Setup Adjustment of Multiple Lots Using a Sequential Monte Carlo Method

Zilong LIAN

Department of Industrial and Manufacturing Engineering
The Pennsylvania State University
University Park, PA 16802

Bianca M. COLOSIMO

Dipartimento di Meccanica–Politecnico di Milano
Piazza Leonardo da Vinci 32
20133 Milano, Italy

Enrique DEL CASTILLO

Department of Industrial and Manufacturing Engineering
The Pennsylvania State University
University Park, PA 16802

A new sequential Monte Carlo (SMC) adjustment method is presented for solving the machine setup adjustment problem when process parameters are unknown. In setup adjustment problems, the mean of the distribution of the quality characteristic of parts can change from lot to lot due to an improper setup operation. It is shown how a first SMC approach has performance equivalent to a recently proposed Markov chain Monte Carlo method but at a small fraction of the computational cost, allowing for on-line control. A second, modified SMC rule that avoids unnecessary adjustments that can inflate the variance is also presented. A simulation approach is presented that allows tuning of the modified SMC rule to provide robust adjustment with respect to the unknown process parameters. Applications in short-run manufacturing processes are discussed.

KEY WORDS: Bayesian hierarchical models; Engineering process control; Random-effects model; Short-run manufacturing.

1. INTRODUCTION

In this article we consider the so-called "setup adjustment problem," first studied by Grubbs (1954). This problem refers to a machine that produces discrete parts in lots or batches, with the possibility of an improper setup operation resulting in an error or offset in the quality characteristic of interest. If there are considerable costs associated with producing off-target relative to the cost of adjusting a process, then adjusting is justified. This problem often arises in discrete machining processes that may experience large lot-to-lot variation. We discuss some instances within the range of applications of this problem in Section 3.

Grubbs proposed two solutions to this problem, depending on whether one considers a single lot of parts or multiple lots of parts with one setup operation before each lot is started. The objective in either case is to minimize the sum of squared deviations from target of the quality characteristic (i.e., the only relevant cost is a quadratic off-target cost). Grubbs's solutions to the setup adjustment problem were recently studied and extended by Trietsch (1998) and by Del Castillo, Pan, and Colosimo (2003a, b). In this article we focus on the second problem studied by Grubbs, the multiple-lot case, in which the initial offset can vary from lot to lot, and propose a Bayesian approach to its solution. One main difference from past efforts is that we make the more realistic assumption of unknown process parameters. The problem is then one of estimation and adjustment, not only of adjustment.

Recent work by Colosimo, Pan, and Del Castillo (2004) presents a Bayesian adjustment rule for the multiple-lot, unknown-parameters case, based on Markov chain Monte Carlo (MCMC) techniques (Gilks, Richardson, and Spiegelhalter 1996). Lian, Colosimo, and Del Castillo (2006) conducted a sensitivity analysis on the performance of this MCMC adjustment rule and modified it to obtain a rule that is more robust with respect to a wider set of process conditions.

Despite the good performance of the modified MCMC setup adjustment rule of Lian et al. (2006), this rule requires substantial computational time at each point in time, an obvious disadvantage if on-line control is needed for a process in which the time between parts is relatively short. This motivates the approach taken in the present article. As an alternative to MCMC methods, we consider sequential Monte Carlo (SMC) methods applied to the multiple-lot, unknown-parameters, setup-adjustment problem.

SMC methods also rely on Monte Carlo algorithms for the solution of Bayesian inference problems in which posterior distributions of the unknown parameters are created numerically from the generation of a large number of random variates or "particles." SMC techniques use the observation available at time $t+1$ to update the previous posterior distributions at time t (the priors at time $t+1$). The new posterior is then used to obtain updated inferences that are useful for adjustment purposes. In contrast to SMC techniques, MCMC methods do not use this sequential updating approach. Every time that a new observation is available, MCMC starts from the prior distributions available before observing data and passes through all data up to the current time to derive posterior distributions. Clearly, the computational efficiency of SMC techniques is relevant in production environments in which the time between consecutive parts is short and rapid on-line adjustments are required.

TECHNOMETRICS, AUGUST 2006, VOL. 48, NO. 3
DOI 10.1198/004017006000000066

A relatively new area of research, SMC techniques have appeared in the last few decades in the statistical literature under such names as sampling/importance resampling (SIR), bootstrapping filters, and particle filters, among others. These techniques have great potential in process control applications where parameters are unknown and rapid on-line computations are necessary.

The remainder of this article is organized as follows. Section 2 presents the statistical model that we consider for the multiple-lot adjustment problem and techniques that were previously proposed for the solution of its different versions. Section 3 contains a discussion of areas of applications of the proposed model and methods. Section 4 presents a first rule based on SMC techniques and indicates the reasons that lead us to the particular choice of SMC techniques that we recommend. This section also presents an illustrative example of the use of the proposed rule in an air-bending process. Section 5 presents a study of the performance of the first SMC adjustment rule compared with other existing rules for setup adjustment. The effects resulting from changing the prior parameters and changing the prior distributions themselves are studied. A problem when using the first SMC method (inflation in variance) is identified when the mean of the lot-to-lot offset distribution is actually 0. To avoid this problem, Section 6 presents a second, modified SMC adjustment rule with improved performance. Finally, Section 7 summarizes the conclusions of this research.

2. PROCESS MODEL AND EXISTING ADJUSTMENT PROCEDURES

Consider a multilot discrete-part manufacturing process, in which an initial offset can occur during the setup of each lot. The assumed statistical model can be expressed as

$$Y_{ij} = \theta_{ij} + \upsilon_{ij}, \tag{1}$$

$$\theta_{ij} = \theta_{i(j-1)} + U_{i(j-1)}, \tag{2}$$

$$\theta_{i0} \sim (\mu, \sigma_\theta^2), \tag{3}$$

$$\upsilon_{ij} \sim (0, \sigma_\upsilon^2), \tag{4}$$

where $i = 1, \dots, I$ is the lot or batch index, $j = 1, \dots, J$ is the part index, and Y_{ij} is the deviation from the nominal value for the quality characteristic associated with the jth part in the ith batch [part (i, j)]. Parameter θ_{ij} is the unknown mean deviation for the jth part in the ith batch. $\theta_{i0} \sim (\mu, \sigma_\theta^2)$ represents the initial offset due to setup errors; thus σ_θ^2 is the lot-to-lot variance. No specific distributional assumption is made on θ_{ij}, because the SMC procedure does not require it. The iid part-to-part errors, $\upsilon_{ij} \sim (0, \sigma_\upsilon^2)$, are due to the combined effect of the intrinsic variability in the process and in the measurement system. Normality was used for the distribution of these errors in the simulations given later, but this is not an assumption of the model.

Let U_{ij} be the adjustment made after the jth part in the ith lot is produced (i.e., after Y_{ij} is observed). It can be represented as the difference between the levels of two consecutive control variables or setpoints, that is, $U_{ij} = u_{ij} - u_{i(j-1)}$. Equation (2) can then be rewritten as

$$\theta_{ij} = \theta_{i0} + u_{i(j-1)}, \tag{5}$$

where we assume that $u_{ij} = 0$ for $j < 0$. An *adjustment policy* can then be represented by the series u_{ij} $(i \geq 1, j \geq 0)$ or by a rule (the *adjustment rule*), a mathematical expression that indicates how to set these values. Note how for $j = 0$, u_{i0}, the initial setpoint, must be selected as well. As shown later, some existing adjustment rules simply set u_{i0} to the "neutral" value of 0, in the absence of prior information on the distribution of the offsets. In contrast, the class of Bayesian rules that we propose allows us to use a better initial setpoint in each lot, "anticipating" the initial error of each lot.

The parameters μ, σ_θ, and σ_υ in equations (1)–(4) are assumed to be unknown. This places the resulting adjustment problem in the adaptive control category (see, e.g., Aström and Wittenmark 1994). Note that Kalman filtering schemes are not applicable, because the process variances are unknown. The adjustments are assumed to contain no error. In manufacturing applications, the adjustments typically constitute a simpler operation than setting up a new lot, for which errors are more common.

We assume that we can always set the desired mean deviation to be 0; that is, over the region of operation of the process, there are no constraints on the magnitudes of the adjustments. The performance of an adjustment rule is evaluated by a quadratic symmetric loss function over all parts, given by

$$C = \sum_{i=1}^{I} \sum_{j=1}^{J} Y_{ij}^2. \tag{6}$$

In this article we do not consider adjustment costs. [For treatments of adjustment costs and errors in the adjustments in (known parameter) setup adjustment problems, see Trietsch 1998; Del Castillo et al. 2003a.] We further discuss the applicability of these assumptions in Section 3.

Several adjustment rules can be applied to this problem based on different assumptions. For instance, we can apply *Grubbs's harmonic rule* (Grubbs 1954), where $U_{ij} = -\frac{1}{j} Y_{ij}$ $(U_{i0} = 0 \; \forall \, i)$. This is optimal if the initial offsets θ_{i0} are totally unpredictable [or $\mu = 0$ and $\sigma_\theta^2 \to \infty$ in (3)]. In practice, however, the initial offsets follow a certain distribution (3) that can be inferred and used to better predict the offsets. Note how the first part in each lot goes uncontrolled.

Grubbs's extended rule (Grubbs 1954) can also be applied. This, given by $U_{ij} = -Y_{ij}/(j + \sigma_\upsilon^2/\sigma_\theta^2)$, requires knowing $\sigma_\upsilon^2/\sigma_\theta^2$ and assumes that $\mu = 0$.

A recently proposed rule that also could be applied is the *Bayesian adjustment rule* of Colosimo et al. (2004), in which desired parameters are estimated sequentially through posterior distributions using MCMC, in particular, Gibbs sampling, and the adjustments are made accordingly. This method, in contrast to Grubbs's rules, makes it possible to eventually predict the offset so that the first part in each lot is controlled.

In later sections we contrast these methods with the proposed SMC adjustment rule. We next discuss the application of the model and of the proposed solution methods in practice.

3. APPLICABILITY OF THE SETUP ADJUSTMENT PROBLEM AND OF THE PROPOSED SOLUTION IN INDUSTRIAL PRACTICE

The model given by (1)–(4) is essentially [after a simple transformation; see (7)] a one-way random-effects model, common in the SPC literature on variance components (see, e.g., Woodall and Thomas 1995) and used for multiple-lot setup adjustment by Colosimo et al. (2004) and Lian et al. (2006). The model is in fact considerably more flexible than a standard one-way random-effects model, because no restrictions are placed on the distribution of the random effects.

One of the key assumptions of the model is the lack of serial correlation in the observations. In addition, the solution that we propose provides significant savings over existing methods when J, the lot size, is small. Finally, the cost function assumes no adjustment costs. We now discuss how this set of assumptions and conditions holds true for an important class of real manufacturing processes. An actual illustration of the solution method is presented in Section 4, where the first SMC method that we propose is given.

In general, the assumed model provides a good approximation of reality in discrete-part manufacturing processes that do not exhibit process or noise dynamics (Del Castillo 2002). An example where the model assumptions hold is precision sheet metal bending, schematically represented in Figure 1. In air bending, a blank sheet is supported by two shoulders of a stationary die and the depth of the punch stroke (which acts as a controllable parameter) determines the bend angle obtained on the workpiece after unloading (the response variable). The actual angle obtained on the unloaded sheet depends on the "springback," that is, the elastic recovery of the original geometry, as represented in Figure 1. Springback is affected mainly by the thickness and the mechanical properties of the blank sheet, which vary from batch to batch because of different suppliers or different conditions of the supplier's process (Elkins and Sturges 2001). Therefore, the observed bent angles will often be biased with respect to the target angle, and the punch depth can be used to compensate for this offset. Each time a lot of a new material is to be bent, a setup operation needs to be done which involves further lot-to-lot variability. The lot sizes of parts bent in precision machining are typically small (sizes of fewer than 10 parts are common). Siu (2004) recently studied the effect of controlling the punch depth in sheet metal bending (see Fig. 1). He showed experimentally how the process is characterized by the absence of dynamics and a natural variability of the bent angle that can be modeled as white noise.

Another important example in which the model assumptions hold and fabrication is done in small lots is in metal machining processes in the aerospace sector. Koons and Luner (1991) reported a SPC study at McDonnell Aircraft Co. on an end-milling process where "typical lot sizes range from 12 to 48 units." They noted how "all of the isolated special causes of variation that were identified were traceable to improper machine setup," and how the basic corrective action was to improve the setup practices. They further noticed how operators made adjustments because the first few pieces were sometimes off-target. They concluded that "modifications to the process, such as adjusting machines, may eliminate the effects" of the observed setup error. However, their article was on process monitoring, and the adjustment procedures were not discussed further. This seems to be an ideal scenario for the proposed models and methods.

Setup adjustment methods, particularly multivariate ones (Del Castillo et al. 2003a), also have applications in certain semiconductor manufacturing processes (Moyne, Del Castillo, and Hurwitz 2000). Consider, for instance, a photolithography process. Test policies for this process involve processing a wafer from a lot, inspecting it, adjusting the parameters of the process if necessary, and processing a new wafer until the stepper machine is "qualified" (Akcalt, Nemoto, and Uzsoy 2001). Usually, several short lots are processed in a stepper machine, with reticle changes that can induce setup errors between lots. The adjusting procedure can be based on setup adjustment techniques.

Although the main aim of the methods presented here is application in high-precision short-run manufacturing processes, there are other applications beyond what we have described as the setup adjustment problem. [In the manufacturing literature, the setup adjustment problem is sometimes called "process positioning" (Bjorke 1989).] For example, Grubbs (1954) described potential applications to calibration and even to artillery.

In all of these applications, the adjustment costs are minor compared with the off-target cost, as adjustments imply changing a setting in a computer-controlled machine; thus they can be neglected, in accordance with the assumed cost function (6).

4. A SEQUENTIAL MONTE CARLO APPROACH

The model described by (1)–(4) can be rewritten as a one-way random-effects model by substituting (5) in (1), thus ob-

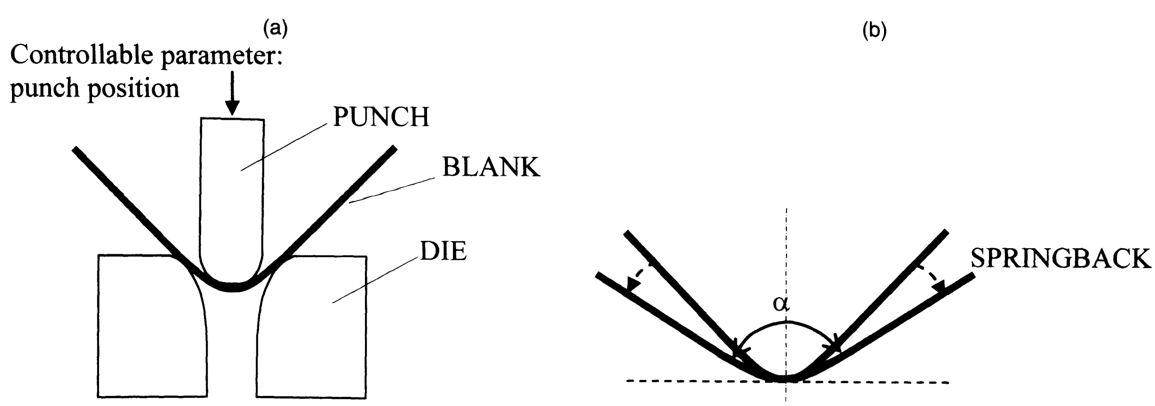

Figure 1. Schemes of the Air-Bending Process (a) and the Springback Phenomenon (b), Where α Represents the Bent Angle.

taining

$$Y_{ij} = \theta_{i0} + u_{i(j-1)} + \upsilon_{ij}.$$

From this last equation, we define the transformed variable,

$$X_{ij} = Y_{ij} - u_{i(j-1)} = \theta_{i0} + \upsilon_{ij}, \quad (7)$$

which can be computed each time after a new part is processed. Equation (7) together with (3) and (4) form a traditional one-way random-effects model—provided that the random effects are normal—and adjustments can be made as was done by Colosimo et al. (2004) at the following two levels:

1. Before processing a new lot i (where $i \geq 3$), the initial setpoint u_{i0} that must be set on the machine before processing lot i is based on the predictive distribution of the setup offset, that is,

$$U_{i0} = u_{i0} - u_{i(-1)} = u_{i0}$$
$$= \begin{cases} -\hat{\mu}|D_{(i-1)J} & \text{for } i \geq 3 \\ 0 & \text{for } i = 1, 2, \end{cases} \quad (8)$$

where $\hat{\mu}|D_{(i-1)J}$ is the mean of the posterior distribution $p(\mu|D_{(i-1)J})$ and

$$D_{(i-1)J} = \{x_{11}, \ldots, x_{(i-1)1}, \ldots, x_{(i-1)J}\}$$

denotes all of the (transformed) data observed before lot i starts (i.e., after the last part in the previous lot $i-1$ is observed).

2. After observing at least one part in a lot, adjustments are based on the posterior distribution of the offset in the current lot,

$$U_{ij} = -\hat{\theta}_{ij}|D_{ij} \quad \text{for } j = 1, 2, \ldots, J-1 \qquad \text{or}$$
$$u_{ij} = -\hat{\theta}_{i0}|D_{ij}, \quad (9)$$

where $\hat{\theta}_{i0}|D_{ij}$ is the mean of the posterior distribution $p(\theta_{i0}|D_{ij})$ and $D_{ij} = \{x_{11}, \ldots, x_{ij}\}$ denotes all of the (transformed) data observed before the $(j+1)$th part in the ith lot is processed.

In this article the following priors were used:

$$\mu|\mu_0, \sigma_0^2 \sim \mathrm{N}(\mu_0, \sigma_0^2), \quad (10)$$

$$\sigma_\theta^2|\mu_1, \sigma_1 \sim \mathrm{LN}(\mu_1, \sigma_1^2), \quad (11)$$

and

$$\sigma_\upsilon^2|\mu_2, \sigma_2 \sim \mathrm{LN}(\mu_2, \sigma_2^2), \quad (12)$$

where μ_0, σ_0^2, μ_1, σ_1^2, μ_2, and σ_2^2 are known constants and LN stands for a lognormal distribution. A lognormal prior for the variances was chosen because (a) SMC does not require the use of conjugate priors; (b) the LN provides more "robust" behavior than other priors, as explained in Section 5.3; (c) making a conjugate inverse-gamma "noninformative" is actually difficult, a fact that has not been recognized until recently (Spiegelhalter, Abrams, and Myles 2004); and, in contrast, (d) the degree of information of the prior can be easily changed using the lognormal.

4.1 Overview of the First SMC Method and Illustrative Example

To implement the adjustment rule, the posterior distributions of the process parameters must be computed at each step when a new part is observed. In this case the MCMC setup adjustment method (Colosimo et al. 2004) becomes time-inefficient as the size of the dataset grows. With SMC methods, the computational effort does not grow with the size of the total dataset.

SMC methods consist of generating a set of draws or "particles" from the prior distributions of unknown parameters and associating a weight with each set of particles. These weights are sequentially updated each time new data are observed, as described in Section 4.2. Despite the benefits of sequential updating, this method poses some implementation difficulties in practice. The problem arises if the generated particles remain the same throughout all iterations and only the weights or the frequencies associated with the particles change. In this case, a phenomenon known as the "degeneracy" of the sample can arise. Degeneracy means that after some iterations of SMC, just a few of the N original particles will have weights >0, producing biased estimates of the distribution of the parameters. The degeneracy problem is particularly severe when the original prior distributions have large variances or high dimension, when the number of particles is small, or when the observed dataset is large. The degree of this type of degeneracy due to an "impoverished" sample can be monitored using the effective sample size (ESS) statistic (Kong, Liu, and Wong 1994; Ridgeway and Madigan 2002), as described in Section 4.2.

Recently, Balakrishnan and Madigan (2004) proposed a one-pass particle filtering (1PFS) algorithm, in which a "rejuvenation" step is used to disperse the particles when ESS drops below a specific level and reduce in this way the degeneracy. In this article we use the rejuvenation step from the 1PFS algorithm and apply it to multilot setup adjustment problems for distribution computations. It turns out that this regeneration step is crucial for successful implementation of the SMC methods.

A MATLAB computer program that implements the SMC method as described in this section can be downloaded from the website given at the end of the article. A detailed algorithmic description of the estimation steps in the SMC method is given in Section 4.2. We first illustrate the use of the method in actual practice.

Example. We apply the SMC procedure to the experimental data obtained by Siu (2004), who studied the air-bending process described in Section 3. The controllable factor is the punch depth (position), and the response of interest to control is the angle, α, of the bend. Whenever a sheet of a different material is to be processed, a new setup operation is necessary.

Several different materials are to be bent in lot sizes of five parts. The collection of all materials (and thicknesses), in addition to improper setup operations by operators, implies a distribution on the lot-to-lot means, such as (3). In addition, measurement error and part-to-part variability imply a distribution like expression (4). A common target angle in air bending is 90 degrees, which we use, so the deviations from target are $Y_{ij} = \alpha_{ij} - 90$ and follow the model (1)–(4), where we assume normality in (3) and (4). The priors of σ_θ^2 and σ_υ^2 were set with

Table 1. Illustration of First SMC Adjustment Rule, Air-Bending Example

i	j	u_{ij}	U_{ij}	d_{ij}	Y_{ij}	α_{ij}	*Int*	*K*	*Material*
1	0	0		1.63			−156	150.85	
	1	−5.10	−5.10	1.60	5.26	95.26	−156	150.85	.025 steel
	2	−5.16	−.06	1.60	.11	90.11	−156	150.85	.025 steel
	3	−5.33	−.17	1.60	.46	90.46	−156	150.85	.025 steel
	4	−5.36	−.03	1.60	.09	90.09	−156	150.85	.025 steel
	5	−5.36		1.60	.12	90.12	−156	150.85	.025 steel
2	0	0		1.70			−153.3	143	
	1	−4.11	−4.11	1.67	2.24	92.24	−153.3	143	.02 steel
	2	−3.63	.48	1.68	−1.50	88.50	−153.3	143	.02 steel
	3	−3.23	.40	1.68	−1.53	88.47	−153.3	143	.02 steel
	4	−2.96	.27	1.68	−1.16	88.84	−153.3	143	.02 steel
	5	−2.96		1.68	−.82	89.18	−153.3	143	.02 steel
3	0	−3.92		1.60			−156.9	151.9	
	1	−4.00	−.08	1.60	.22	90.22	−156.9	151.9	.032 aluminum
	2	−4.08	−.08	1.60	.27	90.27	−156.9	151.9	.032 aluminum
	3	−4.17	−.09	1.60	.36	90.36	−156.9	151.9	.032 aluminum
	4	−4.17	0	1.60	−.01	89.99	−156.9	151.9	.032 aluminum
	5	−4.17		1.60	−.20	89.80	−156.9	151.9	.032 aluminum
4	0	−4.08		1.67			−153.3	143	
	1	−4.51	−.43	1.67	.91	90.91	−153.3	143	.02 steel
	2	−4.58	−.07	1.67	.20	90.20	−153.3	143	.02 steel
	3	−4.47	.11	1.67	−.43	89.57	−153.3	143	.02 steel
	4	−4.53	−.06	1.67	.24	90.24	−153.3	143	.02 steel
	5	−4.53		1.67	.50	90.50	−153.3	143	.02 steel
5	0	−4.31		1.60			−156	150.85	
	1	−4.27	.04	1.60	−.04	89.96	−156	150.85	.025 steel
	2	−4.15	.12	1.60	−.37	89.63	−156	150.85	.025 steel
	3	−4.05	.10	1.60	−.41	89.59	−156	150.85	.025 steel
	4	−4.00	.05	1.60	−.25	89.75	−156	150.85	.025 steel
	5	−4.00		1.60	.03	90.03	−156	150.85	.025 steel

the following approach. A 95% credible region on the parameter was set on values believed to be a priori likely for each parameter. The two percentiles provide two equations from which the parameters μ_i and σ_i in (11) and (12) can be obtained. For the air-bending example, it was believed a priori that the lot-to-lot variance σ_θ^2 was in the range (.1, 10) and that the measurement noise variance σ_υ^2 was much smaller, in the range (.1, 1.0). From the cumulative distribution function of the lognormal and the two 95% percentiles, we obtain $\mu_1 = 1.1276$, $\sigma_1 = .7142$, $\mu_2 = -.9504$, and $\sigma_2 = .5528$.

An important practical problem when applying a feedback controller is appropriately scaling the "gains," that is, the coefficients multiplying the controllable factor(s). In expression (7), the gain is evidently 1.0, but this will not be true in general, and a transformation is necessary. Siu (2004) fitted simple linear regression models between the observed bend angle, α, and the actual depth of cut, d, of the form $\alpha = int + Kd$, where int is some intercept and K is the actual gain. These models depend on the material type; the corresponding intercepts and gains are given in Table 1. Depending on the material type, the relation between the scaled controllable factor, u, and the actual punch depth, d, is given by $d_{ij} = (u_{ij} + T - int)/K$, where T is the target (90 degrees in this illustration). Table 1 shows the adjustments and observed angles for five different lots of five parts using the SMC rule. (The Matlab program was used to compute the adjustments U_{ij}.) Figure 2 shows a time graph of the setpoints (unscaled d_{ij} and scaled u_{ij}) and the corresponding observed angles α_{ij}. As can be seen, the largest adjustment (largest change in the setpoint) is in the first part in each lot, with the SMC rule able to "anticipate" the error in the first part in each lot starting from lot 3.

Figure 3 shows graphs of the posterior distributions obtained by the SMC procedure at the end of the five lots. This is additional useful information for the process engineer, because it

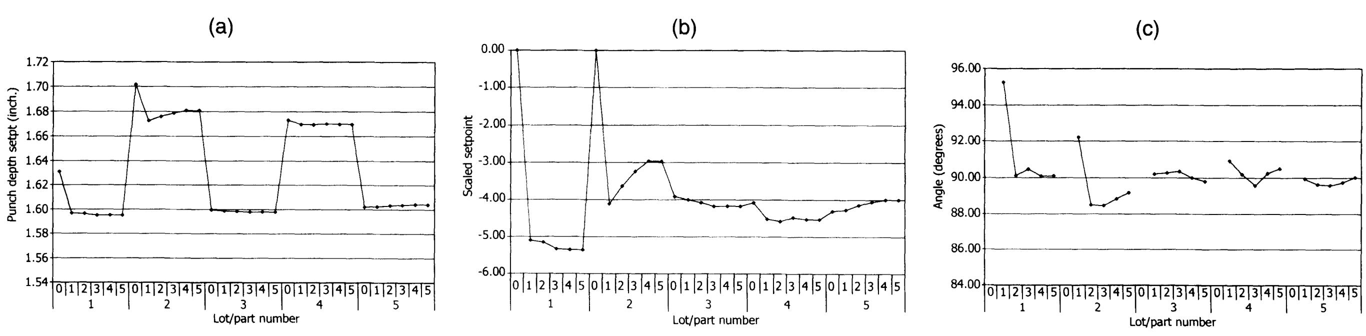

Figure 2. Punch Depth Setpoints d_{ij} (a), Scaled Setpoints u_{ij} (b), and Observed Angles α_{ij} (c), Air-Bending Example.

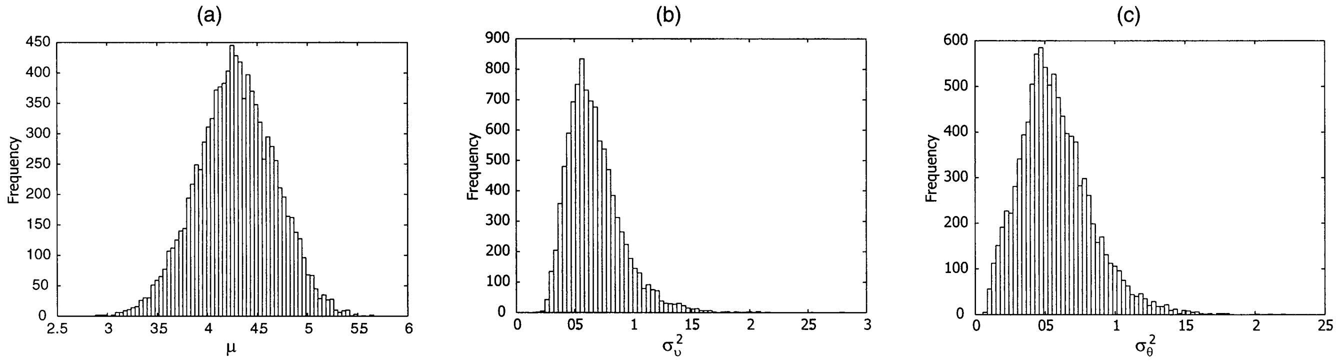

Figure 3. Final Posterior Distributions of μ (a), σ_υ^2 (b), and σ_θ^2 (c) Obtained at the End of the Fifth Lot, Air-Bending Example.

can be used to center the machine once sufficient precision has been reached in the lot-to-lot offset μ.

4.2 Algorithmic Details of the First SMC Adjustment Method

To better describe in detail our basic (first) SMC adjustment rule, consider (8) and (9). The adjustments are selected based on the posterior distributions $p(\mu|D_{(i-1)J})$ and $p(\theta_{i0}|D_{ij})$. In this section we use a more general notation in which $\psi = \{\psi_1, \dots, \psi_p, \dots, \psi_P\}$ denotes the vector of P unknown parameters in the model and t represents the total number of parts machined thus far at a given time. In other words, if we consider the time instant in which the jth part of the ith lot has been completed, then t is given by $t = (i-1) \times J + j$. Consequently, $D_t = \{x_1, x_2, \dots, x_t\}$ represents all of the data observed by time t. At time $t = (i-1) \times J + j$, the vector ψ comprises all unknown parameters $\{\theta_{10}, \theta_{20}, \dots, \theta_{i0}, \mu, \sigma_\theta^2, \sigma_\upsilon^2\}$, and the posterior distribution $p(\psi|D_t)$ can be computed using Bayes's theorem,

$$p(\psi|D_t) = \frac{p(\psi)p(D_t|\psi)}{\int p(\psi)p(D_t|\psi)\,d\psi} = \frac{q(\psi|D_t)}{m(D_t)} \propto q(\psi|D_t), \qquad (13)$$

where $p(D_t|\psi)$ is the likelihood of observing data in D_t, $p(\psi)$ is the prior distribution of unknown parameters, $m(D_t) = \int p(\psi)p(D_t|\psi)\,d\psi$ is the marginal distribution that is acting as a normalizing constant, and $q(\psi|D_t) = p(\psi)p(D_t|\psi)$ represents the *unnormalized* posterior density. A closed form of this posterior distribution is available only under specific conditions (e.g., using conjugate analysis for simple, nonhierarchical problems). For the hierarchical one-way random-effects model that we are dealing with, no closed form is available for the posterior.

Simulation-based approaches allow us to circumvent the problem of directly computing the posterior in (13) by drawing samples from this posterior. If N iid samples, $\psi^{(n)}$ $(n = 1, 2, \dots, N)$, are drawn from the posterior distribution $p(\psi|D_t)$, then any desired estimate can be computed using Monte Carlo integration,

$$\begin{aligned} J &= E(h(\psi)|D_t) \\ &= \int h(\psi)p(\psi|D_t)\,d\psi \approx \hat{J} = \frac{1}{N}\sum_{n=1}^{N} h(\psi^{(n)}), \end{aligned} \qquad (14)$$

where, for example, if $h(\psi) = \psi$, (14) allows computation of the posterior mean.

Therefore, the problem now is how to sample draws from the posterior distribution. We used the sequential importance sampling (SIS) approach, a particular SMC method (Doucet, de Freitas, and Gordon 2001). This method is a sequential variant of the SIR procedure presented by Rubin and described by Smith and Gelfand (1992).

The SIR approach consists of drawing N samples, called "particles," of all unknown parameters $\psi = (\psi_1, \dots, \psi_P)$ from a convenient distribution, $g(\psi)$, which is called the *importance sampling distribution*, provided that its support contains the support of the posterior from which we want to sample. By using these draws, desired posterior estimates can be derived by slightly manipulating (14) as

$$\begin{aligned} J = E(h(\psi)|D_t) &= \int h(\psi)\frac{p(\psi|D_t)}{g(\psi)}g(\psi)\,d\psi \approx \hat{J} \\ &= \frac{1}{N}\sum_{n=1}^{N} h(\psi^{(n)})w_t^{(n)}, \end{aligned} \qquad (15)$$

where

$$w_t^{(n)} = \frac{p(\psi^{(n)}|D_t)}{g(\psi^{(n)})}, \qquad n = 1, \dots, N, \qquad (16)$$

are weights that should be computed for each draw sampled from $g(\cdot)$. This approximation improves as N grows.

As often happens in Bayesian analysis, the posterior is known up to the normalizing constant, which implies that, given the nth particle $(n = 1, \dots, N)$, we cannot compute $p(\psi^{(n)}|D_t)$ in (16) but can compute the unnormalized density $q(\psi^{(n)}|D_t) \propto p(\psi^{(n)}|D_t)$. In this case, we can still use the importance sampling approach by first computing the unnormalized weights as

$$w_t^{(n)} = \frac{q(\psi^{(n)}|D_t)}{g(\psi^{(n)})} \qquad (17)$$

and then normalizing them to get the desired estimates,

$$J = E(h(\psi)|D_t) \approx \hat{J} = \frac{1/N\sum_{n=1}^{N} h(\psi^{(n)})w_t^{(n)}}{1/N\sum_{n=1}^{N} w_t^{(n)}}. \qquad (18)$$

A further step, called "resampling," usually allows better approximation of the desired estimates. Given the set of generated samples, the resampling step draws from the discrete distribution $\psi^{(1)}, \dots, \psi^{(n)}, \dots, \psi^{(N)}$ allocating mass $w_t^{(n)}$ to the nth sample.

When data arrive sequentially, SIS uses past particles generated before observing the last outcome from the process. To show how SIS works, consider the special case in which the importance sampling function $g(\psi)$ is the prior distribution $p(\psi)$ in (13). Because $q(\psi|D_t)$ is the product of the likelihood times the prior (13), the weights given by (17) can be rewritten as

$$w_t^{(n)} = \frac{q(\psi^{(n)}|D_t)}{g(\psi^{(n)})} = \frac{p(D_t|\psi^{(n)})p(\psi^{(n)})}{p(\psi^{(n)})} = p(D_t|\psi^{(n)}). \quad (19)$$

As shown by Smith and Gelfand (1992), this resampling strategy simply means that more weight is given to prior samples that are more likely to occur. In this case, the unnormalized weight associated with the nth particle can be computed as

$$\begin{aligned} w_t^{(n)} &= p(D_t|\psi^{(n)}) = p(x_t|\psi^{(n)})p(D_{t-1}|\psi^{(n)}) \\ &= p(x_t|\psi^{(n)})w_{t-1}^{(n)}. \end{aligned} \quad (20)$$

Therefore, the weight associated with the nth particle changes according to the likelihood of observing x_t evaluated at each particle $\psi^{(n)}$. This expression can be used to sequentially update the weights each time that new data are observed and thus is applicable in situations in which data arise sequentially.

A problem associated with SMC approaches relates to the "degeneracy" of the sample of particles. When the N particles are generated from noninformative priors, most of these particles (those that are less likely, given the data observed) will have weights equal to 0 after few iterations. In these cases particles associated with weights >0 will be very few, say $N^* \ll N$. This "degeneration" of the initial number of N particles to a much smaller number, N^*, will deeply affect the effectiveness of the SMC approach, inducing biased estimates of the unknown parameters. This problem is severe when the original prior distributions have large variances or high dimension, when the number of particles is small, or when the observed dataset is large (i.e., the number of iterations of the SMC method is large).

The degree of this type of degeneracy due to an "impoverished" sample (Balakrishnan and Madigan 2004) can be monitored using the ESS (Kong et al. 1994; Ridgeway and Madigan 2002). At time index t, this is given by

$$ESS_t = \frac{N}{1 + N^2 \operatorname{var}(w_t^{(n)})}, \quad (21)$$

where N is the original number of particles and $w_t^{(n)}$ is the weight associated with particle n at time t. In (21) the degree of degeneracy is described by the variance of the weights $w_t^{(n)}$. If at time t just few particles have associated weights >0, then the variance, $\operatorname{var}(w_t^{(n)})$, will be large, resulting in a smaller ESS_t.

To overcome the degeneracy problem, a rejuvenation step of the N particles can be performed using the 1PFS algorithm described by Balakrishnan and Madigan (2004). This involves approximating the posterior densities of particles with a shrinkage kernel-smoothing method. After the rejuvenation step, a resampling step is then performed. In this case the new draws are resampled from the rejuvenated particles.

The basic SMC algorithm, including the test on degeneracy and possible rejuvenation step, is summarized as follows:

Algorithm for estimating the process parameters μ, θ_{i0}, σ_θ^2, and σ_υ^2. At the beginning of the process:

- Draw N random numbers $\mu^{(n)}$ from the prior distribution of μ.
- Draw N random numbers $\sigma_\theta^{2(n)}$ from the prior distribution of σ_θ^2.
- Draw N random numbers $\sigma_\upsilon^{2(n)}$ from the prior distribution of σ_υ^2, $n = 1, 2, \dots, N$.
- Create an initial weight vector $(w^{(1)}, w^{(1)}, w^{(2)}, \dots, w^{(N)})$, where $w^{(n)} = 1/N$ for all n.

Iterations throughout all lots/parts processed:

For $i = 1, 2, 3, \dots, I$, generate one random number, $\theta_{i0}^{(n)}$, from the distribution $\mathrm{N}(\mu^{(n)}, \sigma_\theta^{2(n)})$ for each n.

For $j = 1, 2, 3, \dots, J$:

- Obtain the new observation y_{ij} and calculate the new variable $x_{ij} = y_{ij} - u_{i(j-1)}$.
- Calculate the likelihood of the nth particle [combination $(\mu^{(n)}, \theta_{i0}^{(n)}, \sigma_\theta^{2(n)}, \sigma_\upsilon^{2(n)})$],

$$L^{(n)} = \frac{1}{\sqrt{\sigma_\upsilon^{2(n)}}} \exp\left\{-\frac{(x_{ij} - \theta_{i0}^{(n)})^2}{2\sigma_\upsilon^{2(n)}}\right\}.$$

- Update the new weight vector and normalize it:

$$w^{(n)} \leftarrow w_n \times L^{(n)}, \quad \text{then}$$

$$w^{(n)} \leftarrow w^{(n)} \Big/ \sum_{l=1}^{N} w^{(l)}.$$

- Obtain the new parameter estimators:

$$\hat\theta_{i0}|D_{ij} = \sum_{n=1}^{N} w^{(n)}\theta^{(n)} \qquad \text{and} \qquad \hat\theta_{ij}|D_{ij} = \hat\theta_{i0}|D_{ij} - u_{ij},$$

$$\hat\sigma_\upsilon^2|D_{ij} = \sum_{n=1}^{N} w^{(n)}\sigma_\upsilon^{2(n)}, \qquad \hat\mu|D_{ij} = \sum_{n=1}^{N} w^{(n)}\mu^{(n)},$$

and

$$\hat\sigma_\theta^2|D_{ij} = \sum_{n=1}^{N} w^{(n)}\sigma_\theta^{2(n)}.$$

- Calculate the effective sample size factor,

$$ESS = \frac{N}{1 + N^2 \operatorname{var}(w^{(n)})}.$$

If $ESS < N/2$, then rejuvenate N particles using the 1PFS algorithm (Balakrishnan and Madigan 2004).

We now study the performance of this basic SMC rule compared with previous approaches. A modified SMC rule is presented in Section 6.

5. COMPARISON WITH PREVIOUS ADJUSTMENT RULES AND PERFORMANCE ANALYSES

5.1 Performance Comparison With Grubbs's Harmonic Rule and the MCMC Adjustment Rule

Grubbs's harmonic rule does not require knowing any process parameters and is used as a benchmark in this section. The relative savings in quadratic cost induced by the proposed SMC adjustment rule compared with Grubbs's harmonic rule are

$$S_s = \frac{C_h - C_s}{C_h} = 1 - \frac{C_s}{C_h}, \tag{22}$$

where C_h is the quadratic cost of a process adjusted by Grubbs's harmonic rule and C_s is the quadratic cost of the process adjusted using the SMC method. The savings obtained by using the MCMC approach relative to using the harmonic rule (S_m) were computed similarly.

To compare the SMC and the MCMC adjustment rules with Grubbs's rule in various production situations, 16 different cases were simulated, as given in Table 2. The lot number, I, was fixed at 20. The prior distribution of μ is usually easy to choose. Here μ_0 was set to 0, and σ_0^2 was chosen to be 10,000. To investigate the impacts of the prior distributions for σ_υ^2 and σ_θ^2 on the adjustment performance, nine additional scenarios were created, as given in Table 3. A prior can be considered to be relatively accurate when its mode is close to the true parameter (prior scenarios 1–3) or to be inaccurate otherwise (underestimated in scenarios 4 and 5, overestimated in scenarios 7–9). A higher variance gives a more vague prior (e.g., scenarios 3, 6, and 9), and a smaller variance represents more confidence on the prior (e.g., scenarios 1, 4, and 7).

For each combination of case (Table 2) and prior scenario (Table 3), 10 replications were created to compute S_s given in (22). Similarly, 10 independent replications for each combination of cases and scenario were created to compute S_m. The averages of the saving rates and their (frequentist) 95% confidence intervals were then calculated; these are reported in Figures 4, 5, and 6 for prior scenarios 1, 4, and 9. The results for the other prior scenarios are omitted because they are close to those of scenario 1 or 9.

Table 2. Cases of Parameters Characterizing the True Behavior of the Process Tried

Case	J	σ_υ	μ	σ_θ
1	20	4	4	2
2	20	4	4	4
3	20	4	0	2
4	20	4	0	4
5	20	2	4	2
6	20	2	4	4
7	20	2	0	2
8	20	2	0	4
9	5	4	4	2
10	5	4	4	4
11	5	4	0	2
12	5	4	0	4
13	5	2	4	2
14	5	2	4	4
15	5	2	0	2
16	5	2	0	4

Table 3. Scenarios for Prior Distributions Tried

Scenario	Mode(σ_υ^2)	Mode(σ_θ^2)	Variance of σ_υ^2 and σ_θ^2
1	σ_υ^2	σ_θ^2	$.25 \times \text{mode}^2$
2	σ_υ^2	σ_θ^2	$2 \times \text{mode}^2$
3	σ_υ^2	σ_θ^2	$100 \times \text{mode}^2$
4	$.5 \times \sigma_\upsilon^2$	$.125 \times \sigma_\theta^2$	$.25 \times \text{mode}^2$
5	$.5 \times \sigma_\upsilon^2$	$.125 \times \sigma_\theta^2$	$2 \times \text{mode}^2$
6	$.5 \times \sigma_\upsilon^2$	$.125 \times \sigma_\theta^2$	$100 \times \text{mode}^2$
7	$8 \times \sigma_\upsilon^2$	$2 \times \sigma_\theta^2$	$.25 \times \text{mode}^2$
8	$8 \times \sigma_\upsilon^2$	$2 \times \sigma_\theta^2$	$2 \times \text{mode}^2$
9	$8 \times \sigma_\upsilon^2$	$2 \times \sigma_\theta^2$	$100 \times \text{mode}^2$

As can be observed, for each case the intervals on the average relative savings S_s and S_m overlap for the different prior scenarios. Although the number of replications is small, there appears to be little difference in the relative savings obtained applying the two Bayesian adjustment rules (MCMC and SMC). However, the time required to perform SMC is very different from the time required by MCMC. This time comparison is discussed in the next section.

To provide more insight into the effect of the prior scenarios, Figure 7 reports the 95% confidence intervals on S_s for all of the cases and prior scenarios 1, 4, and 9.

Among the nine different scenarios for the prior distributions, scenario 4 is the only one that has a significant difference in average savings from the other scenarios. The prior settings on σ_θ^2 and σ_υ^2 may affect the convergence of the variance component estimators, but their effect on estimating the offsets is almost always negligible. Only in scenario 4 do we observe an effect of the prior distributions of the variances on our ability to estimate the offsets. This occurs because in this scenario the modes of the prior distributions are smaller than the true values and the variances of the priors are small. This makes the density have a thin right tail, so that the probability density around the true value is low. This requires a longer time for the mean of the posterior distribution to converge to its true value.

The prior distribution of the initial offset at the beginning of each lot depends on the prior distributions of μ and σ_θ^2, which are given by $\theta_{i0}|\mu, \sigma_\theta^2 \sim \text{N}(\mu, \sigma_\theta^2)$. A prior that indicates a very

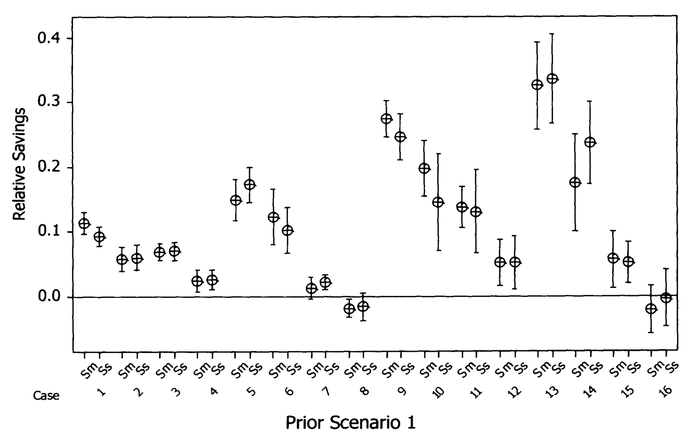

Figure 4. SMC Relative Savings Rates (S_s) and MCMC Savings Rates (S_m) for All Process Cases (Table 2) and Prior Scenario 1 (Table 3). Lines represent the 95% confidence intervals.

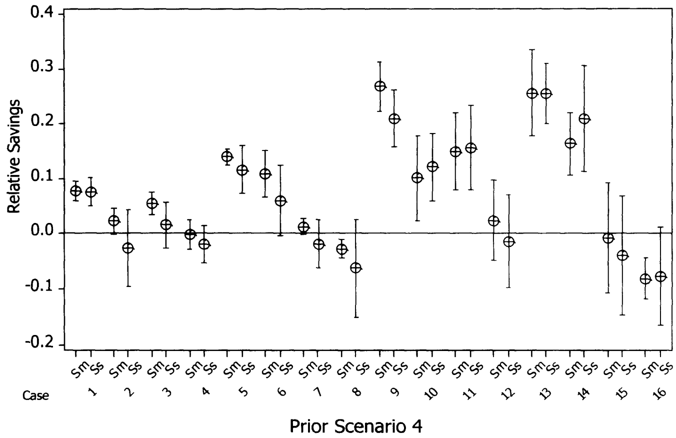

Figure 5. SMC Relative Savings Rates (S_s) and MCMC Savings Rates (S_m) for All Process Cases (Table 2) and Prior Scenario 4 (Table 3). Lines represent the 95% confidence intervals.

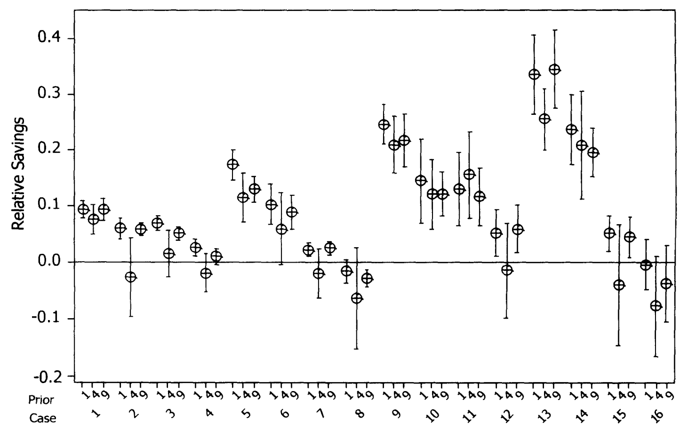

Figure 7. SMC Relative Savings Rates (S_s) for Select Process Cases (Table 2) and Priors (Table 3). Lines represent the 95% confidence intervals.

small σ_θ^2 gives extra confidence to the prior distribution of θ_{i0} (i.e., its variance is small). This obviously can be dangerous if the difference between the actual setup offset θ_{i0} and the estimate of μ is large. Creating definite rules on how to determine the prior distributions of the variances is difficult; one suggestion is to use a credibility interval as in the example of Section 4.1.

Several conclusions about the comparison between the three adjustment rules can be drawn based on these results:

- The SMC method showed advantages over the harmonic rule that are comparable to those obtained with the MCMC adjustment rule.
- In cases when an appropriate prior was chosen (i.e., not in scenario 4), the SMC method showed significant advantages against the harmonic rule, except for cases 8 and 16. In these two cases the means of the initial offsets are 0, and there are large lot-to-lot variances, σ_θ^2. These conditions are close to the assumptions on which the harmonic rule is based and for which it works best. On the other hand, extra errors are introduced in estimating μ by the SMC method when there is a large variance σ_θ^2, and this increases the total quadratic cost. Even in such disadvantageous cases, the performance of the SMC rule is still close to that of the harmonic rule (average savings are <0, but the confidence intervals include 0).
- The percentage savings when using the SMC rule decreases as the number of parts per lot, J, increases. This is because for the assumed process model, large adjustments are needed only in the first few parts in each lot. Once the process is centered, the remaining variability is uncontrollable. For a large lot, any of the adjustment rules that we discussed earlier provides adjustments that rapidly tend to 0. The total cost, C, includes the cost of all variability sources for all IJ parts. Thus if J is large, then in only a few of the J parts (the first ones in each lot, particularly the very first one) will there be an opportunity for a rule to outperform Grubbs's harmonic rule. Because we are measuring percentages of savings across all IJ observations, this percentage cannot be large for large J. One way to make the uncontrollable part of the total cost C evident is to partition it as

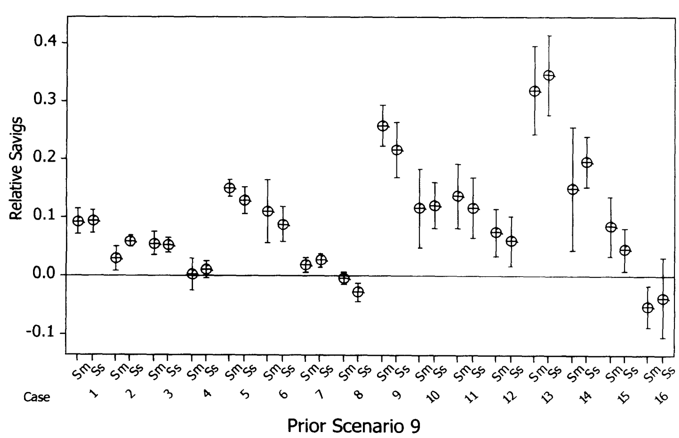

Figure 6. SMC Relative Savings Rates (S_s) and MCMC Savings Rates (S_m) for All Process Cases (Table 2) and Prior Scenario 9 (Table 3). Lines represent the 95% confidence intervals.

$$
\begin{aligned}
C &= \sum_{i=1}^{I}\sum_{j=1}^{J} Y_{ij}^2 \\
&= \sum_{i=1}^{I}\sum_{j=1}^{J}(\theta_i + U_{ij})^2 + \sum_{i=1}^{I}\sum_{j=1}^{J} 2(\theta_i + U_{ij})v_{ij} \\
&\quad + \sum_{i=1}^{I}\sum_{j=1}^{J} v_{ij}^2, \qquad (23)
\end{aligned}
$$

where the variance due to the last term clearly is not controllable. Once U_{ij} has converged to the true value of $-\theta_i$, which typically occurs after a few parts in each lot, the last term dominates the total cost C. (This explains why in cases 1–4, where σ_v^2 is at its high level and $J = 20$, the savings offered by the SMC rule compared with Grubbs's rule, are among the lowest, because we are comparing *total* cost.) To see more clearly that greatest savings occurs in the first few parts per lot, consider Figure 9(a) in Section 6, which shows the total quadratic cost per part obtained when processing lots under case 9, scenario 2.

Similar comparisons between the SMC rule and Grubbs's extended rule were performed, but the results are not reported here. The extended rule was competitive when μ was actually 0 (or known), in accordance with one of its underlying assumptions for which it is optimal. For cases when μ differs from 0 and was unknown, the SMC rule was significantly better (see Lian 2004 for details).

5.2 Computational Time Comparison of the Bayesian Approaches: SMC versus the MCMC Adjustment Rules

To outline the computational savings that can be achieved if the SMC algorithm is used instead of the MCMC approach, comparisons were made of the time required by the two algorithms to compute the adjustment setting after each new part has been observed. This is the only time that can not be masked in real time applications, because during this time interval the machine has to wait to receive as input the new setting of the control parameter that must be used to machine the next part.

Both the MCMC and the SMC algorithm were run on a Pentium 4. The SMC approach was implemented in Matlab (following the algorithm described in Sec. 4.2) and using 10,000 particles. The MCMC approach was performed using WinBUGS (Spiegelhalter, Thomas, and Best 2003) called by a script that uses a "burn-in" period equal to 1,000. Then 10,000 additional iterations of the chain were run to compute the required estimates from the steady-state distribution.

Figure 8 shows histograms of the time required at each adjustment step by the MCMC and the SMC algorithms for processing 20 lots of 20 parts (process case 1 in Table 2 and prior scenario 4 in Table 3). Similar behaviors were observed for other cases and prior scenarios and are not reported here.

As can be observed in Figure 8, the time required at each step by the MCMC ranges from 8 to 12 seconds. This time is required to scan through the dataset of observations already collected and perform the required estimates. In contrast, the computational time required by the SMC algorithm is almost always less than 1 second, whereas for just a few iterations the time required ranges from 9 to 12 seconds. These few iterations are those in which the rejuvenation step is required to avoid the degeneracy problem (Balakrishnan and Madigan 2004).

The comparison of computational time clearly shows the advantages related to the use of the proposed SMC algorithm. Referring again to the bending process illustrated in Section 4.1, the cycle time for an automated manufacturing cell is on the order of 6 seconds (Bollheimer 2004). It is clear that in this type of application, the MCMC adjustment rule is impractical, whereas the SMC approach represents a viable adjustment procedure.

5.3 Effect of Changing the Type of Prior Distribution in the SMC Procedure

In the previous section it was shown that the first SMC rule and the MCMC adjustment rule give essentially equivalent performance in terms of cost. To study the sensitivity of the SMC approach to the *type* of prior, additional simulation results, not reported here, were conducted for uniform $(0, b)$ priors on the variances. In summary, although the performance of the SMC method is generally not overly sensitive to the choice of prior distribution, we suggest using LN priors. These have a decaying right tail that provides some extra "robustness" not provided by finite-range priors such as the uniform, are easy to assess and calibrate from expert opinion, and, compared with the inverse-gamma priors frequently used for variances, are easier to make "noninformative" if that is what is desired.

6. AVOIDING INFLATION IN VARIANCE DUE TO UNNECESSARY ADJUSTMENTS: A SECOND (MODIFIED) SMC RULE

Two types of process adjustments are made by the SMC method discussed in previous sections. The first type of adjustment is made according to the posterior distribution of θ_{ij} immediately after a new part is observed. These adjustments are similar to those suggested by the harmonic rule. The second type of adjustment is made at the beginning of a lot based on the estimate of μ before any part in that lot is observed. No analogous adjustment is provided by Grubbs's harmonic rule, because the first observation in each lot is uncontrolled.

In the experiments of the previous section, cases 8 and 16 (see Table 2) have $\mu = 0$. In these two cases, extra errors can be introduced by the second type of adjustment (i.e., the quadratic

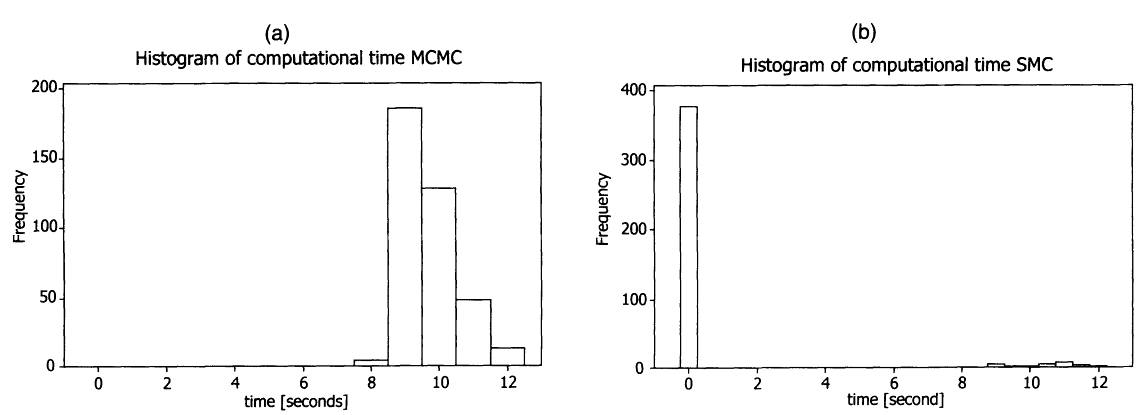

Figure 8. Histograms of the Computational Time per Adjustment (in seconds) Incurred by the MCMC Method (a) and the SMC Method (b), Process Case 1 (Table 2) and Prior Scenario 4 (Table 3).

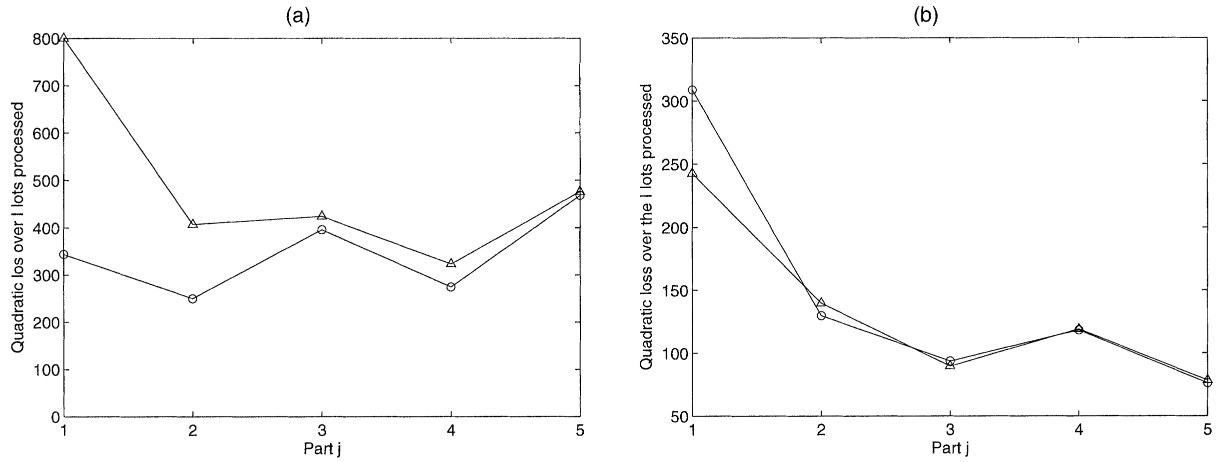

Figure 9. Total Quadratic Loss per Part for a Single Simulation of Two Cases: Case 9 (a) and Case 16 (b), Under Scenario 2 (—△— Grubbs; —○— SMC). Losses shown are totals per part j added across all 20 lots.

cost will be inflated because there is no need to adjust if the lot means are really 0). This problem is obvious from looking at cases 8 and 16 in Figure 7.

The poor performance of the first SMC method also can be observed by considering a case where the method works well compared with a case where it does not work so well due to unnecessary adjustments. Figure 9(a) shows the total quadratic losses per part for case 9 (where $\mu = 4$), and Figure 9(b) shows the total quadratic losses per part for case 16 (where $\mu = 0$), assuming the prior parameters as in scenario 2. In the case 9, the SMC procedure gives considerable savings over Grubbs's harmonic rule, particularly in the first couple of parts, whereas for case 16, where the mean is 0, the SMC actually is more costly than Grubbs's rule.

In a first attempt at reducing this "inflation in variance" problem, one could think of a rule that would never adjust the first part in each lot. The effect of making this modification is illustrated in Figure 10, where the original SMC method is called method 1 and the SMC method omitting all first adjustments in each lot is called method 2. (We explain the third method shortly.) The average savings and confidence intervals were obtained based on 10 replications. The advantages of this modification are significant in all cases where $\mu = 0$, because the extra errors made by the second type of adjustments are eliminated in these cases. But when $\mu \neq 0$, these first adjustments would have been beneficial, so eliminating them in cases when $\mu = 4$ makes the saving rates actually drop significantly.

To eliminate the disadvantages of the second type of adjustments when they are unnecessary ($\mu = 0$) while keeping their advantages when they are beneficial ($\mu \neq 0$), a *conditional first-adjustment SMC method* is proposed. Under this method, at the beginning of lot i ($i > 2$), the first adjustment is executed only when the $100(1 - \alpha)\%$ credibility region of the posterior distribution of μ excludes 0; otherwise the first adjustment is omitted. Note that when $\alpha = 1$, this method works identically to the first SMC method of Section 4.2. When $\alpha = 0$, all of the first adjustments are omitted, as in method 2 in Figure 10.

We recommend selecting the value of α by simulation for given prior distributions. For instance, the graph in Figure 11 was obtained by simulating the adjusted process for a set of real process parameters, in this case those in Table 2 and, for

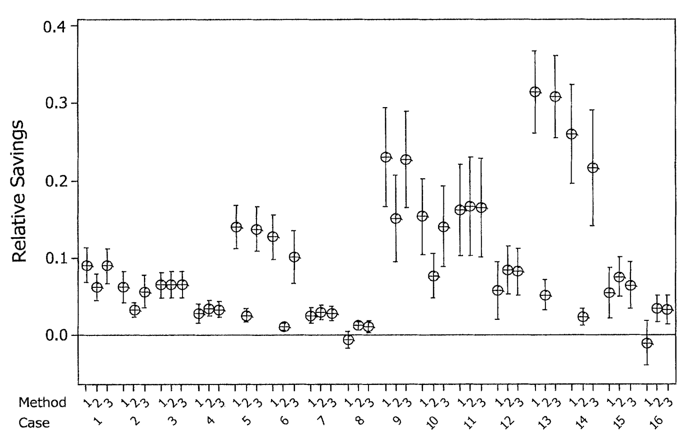

Figure 10. Average Savings Rates Obtained by the First SMC Method (method 1), the SMC Method When Always Omitting the First Adjustment (method 2), and the Conditional First-Adjustment SMC Method (method 3) Compared With the Harmonic Rule Using LN Priors. Lines indicate the 95% confidence intervals.

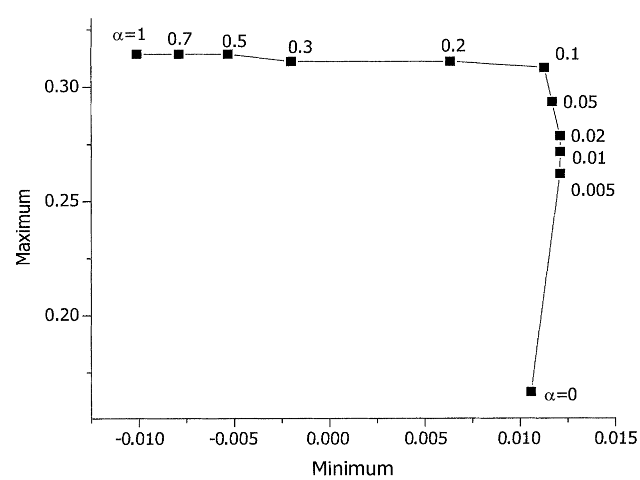

Figure 11. Maximum and Minimum Average Savings When Using the Modified SMC Rule With Different α Values in the 16 Cases of True Parameters Assuming the Priors as in Scenario 2.

illustration, scenario 2 from Table 3. For each case of true parameters, several values of α were used, and simulations were run. The largest and smallest savings obtained for each value of α were saved, giving the points on the graph. From the graph in Figure 11, we would choose $\alpha = .1$, because for that α, the maximum saving rates are very high (these are cases when the method performed best) while the minimum saving rate (cases when the method performed worst) is also large. This choice of α is therefore "robust" with respect to variations in the true parameters of the process, which are unknown to us and are treated as "noise factors" in the simulation experiment. A MATLAB computer program that simulates the process and computes the graph for a given set of true parameters and given LN priors is available from the same website as before. From our numerical experiments, the best value of α is usually within the (.05, .3) interval.

Figure 10 shows the performance of the conditional first-adjustment method with $\alpha = .1$ (method 3) compared with the harmonic rule. Compared with the original SMC method (method 1), the improved conditional first-adjustment method is more robust to variations in the different process conditions. This method retains the substantial savings provided by the original SMC method in the cases where $\mu = 4$ (e.g., process cases 13 and 14), while efficiently eliminating the unnecessary second type of adjustments in cases where μ is actually 0 (e.g., process cases 15 and 16).

We point out that performing a similar simulation analysis to implement the conditional first adjustment using the MCMC rule will require extremely time-consuming simulations, which the SMC rule avoids.

7. CONCLUSIONS

An adjustment rule based on sequential Monte Carlo techniques for the setup adjustment problem was presented. The method was contrasted against Grubbs's harmonic rule and a MCMC method. Bayesian methods starting with vague priors have similar behavior in the first few lots as the harmonic rule, but they keep updating the posterior distributions of the parameters as more observations are available, from which increasingly better adjustments are made. The SMC method is considerably more time-efficient than the MCMC method, with computational savings growing as the size of the dataset grows. At the same time, it provides equivalent performance to the MCMC rule.

The performance of the first SMC method was analyzed with respect to changes in the prior parameters and changes in the type of prior. It was shown that unless the prior reflects strongly inaccurate information with respect to the true parameters, the performance of the first SMC method is superior to Grubbs's harmonic rule, because the method is able to "anticipate" the first observation in each lot. This analysis also provided justification for using the proposed lognormal priors for the variances, which are easy to assess in practice and were shown to provide some extra protection that other priors lack.

A second, modified SMC rule was also presented to avoid variance inflation due to unnecessary adjustments, a situation that occurs if the lot-to-lot mean μ is actually 0. In the modified rule, redundant adjustments are omitted when one can conclude that $\mu \simeq 0$ based on the posterior distribution of μ and a credibility parameter α, which is user-selected. A Taguchi-like simulation approach was presented to help the user select the value of α.

Computer programs that implement the SMC method and that help select the value of α in the modified SMC method can be downloaded from *http://www2.ie.psu.edu/Castillo/research/EngineeringStatistics/*.

[Received November 2004. Revised December 2005.]

REFERENCES

Akcalt, E., Nemoto, K., and Uzsoy, R. (2001), "Cycle Time Improvements for Photolithography Processes in Semiconductor Manufacturing," *IEEE Transactions on Semiconductor Manufacturing*, 14, 48–56.

Aström, K. J., and Wittenmark, B. (1994), *Adaptive Control* (2nd ed.), Englewood Cliffs, NJ: Prentice-Hall.

Balakrishnan, S., and Madigan, D. (2004), "A One-Pass Sequential Monte Carlo Method for Bayesian Analysis of Massive Datasets, technical paper, Rutgers University, Dept. of Statistics; available at *http://www.stat.rutgers.edu/~madigan/PAPERS/*.

Bjorke, O. (1989), *Computer-Aided Tolerancing* (2nd. ed.), New York: ASME Press.

Bollheimer, M. (2004), "Bending and Handling Tubes: Tailoring Equipment to an Application," *TPJ—The Tube & Pipe Journal*, July, available at *http://www.thefabricator.com/Articles/Tube_And_Pipe_Article.cfm?ID=927*.

Colosimo, B. M., Pan, R., and Del Castillo, E. (2004), "A Sequential Markov Chain Monte Carlo Approach to Setup Adjustment of a Process Over a Set of Lots," *Journal of Applied Statistics*, 31, 499–520.

Del Castillo, E. (2002), *Statistical Process Adjustment for Quality Control*, New York: Wiley.

Del Castillo, E., Pan, R., and Colosimo, B. M. (2003a), "An Unifying View of Some Process Adjustment Methods," *Journal of Quality Technology*, 35, 286–293.

——— (2003b), "Small-Sample Comparison of Some Process Adjustment Methods," *Communications in Statistics, Part B—Simulation and Computation*, 32, 923–942.

Doucet, A., de Freitas, N., and Gordon, N. (eds.) (2001), *Sequential Monte Carlo Methods in Practice*, New York: Springer-Verlag.

Elkins, K. L., and Sturges, R. H. (2001), "Design of a Sensor for On-Line Measurement of Loaded Bend Angle for Pressbrake Control," *Robotics and Computer-Integrated Manufacturing*, 17, 329–340.

Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. (eds.) (1996), *Markov Chain Monte Carlo in Practice*, New York: Chapman & Hall.

Grubbs, F. E. (1954), "An Optimum Procedure for Setting Machines or Adjusting Processes," *Industrial Quality Control*, 11, 6–9; *Journal of Quality Technology*, 15 (1983), 186–189.

Kong, A., Liu, J., and Wong, W. (1994), "Sequential Imputation and Bayesian Missing Data Problems," *Journal of the American Statistical Association*, 89, 278–288.

Koons, G. F., and Luner, J. J. (1991), "SPC in Low-Volume Manufacturing: A Case Study," *Journal of Quality Technology*, 23, 287–295.

Lian, Z. (2004), "Bayesian Statistical Process Adjustment for Unknown Parameter Systems," unpublished doctoral thesis, Pennsylvania State University, Dept. of Industrial & Manufacturing Engineering.

Lian, Z., Colosimo, B. M., and Del Castillo, E. (2006), "Setup Error Adjustment: Sensitivity Analysis and a New MCMC Control Rule," *Quality and Reliability Engineering International*, 22, 403–418.

Moyne, J., Del Castillo, E., and Hurwitz, A. (eds.) (2000), *Run-to-Run Process Control for Semiconductor Manufacturing*, Boca Raton, FL: CRC Press.

Ridgeway, G., and Madigan, D. (2002), "A Sequential Monte Carlo Method for Bayesian Analysis of Massive Datasets," *Journal of Knowledge Discovery and Data Mining*, 7, 301–319.

Siu, T.-S. (2004), "Cycle-to-Cycle Feedback Control of Manufacturing Processes," unpublished master thesis, Massachusetts Institute of Technology, Dept. of Mechanical Engineering; available from MIT OpenCourseware at *http://ocw.mit.edu/OcwWeb/Mechanical-Engineering/2-830JSpring2004/Readings/index.htm*.

Smith, A. F. M., and Gelfand, A. E. (1992), "Bayesian Statistics Without Tears: A Sampling-Resampling Perspective," *The American Statistician*, 46, 84–88.

Spiegelhalter, D. J., Abrams, K. R., and Myles, J. P. (2004), *Bayesian Approaches to Clinical Trials and Health Care Evaluation*, New York: Wiley.

Spiegelhalter, D. J., Thomas, A., and Best, N. (2003), Win-BUGS Version 1.4 User Manual; available at *http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml*.

Trietsch, D. (1998), "The Harmonic Rule for Process Setup Adjustment With Quadratic Loss," *Journal of Quality Technology*, 30, 75–84.

Woodall, W. H., and Thomas, E. V. (1995), "Statistical Process Control With Several Components of Common Cause Variability," *IIE Transactions*, 27, 757–764.