtain each parameter's full conditional distribution.

To perform Gibbs sampling, the sampled chain should satisfy the ergodicity property to ensure convergence. However, in the online estimation framework, if the exact estimated posteriors are used, the ergodicity property of cutoff parameter γ_j cannot be guaranteed. In its full conditional distribution below, it can be seen that the sampling would be constrained by the estimated latent variables before $\{\hat{y}_s, s \le t\}$; while $\{\hat{y}_s, s \le t\}$ are only the estimated values, not the true ones (true values are never known), the generated sample of γ_j might be unable to reach the whole space of its true posterior distribution.

$$f\left(\gamma_j \,|a, \{\gamma_i, i \neq j\}, y_{t+1}, Y_{t+1}, Y_t, \ldots, Y_1\right) \propto \begin{cases} f\left(\gamma_j \,|\, Y_t, \ldots, Y_1\right) \times I\left(\gamma_{j-1} < \gamma_j < y_{t+1}\right), & j = Y_{t+1} - 1, \\ f\left(\gamma_j \,|\, Y_t, \ldots, Y_1\right) \times I\left(y_{t+1} < \gamma_j < \gamma_{j+1}\right), & j = Y_{t+1}, \\ f\left(\gamma_j \,|\, Y_t, \ldots, Y_1\right) \times I\left(\gamma_{j-1} < \gamma_j < \gamma_{j+1}\right), & o.w. \end{cases}$$

$$\begin{cases} N\big(\mu_{\gamma_j,0}, \sigma_{\gamma_j,0}^2\big) \times I\left(\max\limits_{s \le t} \{\hat{y}_s \,|\, Y_s = j-1\} < \gamma_j < \min\limits_{s \le t} \{\hat{y}_s \,|\, Y_s = j\}\right) \times I\left(\gamma_{j-1} < \gamma_j < y_{t+1}\right), & j = Y_{t+1} - 1, \\ N\big(\mu_{\gamma_j,0}, \sigma_{\gamma_j,0}^2\big) \times I\left(\max\limits_{s \le t} \{\hat{y}_s \,|\, Y_s = j-1\} < \gamma_j < \min\limits_{s \le t} \{\hat{y}_s \,|\, Y_s = j\}\right) \times I\left(y_{t+1} < \gamma_j < \gamma_{j+1}\right), & j = Y_{t+1}, \\ N\big(\mu_{\gamma_j,0}, \sigma_{\gamma_j,0}^2\big) \times I\left(\max\limits_{s \le t} \{\hat{y}_s \,|\, Y_s = j-1\} < \gamma_j < \min\limits_{s \le t} \{\hat{y}_s \,|\, Y_s = j\}\right) \times I\left(\gamma_{j-1} < \gamma_j < \gamma_{j+1}\right), & o.w. \end{cases}$$

One solution to this problem is that at the end of each step the estimated posterior is approximated by a parametric distribution that could only have value in the entire real axis. Solla and Winther (1999) presented an optimal online learning method using a Bayesian approach. The exact posterior distribution is approximated by a simple parametric distribution and each new observation is utilized for posterior update. Following the method proposed by Solla and Winther (1999), we now propose an updating framework for the posteriors.

Here the joint posterior $f(\theta | Y_t, \ldots, Y_1)$ is approximated by a multivariate normal distribution $N(\theta | \boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t)$, where $\boldsymbol{\mu}_t$ is the mean vector of θ, also treated as the estimates of θ at step t, and $\boldsymbol{\Sigma}_t$ is the covariance matrix. Notice that here we assume the covariance matrix to be diagonal, due to computation simplicity, which implies that a and cutoff parameters are independent; thus, the sample mean and variance of each parameter could be used as the estimates. When a new categorical observation Y_{t+1} arrives, the current posterior now becomes

$$f(\theta|(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t), Y_{t+1}) \propto f(Y_{t+1}|\theta) \times N(\theta|\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t),$$

which can be decomposed to

$$f(\theta|(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t), Y_{t+1}) \propto f(Y_{t+1}|a, \boldsymbol{\gamma}) \times N(a|\mu_{a,t}, \sigma^2_{a,t}) \times N(\boldsymbol{\gamma}|\boldsymbol{\mu}_{\gamma,t}, \boldsymbol{\Sigma}_{\gamma,t}), \qquad (3)$$

in which $\mu_{a,t}$ is the mean of a, $\boldsymbol{\mu}^{\mathrm{T}}_{\gamma,t}$ is the mean of vector of $\boldsymbol{\gamma}$, $\sigma^2_{a,t}$ is the variance of a, $\sigma^2_{\gamma_i,t}$ is the variance of γ_i, $\mu_t = (\mu_{a,t}, \mu^{\mathrm{T}}_{\gamma,t})^{\mathrm{T}}$, and $\boldsymbol{\Sigma}_t = \mathrm{diag}(\sigma^2_{a,t}, \sigma^2_{\gamma_1,t}, \ldots, \sigma^2_{\gamma_{c-1},t})$. Then the updated parametric posterior $N(\theta|\boldsymbol{\mu}_{t+1}, \boldsymbol{\Sigma}_{t+1})$ is approximated by $f(\theta|(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t), Y_{t+1})$, by choosing the parameter values $(\boldsymbol{\mu}_{t+1}, \boldsymbol{\Sigma}_{t+1})$ to be equal to the mean and covariance matrix of $f(\theta|(\boldsymbol{\mu}_t, \sigma^2_t), Y_{t+1})$. Such an approximation minimizes the Kullback-Leibler distance between the two distributions (Saad, 1998).

Then the full conditional distributions need to be calculated. In this application, the conditional distributions of three sets of parameters are to be estimated; that is, the distribution of the latent variable, $f(y_{t+1}|a, \boldsymbol{\gamma}, (\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t), Y_{t+1})$, the distribution of the intercept parameter $f(a|y_{t+1}, \boldsymbol{\gamma}, (\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t), Y_{t+1})$, and the distributions of the cutoff parameters, $f(\gamma_j|a, \{\gamma_i, i \neq j\}, y_{t+1}, (\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t), Y_{t+1}), (j = 1, \ldots, c-1)$.

At each step, once the categorical variable, Y_{t+1}, is observed, we first investigate the full conditional distribution of y_{t+1} using the following formula:

$$f(y_{t+1}|a, \boldsymbol{\gamma}, (\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t), Y_{t+1}) \propto N(a + bu_t + cx_{t+1}, \sigma^2) \times I(\gamma_{Y_{t+1}-1} < y_{t+1} < \gamma_{Y_{t+1}}), \qquad (4)$$

which is a normal distribution truncated by the boundaries of the category that it falls within.

Based on the joint posterior distribution and Bayes' theorem, the full conditional distribution of a can be written as

$$f(a|y_{t+1}, \boldsymbol{\gamma}, (\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t), Y_{t+1}) \propto N(\mu_{a,t}, \sigma^2_{a,t}) \times f(y_{t+1}, Y_{t+1}|a, \boldsymbol{\gamma}). \qquad (5)$$

given that

$$f(y_{t+1}, Y_{t+1}|a, \boldsymbol{\gamma}) = f(y_{t+1}|a, \boldsymbol{\gamma}) \times f(Y_{t+1}|y_{t+1}, a, \boldsymbol{\gamma}) \propto N\left(a + bu_t + cx_{t+1}, \sigma^2\right) \times I\left(\gamma_{Y_{t+1}-1} < y_{t+1} < \gamma_{Y_{t+1}}\right).$$

If we treat Equation (5) as a function of a and we move all constants independent of a, the density function of a can be reduced to

$$f\left(a\,|y_{t+1}, \boldsymbol{\gamma}, (\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t), Y_{t+1}\right) \propto N\left(\left(\frac{\mu_{a,t}}{\sigma^2_{a,t}} + \frac{y_{t+1} - bu_t - cx_{t+1}}{\sigma^2}\right)\Bigg/ \left(\frac{1}{\sigma^2_{a,t}} + \frac{1}{\sigma^2}\right), 1\Bigg/\left(\frac{1}{\sigma^2_{a,t}} + \frac{1}{\sigma^2}\right)\right), \qquad (6)$$

where $\mu_{a,0} = a_0$, and $\sigma^2_{a,0} = \sigma^2_a$.

The full conditional distribution of the cutoff parameters can be derived similarly, with an additional constraint condition $\gamma_{\min} < \gamma_1 < \gamma_2 < \cdots < \gamma_{c-1} < \gamma_{\max}$, where $\gamma_{\min}$ and $\gamma_{\max}$ are the lower and upper bounds of $\boldsymbol{\gamma}$. Thus, the samples of each cutoff point drawn from Gibbs sampling should also satisfy the restriction, which guarantees that the estimates of cutoff parameters, the sample means from Gibbs sampling, never overlap with each other. The full conditional distribution of γ_j used in Gibbs sampling is

$$f(\gamma_j|a, \{\gamma_i, i \neq j\}, y_{t+1}, (\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t), Y_{t+1}) \propto \begin{cases} N\left(\mu_{\gamma_j,t}, \sigma^2_{\gamma_j,t}\right) \times I(\gamma_{j-1} < \gamma_j < y_{t+1}), & j = Y_{t+1} - 1, \\ N\left(\mu_{\gamma_j,t}, \sigma^2_{\gamma_j,t}\right) \times I(y_{t+1} < \gamma_j < \gamma_{j+1}), & j = Y_{t+1}, \\ N\left(\mu_{\gamma_j,t}, \sigma^2_{\gamma_j,t}\right) \times I(\gamma_{j-1} < \gamma_j < \gamma_{j+1}), & o.w. \end{cases} \qquad (7)$$

3.3. *Online parameter estimation procedures*

Now that the full conditional distributions are ready, we can now outline the procedure to estimate and update parameters at each step when a new observation arrives.

When a new categorical observation Y_{t+1} is collected, the Gibbs sampling procedure starts to sample y_{t+1}, a, and $\gamma_1, \ldots, \gamma_{c-1}$ repeatedly for a sufficiently large number of times. The updated parameter values in the approximated normal distribution $N(\theta|\boldsymbol{\mu}_{t+1}, \boldsymbol{\Sigma}_{t+1})$ can be obtained by calculating the sample mean and variance of a and $\gamma_1, \ldots, \gamma_{c-1}$ with the initial samples removed. The sampling process for each run of online estimation can be outlined as follows:

Step 1. Sample one y_{t+1} from Equation (4).

Step 2. Using the approximated normal distribution of a, $N(\mu_{a,t}, \sigma^2_{a,t})$ estimated in the previous run as a prior and y_{t+1} sampled from Step 1, calculate the conditional distribution of a for run $t+1$ in Equation (6).

Step 3. Sample an a from its conditional distribution obtained from Step 2.

Step 4. Using the approximated normal distribution of γ_j, $N(\mu_{\gamma_j,t}, \sigma^2_{\gamma_j,t})$ estimated in the previous run as a prior, y_{t+1} sampled from Step 1 and a from Step 3, calculate the conditional distribution of γ_j for run $t+1$ in Equation (7).

Step 5. Sample $\gamma_1, \ldots, \gamma_{c-1}$ in sequence from their respective conditional distribution, with one element at a time.

Step 6. Using the newly sampled a and $\boldsymbol{\gamma}$, update the conditional distribution of y_{t+1} and go back to Step 1.

Step 7. Repeat Steps 1–6 until reaching a sufficiently large number of times and the Gibbs sampling procedure is stable.

The above steps are conducted when the sample at step $t+1$ arrives. With a large number of sequentially drawn samples, we can calculate the mean and variance of a and γ_j

and obtain the posterior distributions of these parameters, $N(\mu_{a,t+1}, \sigma^2_{a,t+1})$ and $N(\mu_{\gamma_j,t+1}, \sigma^2_{\gamma_j,t+1})$.

Compared with the estimates obtained at Step t, the new updates contain information conveyed by observation Y_{t+1}. When another observation, Y_{t+2}, becomes available, the whole procedure will be repeated again to further integrate the information conveyed by Y_{t+2}. With new observations arriving continuously, the estimates of parameters are expected to approach their respective true values.

3.4. *A Bayesian controller*

To control the process on target and compensate for initial bias, at the end of each run, a recipe should be generated under a specific criterion to minimize process variability. Denote the target of the process in Equation (1) as T. We define the following quadratic loss function conditioning on all historical information as the objective of process adjustment at step $t+1$:

$$L = E[(y_{t+2} - T)^2 \mid Y_{t+1}, \ldots, Y_1]. \tag{8}$$

Equation (1) shows that $y_{t+2} = a + bu_{t+1} + cx_{t+2} + \varepsilon_{t+2}$ and $E(\varepsilon^2_{t+2}) = \sigma^2$; thus, it follows that:

$$L = (a - cx_{t+2} - T)^2 + b^2u^2_{t+1} + 2(a - cx_{t+2} - T)bu_{t+1} + \sigma^2.$$

Taking the partial derivative of the above equation with respect to u_{t+1} as zero and replacing the unknown parameter with its estimate leads to the optimal control action:

$$u_{t+1} = \frac{T - a^{(t+1)} - cx_{t+2}}{b}, \tag{9}$$

where $a^{(t+1)}$ is the estimation of a at step $t+1$. That is, at each run, using Equation (9) to adjust a process, the quality loss function defined in Equation (8) is expected to be minimized.

As the above controller is derived using a Bayesian framework, we call it a Bayesian controller. It is interesting to see that the Bayesian controller is similar to the popular EWMA controller. Both controllers use a continuously updated intercept parameter to calculate control actions. However, the EWMA controller updates a using an EWMA equation that relies on continuous observations, whereas the Bayesian controller updates a using a Bayesian framework that relies on categorical observations.

4. Performance studies

In this section, we investigate the performance of the proposed method and compare it with existing methods.

The EWMA controller is a popular choice to control R2R processes. It is known that the implementation of an EWMA controller requires accurate numerical readings to be available. Therefore, it is unfair to compare it with the proposed method, which uses less accurate categorical observations. However, in the following, we still show the performance of the EWMA controller in controlling a simulated process and compare the difference between the EWMA and Bayesian controllers.

Wang and Tsung (2007) proposed a categorical controller that generates control actions using categorical observations in a R2R process. The authors assumed that the process model is built based on historical data; the model is not updated once established. In this work, we assume that initial bias exists and parameters are updated gradually. Although it is somewhat unfair to compare these two methods, we still show the performance of categorical controller for benchmarking.

Lin and Wang (2010) proposed an adjusted ML method for online parameter estimation and process adjustment using categorical observations. Let θ_t be an estimate obtained after the tth run and θ_{t+1} be an estimate obtained after the $(t+1)$th run. Unlike the Bayesian method, the adjusted ML method updates estimates of θ through maximizing the following objective function:

$$F(\theta_{t+1}) = \eta \times \log P_{\theta_{t+1}}(Y = j) - \frac{1}{2}(\theta_{t+1} - \theta_t)^2,$$

where $\log P_{\theta_{t+1}}(Y = j)$ is the log-likelihood function under new parameters θ_{t+1}. The distance between new and original estimates, $(\theta_{t+1} - \theta_t)^2$, is used as a penalty to balance the magnitude of parameter changes and log-likelihood of new observations. The tuning parameter, η, is the learning rate to control the updating speed. Using a first-order Taylor series for approximation and equaling the first-order derivative to zero yields:

$$\theta_{t+1} = \theta_t + \eta \times \frac{\mathrm{d}(\log P_{\theta_t}(Y = j))}{\mathrm{d}\theta}. \tag{10}$$

Therefore, all parameters can be updated in a recursive manner when new observations become available. The adjusted ML method can estimate all parameters in the model, including the cutoff parameters. Therefore, in the following, we also compare the proposed method with the adjusted ML method.

For all of the studied cases, the true model was set to be the same as the one in Lin and Wang (2010). That is, the target process follows Equation (1) with $a = 60$, $b = 2$, $c = 0.1$, and a standard deviation $\sigma = 3$. The process target T equals 400, and four cut-points, 396, 398, 401, and 405, are used to classify output y_t into five mutually exclusive categories; that is, $\boldsymbol{\gamma} = [396, 398, 401, 405]$. Here the thickness before lapping x_t is assumed to obey an uniform distribution in the interval [450, 550].

4.1. *A study of parameter estimation performance*

To simulate possible initial bias in a process, in the first study, we assumed that the prior mean of a and $\boldsymbol{\gamma}$ were 50 and $[394, 396, 403, 407]^{\mathrm{T}}$, respectively, and their standard

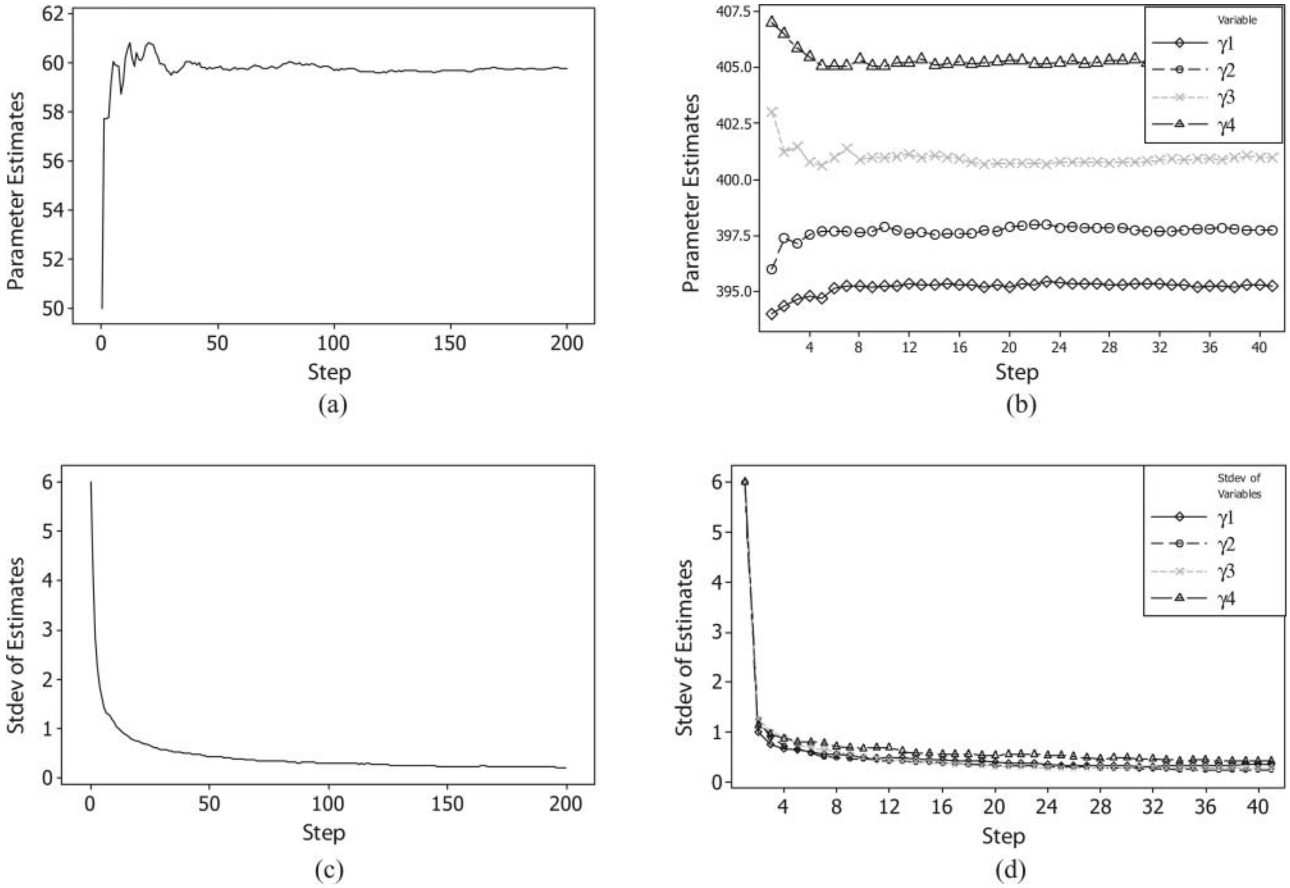

Fig. 2. Trajectories of estimated parameters: (a) the mean of a; (b) the means of γ; (c) standard deviation of a; and (d) standard deviation of γ.

deviations were all six. In addition, considering real scenarios, the cut-points were restricted such that $392 < \gamma_1 < \gamma_2 < \gamma_3 < \gamma_4 < 409$. The process was simulated to over 200 steps, which corresponded to 200 runs in the lapping process.

The Gibbs sampling method was set to repeat 10 000 times whenever a new observation was generated, and the last 5000 samples were used to calculate the marginal distribution of each unknown parameter. That is, the chain length was 10 000 and the burn-in period was 5000. Geweke's convergence diagnosit was used to check if this setting was able to ensure the convergence. Since the p-values for all of the estimated parameters in testing the first 10% and the last 50% samples of a single chain after the burn-in period were all larger than 0.1, we concluded that this choice was reasonable.

Figure 2 shows the trajectories of the mean and standard deviation of estimated a and cutoff values. We can see from Figs. 2(a) and 2(b) that the estimated parameters approach their true values gradually as categorical observations are collected run by run. Oscillation may exist in an early stage, since at the beginning there are an insufficient number of samples and the information contained in the categorical data is comparatively rough. Nevertheless, after around 40 steps, the estimates are already very close to their true values. From Figs. 2(c) and 2(d), it can be clearly seen that the standard deviations of the estimated parameters decrease sharply first and then gradually to reach a very small value, which shows that the online algorithm can give increasingly accurate estimates with smaller variance using a stream of categorical data.

As the adjusted ML method in Equation (10) can be used to estimate all of the parameters using categorical observations, in the following we compare its performance with the proposed Bayesian framework in terms of parameter estimation accuracy.

For different initial parameter settings, the process was replicated 100 times; the mean and estimation errors (absolute values between true and estimated parameters) and Mean Squared Error (MSE) at certain steps (10th, 50th, 100th, and 200th steps) of the 100 paths were calculated and are listed in Table 1. Note that the initial values for the adjusted ML method are the initial estimates before collecting any observation, whereas the ones for the Bayesian method refer to the prior means for unknown parameters, with all standard deviations in the prior distributions set to six, which is a very large value and gives no advantage to the Bayesian method.

Table 1. Mean of estimation errors and MSE

			Step			
Method	*Initial values*	*Statistics*	*10*	*50*	*100*	*200*
Adjusted ML	55 [395, 397, 402, 406]	Mean	1.34	0.50	0.27	0.21
		MSE	1.35	0.57	0.41	0.37
	50 [394, 396, 403, 407]	Mean	2.83	1.12	0.52	0.41
		MSE	2.83	1.24	0.87	0.91
	45 [393, 395, 404, 408]	Mean	4.07	1.86	1.17	0.80
		MSE	4.07	1.89	1.25	0.97
Bayesian	55 [395, 397, 402, 406]	Mean	0.58	0.39	0.29	0.20
		MSE	1.03	0.85	0.79	0.74
	50 [394, 396, 403, 407]	Mean	0.69	0.43	0.32	0.21
		MSE	0.93	0.64	0.54	0.43
	45 [393, 395, 404, 408]	Mean	0.58	0.41	0.31	0.24
		MSE	0.86	0.64	0.52	0.44

It can be seen from Table 1 that for any of the three studied cases, the Bayesian method performs much better than the adjusted ML method in terms of parameter estimation. For most cases, the Bayesian method gives a smaller mean and MSE, especially when the initial values of the estimated parameters are far from their true values. For the Bayesian method, the estimates approach true values very quickly in the first few steps; both the means of the errors and MSE are already very small after 10 steps. Its convergence rate becomes slower later on. Therefore, we conclude that the Bayesian method has a much shorter delay before precise estimates can be achieved and generates more accurate estimations when more categorical observations become available; this algorithm is also quite robust to the choice of initial values.

4.2. *A study of process control performance*

It is argued that when a process is contaminated by only white noise, there is no need to control the process. In Fig. 3, we show the sequences of controlled and uncontrolled output y_t under the same settings as used above.

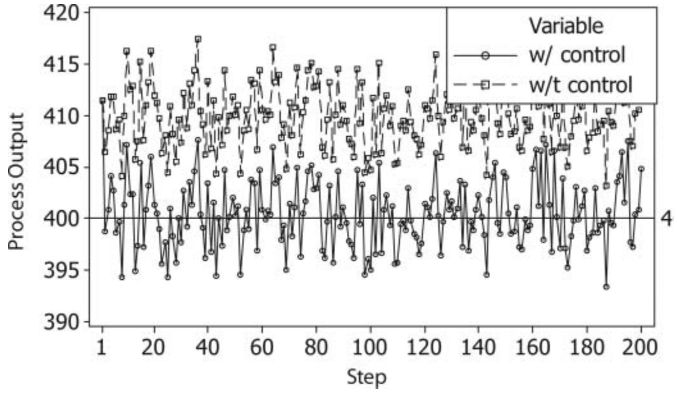

Fig. 3. Trajectories of process output with and without a controller.

Table 2. Control performance comparison of different controllers

			MSE			
			Step			
Controller	*CPU time (s)*	*Initial values*	*10*	*50*	*100*	*200*
Adjusted ML	0.28	55 [395, 397, 402, 406]	5.02	3.57	3.27	3.03
		50 [394, 396, 403, 407]	8.43	5.22	3.76	4.11
		45 [393, 395, 404, 408]	12.59	6.80	5.44	5.02
Bayesian	1320	55 [395, 397, 402, 406]	3.49	3.10	3.10	3.50
		50 [394, 396, 403, 407]	3.15	3.30	2.61	3.09
		45 [393, 395, 404, 408]	3.04	3.16	2.79	2.60
EWMA	0.0023	55, —	3.09	3.14	2.94	3.08
		50, —	3.56	2.98	3.21	3.03
		45, —	3.43	3.10	3.53	3.48
Categorical	0.0021	—	5.97	5.84	6.21	7.03

Figure 3 clearly suggests that the uncontrolled output strongly deviates from the target 6400, whereas the controlled output is almost maintained on target. Therefore, when initial bias is inevitable, it is still necessary to implement controllers to a process.

Next, we compare the control performance of the proposed method with existing controllers, including the categorical controller, the popular EWMA controller, and the adjusted ML method. For a better understanding of the simulation results, the treatments applied to the different controllers should be noted: (i) an accurate process output is provided to the EWMA controllers to generate control actions; and (ii) process parameters are assumed to be exactly known for the categorical controller to work. For comparison, the MSE from the target and the CPU time (obtained on a personal computer with a Core 2 Duo 3G Hz CPU and 2GB DDR2 memory) were calculated and are presented in Table 2.

We can see from Table 2 that comparing the Bayesian and EWMA controllers, their performances are quite close; the EWMA controller can usually compensate for initial bias faster since it has a smaller MSE in early steps. However, after a certain number of runs, when the Bayesian controller obtains a sufficiently accurate estimate of parameters, it also should have competitive performance.

Compared with the adjusted ML and categorical controllers, the Bayesian controller obviously gives the smallest MSE at almost all four steps, especially after 10 steps. Therefore, it shows that the Bayesian controller is more efficient in controlling this process.

To illustrate the differences in the control performances of these methods visually, the sequences of one realization of process outputs are shown in Fig. 4. We can see that the Bayesian and EWMA controllers can compensate for initial bias and return the process output back to the target value faster than the adjusted ML method. The categorical controller shows no bias since we assume all parameters are exactly known in setting up this controller. However, it shows quite a large variation.

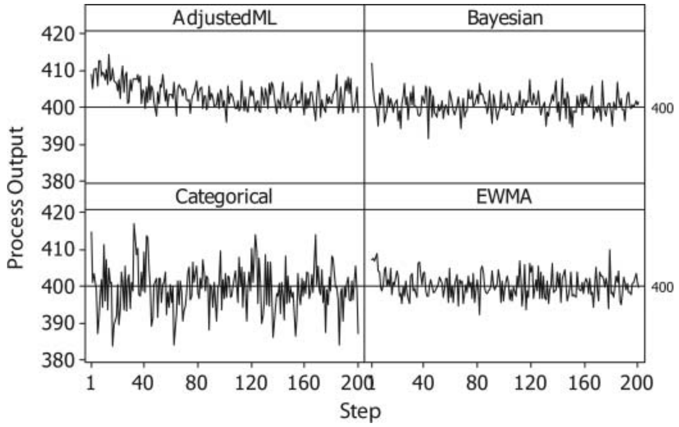

Fig. 4. Trajectories of controlled outputs.

Computational requirements should also be considered in the implementation of an online feedback controller. The CPU time in Table 2 is calculated for the whole 200 runs. It is clear that the Bayesian controller requires much more computation time than the other methods, which is due to the 10 000 nonsimulation length for the Gibbs sampling of each run, whereas the 1320 s for 200 runs implies that for each run, it takes about 6.6 s to generate the control action for the next run. Hence, the computation expense of the Bayesian controller is thought to also be acceptable in practice.

5. Conclusions

New algorithms are needed to model and control an R2R process when only categorical observations are available. This article proposed a Bayesian framework for online parameter estimation and updating when categorical observations are collected gradually. Based on the Bayesian estimation framework, we also derived a Bayesian controller for controlling a process. Simulation studies revealed that compared with existing methods, the proposed method can give better performance in both parameter estimation and process control.

In real applications, accurate numerical observations may become available after a certain delay. If such delayed but accurate information could be used to calibrate the model estimated based on categorical observations, the performance in terms of parameter estimation and process control is expected to be improved. This is an interesting and important topic for practitioners and should be studied in future research.

Acknowledgements

The authors would like to thank the editor and two anonymous referees for their valuable suggestions, which helped to significantly improve this manuscript. This work was supported by the National Natural Science Foundation of China (NSFC) under grant 70802034.

References

Chen, A. and Guo, R.S. (2001) Age-based double EWMA controller and its application to CMP processes. *IEEE Transactions on Semiconductor Manufacturing*, **14**(1), 11–19.

Chipman, H. and Hamada, M. (1996) Bayesian analysis of ordered categorical data from industrial experiments. *Technometrics*, **38**(1), 1–10.

Del Castillo, E. (2006) Statistical process adjustment: a brief retrospective, current status, and some opportunities for further work. *Statistica Neerlandica*, **60**(3), 309–326.

Del Castillo, E. and Hurwitz, A.M. (1997) Run-to-run process control: literature review and extensions. *Journal of Quality Technology*, **29**(2), 184–196.

Girard, P. and Parent, E. (2001) Bayesian analysis of autocorrelated ordered categorical data for industrial quality monitoring. *Technometrics*, **43**(2), 180–191.

Ingolfsson, A. and Sachs, E. (1993) Stability and sensitivity of an EWMA controller. *Journal of Quality Technology*, **25**(4), 271–287.

Jen, C.H. and Jiang, B.C. (2008) Combining on-line experiment and process control methods for changes in a dynamic model. *International Journal of Production Research*, **46**(13), 3665–3682.

Jin, M. and Tsung, F. (2009) Smith–EWMA run-to-run control schemes for a process with measurement delay. *IIE Transactions*, **41**(4), 346–358.

Lawrence, E., Bingham, D., Liu, C. and Nair, V.N. (2008) Bayesian inference for multivariate ordinal data using parameter expansion. *Technometrics*, **50**(2), 182–191.

Lin, J. and Wang, K. (2010) Online parameter estimation and run-to-run process adjustment using categorical observations. *International Journal of Production Research*, **49**(13), 1–15.

Othman, M.K., Dolah, A., Omar, N.A. and Yahya, M.R. (2006) Design of experiment (DOE) for thickness reduction of GaAs wafer using lapping process, in *Proceedings of the 2006 IEEE International Conference on Semiconductor Electronics*, IEEE Press, Piscataway, NJ, pp. 583–585.

Ruegsegger, S., Wagner, A., Freudenberg, J.S. and Grimard, D.S. (1999) Feedforward control for reduced run-to-run variation in microelectronics manufacturing. *IEEE Transactions on Semiconductor Manufacturing*, **12**(4), 493–502.

Saad, D. (1998) *On-line Learning in Neural Networks*, Cambridge University Press, Cambridge, UK.

Shang, Y., Wang, K. and Tsung, F. (2008) An improved run-to-run process control scheme for categorical observations with misclassification errors. *Quality and Reliability Engineering International*, **976**, 1–11.

Solla, S.A. and Winther, O. (1999) Optimal online learning: a Bayesian approach. *Computer Physics Communications*, **121**, 94–97.

Spanos, C.J. and Chen, R.L. (1997) Using qualitative observations for process tuning and control. *IEEE Transactions on Semiconductor Manufacturing*, **10**(2), 307–316.

Tseng, S.T., Tsung, F. and Liu, P.Y. (2007) Variable EWMA run-to-run controller for a drifted process. *IIE Transactions*, **39**(3), 291–301.

Vanli, O.A. and Del Castillo, E. (2009) Bayesian approaches for on-line robust parameter design. *IIE Transactions*, **41**(4), 359–371.

Wang, K. and Tsung, F. (2007) Run-to-run process adjustment using categorical observations. *Journal of Quality Technology*, **39**(4), 312–325.

Wang, K. and Tsung, F. (2008) An adaptive T^2 chart for monitoring dynamic systems. *Journal of Quality Technology*, **40**(1), 109–123.

Wang, K. and Tsung, F. (2010) Recursive parameter estimation for categorical process control. *International Journal of Production Research*, **48**(5), 1381–1394.

Biographies

Jing Lin received her M.E. degree in Management Science and Engineering from Tsinghua University, Beijing, China, in 2010. She has conducted research in statistical quality control and monitoring. Currently she is a business analyst at Baidu.com Inc. Her work focuses on strategy design, evaluation, and optimization.

Kaibo Wang is an Associate Professor in the Department of Industrial Engineering, Tsinghua University, Beijing, China. He received his B.S., and M.S. degrees in Mechatronics from Xi'an Jiaotong University, Xi'an, China, and his Ph.D. in Industrial Engineering and Engineering Management from the Hong Kong University of Science and Technology, Hong Kong. He has published papers in journals such as *Journal of Quality Technology*, *Quality and Reliability Engineering International*, *International Journal of Production Research*, and others. His research interests include statistical quality control, multivariate statistical process control, and data-driven complex system modeling/monitoring/diagnosis/control.