

A Sequential Markov Chain Monte Carlo Approach to Setup Adjustment of a Process Over a Set of Lots

B.M. Colosimo¹, R. Pan², E. del Castillo^{3*}

¹ Dipartimento di Meccanica - Politecnico di Milano
Piazza Leonardo da Vinci 32, 20133 Milano, Italy

²Department of Mechanical & Industrial Engineering
The University of Texas at El Paso, El Paso, TX 79968-0521, USA

³Department of Industrial & Manufacturing Engineering
The Pennsylvania State University, University Park, PA 16802, USA

Revised August 2002

Abstract

The existing procedures for adjusting setup errors in a production process over a set of lots are based on the assumption of known process parameters. In practice, these parameters are usually unknown, especially in short-run production. Due to this lack of knowledge, simpler procedures such as Grubbs' rules (1954, 1983) and discrete integral controllers (also called EWMA controllers) aimed at adjusting the initial offset in each single lot, are typically used. This paper presents an approach for adjusting the initial machine offset over a set of lots when the process parameters are unknown and are iteratively estimated using Markov Chain Monte Carlo (MCMC). As each observation becomes available, a Gibbs Sampler is run to estimate the parameters of a hierarchical normal means model given the observations up to that point in time. The current lot mean estimate is then used for adjustment. If used over a series of lots, the proposed method allows to eventually start adjusting the offset before producing the first part in each lot. The method is illustrated with application to two examples reported in the literature. It is shown how the proposed MCMC adjusting procedure can outperform existing rules.

Key Words: Process Adjustment, Gibbs Sampling, Bayesian hierarchical models, Random effects model, Normal means model, Process control.

*corresponding author

1 Introduction

Modern manufacturing practice has as one of its tenets the reduction of the size of the lots or batches of product (i.e., a tendency for smaller production runs) in order to improve the efficiency of operations and eliminate waste. When parts are processed in small batches, two sources of variability become relevant: the "within-batch" variability and the "between-batches" one. Between-batches variability implies that the quality characteristics produced in each batch or lot can be characterized by different means which represent offsets when measured with respect to a target value. This can be due to "day-to-day changes, shift-to-shift changes, unknown shifts in level of a machine for a given setting due to repair work, etc." (Grubbs 1954, 1983), or more generally, due to an incorrect setup operation. If the offset characterizing a lot were observable without error, it could be removed at once by adjusting the process mean after the first part in the lot has been observed. Unfortunately, within-batch variability does not allow to observe directly the magnitude of the offset, and the adjustment procedure should estimate the offset to effectively reduce scrap and rework.

F. E. Grubbs proposed a procedure that will be called in the following Grubbs' "extended rule" (following Trietsch, 1998), aimed at adjusting initial offsets over a set of lots. In his model, Grubbs assumes the initial offset in each lot can be considered as an occurrence of a random variable with known mean $\mu = 0$ and known variance σ_θ^2 (the between-batch or between-lot variance). Furthermore, the within-batch or within-lot variance σ_v^2 (due to process and measurement errors) is also assumed to be known before starting adjusting. No further offsets or "shifts" are assumed to occur within a lot. If all these assumptions hold, Grubbs' extended rule is optimal if the off-target costs are quadratic and no cost is incurred when performing the adjustments. If off-target costs are based on a 0-1 loss criterion, Sullo and Vandeven (1999) derived the optimal adjustment strategy considering again process parameters are known.

Del Castillo, Pan and Colosimo (2002a) showed how Grubbs' extended rule has a Bayesian interpretation based on a Kalman filter where an *a priori* knowledge of parameters characterizing both batch-to-batch and within-batch distributions is required. When no information on the offset is available, Grubbs' suggested using a different rule, that we will call Grubbs' "harmonic rule" following again Trietsch (1998). This can be derived as a special case of the extended rule when the prior distribution on the initial offset has infinite variance, i.e., no *a priori* knowledge is available (Del Castillo *et al.*, 2002a). Although derived to compensate for an offset while processing a single lot, this rule and the simple integral controller (or EWMA controller) can perform better than the extended rule (Del Castillo *et al.*, 2002b), since optimality of the later rule is no longer guaranteed if the prior distribution of the initial offset is not exactly equal to the true setup error distribution.

The present paper proposes a different approach for process adjustments over a set of batches, when no previous knowledge on the distribution of the setup offsets and on the process variability is available. Hence, the on-line adjustment procedure proposed herein can be useful when starting processing lots of a new product or with a new installed process. The approach is based on hierarchical Bayesian models and uses Markov Chain Monte Carlo (MCMC) to derive estimates used to compute the adjustments at each observation. Therefore, a *sequence* of MCMC runs is conducted, one per adjustment, and the results of each are then used to adjust the process.

It will be shown, based on two examples reported in the literature that the proposed adjustment approach can perform considerably better than competitor rules, in particular, Grubbs' harmonic rule and the EWMA controller, when no a priori knowledge is available.

The remainder of this paper is organized as follows. The problem of adjusting setup errors over a set of lots and its relation with performing inference in a one-way random effects model is presented in Section 2. The basic motivation for using a Bayesian hierarchical model and details on how to compute the adjustments at each observation are presented in Sections 3 and 4, respectively. Implementation of the Markov Chain Monte Carlo algorithm used to compute estimates required (Gibbs sampling) and related convergence issues as applied to the setup adjustment problem are discussed in Section 5. Section 6 contains some performance comparisons between the proposed approach vs. other competing adjustment rules (Grubbs' rule and two EWMA controllers) on the basis of two examples presented in Box and Tiao (1973). Finally, the last section gives conclusions and discusses some directions for further research.

2 Setup offset adjustments over a set of batches

Assume parts are processed in batches or lots and that the quality characteristic of parts in each lot is characterized by a different initial offset or deviation from nominal. Denote by $i = 1, \dots, I$ the lots processed, by $j = 1, \dots, J$ the parts in the lot and by Y_{ij} the quality characteristic observed at the j^{th} part of the i^{th} lot (or at time (i, j) , for short). This quality characteristic is measured with reference to a nominal value T which is assumed, without loss of generality, to be equal to zero. Denote by θ_i the initial (unknown) offset before starting processing the i^{th} lot. It is assumed that the initial offset in each lot is generated as a random realization from a normal distribution with mean μ and variance σ_θ^2 , both unknown, i.e.,

$$\theta_i | \mu, \sigma_\theta^2 \sim N(\mu, \sigma_\theta^2) \quad . \quad (1)$$

Note that contrary to Grubbs' extended rule, parameters are assumed unknown and the mean μ can be different from zero. This can occur due to some systematic errors that affect all the setups performed on the machine.

Before starting processing the first item of the i^{th} lot, assume the controllable factor has been set to the value U_{i0} . This value will be referred in what follows as the "set point" of the machine. The deviation from target for the first part processed is thus characterized by a mean θ_{i1} equal to the initial (unknown) offset θ_i plus the set point decided:

$$\theta_{i1} = \theta_i + U_{i0} \quad .$$

As in Grubbs' model (Grubbs, 1954, 1983), process variability plus measurement errors are modelled together and assumed to be normally distributed with zero mean and variance σ_v^2 . It is further assumed that σ_v^2 is unknown, relaxing the corresponding assumption in Grubbs' extended rule. The quality characteristic observed at the first part in the i^{th} lot is thus given by:

$$Y_{i1} = \theta_{i1} + v_{i1} \quad ,$$

where $v_{ij} \stackrel{iid}{\sim} N(0, \sigma_v^2)$ represents random error. Once the quality characteristic has been observed, the adjustment $\nabla U_{i1} = U_{i1} - U_{i0}$ has to be determined and implemented in order to "cancel" the offset θ_{i1} , i.e.,

$$\nabla U_{i1} = -\widehat{\theta}_{i1} \quad ,$$

where $\widehat{\theta}_{i1}$ is an estimate of the offset of the current lot (lot i) after obtaining the measurement for part 1. Due to this adjustment, the mean of the quality characteristic for the next processed part changes to:

$$\theta_{i2} = \theta_{i1} + \nabla U_{i1} \quad ,$$

and the quality characteristic observed for the second part in the lot is given by:

$$Y_{i2} = \theta_{i2} + v_{i2} \quad .$$

Continuing this way, the quality characteristic of the j^{th} part processed in the i^{th} lot is given by:

$$Y_{ij} = \theta_{ij} + v_{ij} \quad , \tag{2}$$

where:

$$\theta_{ij} = \theta_{ij-1} + \nabla U_{ij-1} \quad , \tag{3}$$

and $\theta_{i0} = \theta_i$. Hence, the next adjustment is selected as:

$$\nabla U_{ij} = -\widehat{\theta}_{ij} \quad . \tag{4}$$

Therefore, adjustments ∇U_{ij} can be determined by estimating the relevant unknown parameter, in this case, the mean of the quality characteristic in the current lot, based on data available up to time (i, j) .

An important relationship with the problem of estimating the means in a one-way random effects model can be seen after some manipulations in expressions (2) to (4). Solving recursively equation (3) and considering that $\nabla U_{ij} = U_{ij} - U_{ij-1}$, the mean of the quality characteristic can be rewritten as:

$$\theta_{ij} = \theta_i + U_{ij-1} \quad . \tag{5}$$

Therefore, the quality characteristic of the j^{th} part in the i^{th} lot can be rewritten as:

$$Y_{ij} = \theta_i + U_{ij-1} + v_{ij} \quad . \tag{6}$$

Considering that at the time Y_{ij} is observed, U_{ij-1} is known, a new variable X_{ij} can be defined as:

$$X_{ij} = Y_{ij} - U_{ij-1} = \theta_i + v_{ij} \quad . \tag{7}$$

Merging the last expression with equation (1) permits to derive an analogy with a one-way random effects model, i.e.,:

$$\begin{aligned} X_{ij} | \theta_i, \sigma_v^2 &\sim N(\theta_i, \sigma_v^2) \quad , \\ \theta_i | \mu, \sigma_\theta^2 &\sim N(\mu, \sigma_\theta^2) \quad . \end{aligned} \tag{8}$$

Finally, expression (4), which defines the adjustment rule or feedback “controller”, can be rewritten from equation (5), thus obtaining:

$$\begin{aligned}\nabla U_{ij} &= -\widehat{\theta}_{ij} = -\widehat{\theta}_i | \mathbf{x}^{ij} - U_{ij-1} \quad \text{or, alternatively,} \\ U_{ij} &= -\widehat{\theta}_i | \mathbf{x}^{ij} \quad ,\end{aligned}\tag{9}$$

where $\mathbf{x}^{ij} = \{x_{11}, x_{12}, \dots, x_{1J}, \dots, x_{i1}, \dots, x_{ij}\}$ represents all (transformed) data observed at the time the estimate of θ_i is computed, and $\widehat{\theta}_i | \mathbf{x}^{ij}$ represents the estimate of the i^{th} mean in the random effects model, given all data available up to time (i, j) . We point out that selecting an adjustment at each step reduces to performing inference about means in an unbalanced one-way random effects model as long as processing of the current lot (i.e., the last lot) has not been completed (so $j < J$). We now consider the Bayesian estimation of the parameters of such model.

3 Why a Bayesian model?

Box and Tiao (1973), Gelman *et al.* (1996) and Carlin and Louis (1996) discussed parallels in random effects model estimation using sampling-based and Bayesian approaches. In particular, consider the estimation of the i^{th} mean, θ_i , in the model given by equation (8). Referring for simplicity to the balanced case, traditional inference based on sampling theory can be performed by first performing an Analysis of Variance (ANOVA). Standard sampling theory results suggest two candidate estimators for the lot means. The first one is the average of data collected in the i^{th} lot:

$$\widehat{\theta}_i^1 = \bar{x}_i = \frac{\sum_j x_{ij}}{J} \quad ,\tag{10}$$

where, following common notation, a dot in the subscript means average is computed over that subscript. The second estimator considers complete pooling and is given by:

$$\widehat{\theta}_i^2 = \bar{x}_{..} = \frac{\sum_i \bar{x}_i}{I} \quad .\tag{11}$$

The selection between these two estimators depends on the outcome of an F test performed to evaluate if means characterizing lots can be considered different. If the test suggests that the ratio of between to within mean squares (MS_b/MS_w) is not statistically significant, the pooled estimator in (11) should be used, while separate estimates (given by equation 10) have to be used otherwise. Depending on the first type error selected for the F test, usual sampling-based approaches imply a sudden shift from none to complete pooling of the independent estimates. As outlined in Gelman *et al.* (1996) the Bayesian approach permits to naturally overcome this choice, since the Bayesian estimator of θ_i , i.e., the posterior mean when adopting a normal prior density, is a weighted combination of the complete and separate estimates and the extreme cases (none or complete pooling) turn out to be special cases of the Bayesian solution. Box and Tiao (1973, pp. 388) pointed out how this has a parallel with certain results in sampling theory, specifically with the work of James and Stein.

The Bayesian approach has some additional advantages over the sampling approach related to the estimation of variance components and to flexibility in departures from assumptions. Variance components estimates can be useful in adjusting for initial offsets. After some lots have been processed, estimates of between-lots and within-lot variances, i.e., σ_θ^2 and σ_v^2 , can be obtained and Grubbs' extended rule can be adopted thereafter. Considering the estimation of the variance components, traditional sampling theory suggests using:

$$\hat{\sigma}_v^2 = MS_w \quad \text{and} \quad (12)$$

$$\hat{\sigma}_\theta^2 = \frac{MS_b - MS_w}{I} \quad (13)$$

A first well-known problem concerns $\hat{\sigma}_\theta^2$, which turns out to be negative if $MS_b < MS_w$. Furthermore, the estimate $\hat{\sigma}_\theta^2$ given by (13) has a complex distribution that induces problems in deriving a confidence interval for this quantity. In contrast, the Bayesian approach permits to tackle easily the lack of normality and/or independence and possible heterogeneity of variances (Box and Tiao, 1973).

Due to the considerations outlined, a Bayesian approach has been adopted for determining the adjustments for an on-line setup adjustment in a lot by lot production process.

4 Bayesian online adjustment of initial offsets

From a Bayesian perspective, the one-way random effects model is a special case of a *hierarchical model*, used in describing multiparameter problems in which parameters are related based on some structure that depends on the specific problem addressed.

The Bayesian model used for the setup adjustment problem can be described by considering the case when the j^{th} part in the i^{th} lot has been just processed and we have to decide the adjustment for the next part in lot i . In this case, the first stage of the hierarchy models the distribution of data conditionally on unknown parameters and is given by the first equation in (8). In this equation, the (transformed) data distribution is given conditionally on two unknown parameters: the current mean θ_i and the variance within lots σ_v^2 . The second stage in the hierarchy specifies the distribution of these unknown parameters. The within lot variance σ_v^2 , which represents process variability plus measurement errors, does not depend on any other parameter. Therefore the probabilistic specification of this parameter at the second stage of the hierarchy models the previous or subjective knowledge on it, i.e. its prior distribution. Adopting conjugacy at each step of the hierarchical model (a common choice for the random effects model, Gelfand *et al.*, 1990; Gelfand and Smith, 1990), the prior distribution for σ_v^2 is given by:

$$\sigma_v^2 | a_2, b_2 \sim IG(a_2, b_2) , \quad (14)$$

where IG represents an Inverse-Gamma distribution and a_2, b_2 are assumed known and have been chosen both equal to 0.001 to model "vague" prior information (Spiegelhalter *et al.*, 1994).

The second unknown parameter modelled at the second stage in the hierarchy is the initial offset θ_i , given by equation (1). As it can be observed, its distribution is given conditionally

on other two parameters, μ and σ_θ^2 . These are *hyperparameters* and are modelled at the third stage in the hierarchy, where the prior distributions on them is reported. Adopting once again conjugacy, priors on these hyperparameters are given by:

$$\mu|\mu_0, \sigma_0^2 \sim N(\mu_0, \sigma_0^2) \quad , \quad (15)$$

$$\sigma_\theta^2|a_1, b_1 \sim IG(a_1, b_1) \quad , \quad (16)$$

where IG is the Inverse-Gamma distribution and $\mu_0, \sigma_0^2, a_1, b_1$ are assumed known. In particular, they were selected according to values suggested in the literature (Spiegelhalter *et al.*, 1994) to model "vague" prior distributions, i.e., $\mu_0 = 0, \sigma_0^2 = 1.0E + 10, a_1 = b_1 = 0.001$.

The random effects model, hereafter called model M_1 , can be described by three stages modelling respectively data, parameters and hyperparameters. Using this approach, the model M_1 can be graphically represented as in Figure 1 and is given by:

Three-stage hierarchical model M_1

- *first stage:* data X_{kp} ($k = 1, \dots, i$ and $p = 1, \dots, J$) with distribution given by the first equation in (8);
- *second stage:* parameters θ_k ($k = 1, \dots, i$) and σ_v^2 with distributions given respectively by equations (1) and (14);
- *third stage:* hyperparameters μ and σ_θ^2 with distributions given respectively by equations (15) and (16).

We point out the similarity between the assumed model and a "normal means" model (Gelfand *et al.*, 1990). Once data from all lots are available, the normal means model allows for different lot sizes and different within-lot variances whereas we assume a constant lot size and the same within-lot variance distribution for all lots. Extension to such case is straightforward but not addressed in this paper.

As already mentioned, the adjustment ∇U_{ij} can be computed using (9) once the deviation from target Y_{ij} has been measured, i.e., it is based on estimating the initial offset θ_i . Using the three-stage hierarchical model, the estimator of θ_i is given by the expected value of the posterior distribution and can be computed if at least two parts from different lots have been already processed (otherwise the variance between lots σ_θ^2 can not be estimated). Therefore adjustments ∇U_{ij} for lots starting from the second lot are given by:

$$U_{ij} = -E(\theta_i | \mathbf{x}^{ij}, M_1) \quad i = 2, \dots, I \quad \text{and} \quad j = 1, \dots, J, \quad (17)$$

where conditioning on M_1 means the estimation is computed using the random effects model M_1 .

Notice how when selecting the adjustment for the $j+1$ part in the i^{th} lot, we have unequal lot sizes since for all the lots already processed ($k = 1, \dots, i-1$) J entries are available (in accordance to our constant lot size assumption) while for the current lot ($k = i$) just $j \leq J$ data have been already collected. However, lot sizes can be treated as equal by considering future data (i.e., $p = j+1, \dots, J$) of the current lot i as if they were missing data (Spiegelhalter *et al.*, 1994).

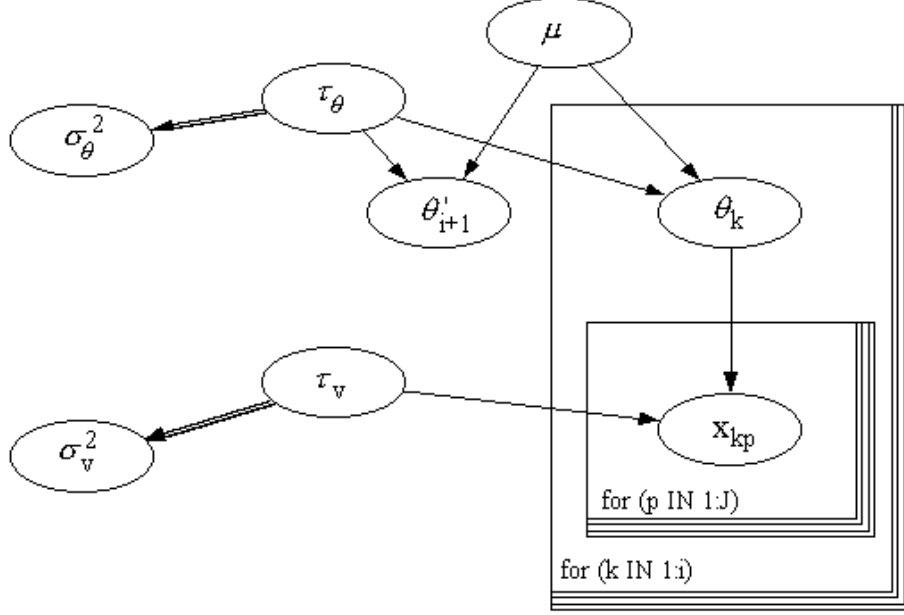


Figure 1: Graphical representation of the three-stage model M_1 , where $\tau_v = 1/\sigma_v^2$ and $\tau_\theta = 1/\sigma_\theta^2$ are the *precisions* of the two normal distributions in the model.

A main advantage of applying Bayesian hierarchical modelling is the possibility to select the initial set-point $U_{i+1,0}$ that has to be set on the machine *before* processing the first part in the next lot (lot $i + 1$). In contrast, Grubbs' harmonic rule requires to collect at least one (uncontrolled) observation in each lot in order to set the controllable variable starting with the second part in each lot. Using the hierarchical model described, the predictive distribution for the offset in the next lot can be derived using information collected in all the previous lots. Therefore, we can set:

$$U_{i+1,0} = -E(\theta'_{i+1} | \mathbf{x}^{iJ}, M_1) \quad \text{for } i = 3, \dots, I \quad (18)$$

where $\theta'_{i+1} | \mathbf{x}^{iJ}, M_1$ is the posterior predictive distribution (Gelman *et al.*, 1996) of a future offset given all the data collected from previously produced lots. Similarly as before, prediction can start from the third lot since at least two lots have to be processed to start estimating the variance between lots.

To start adjusting in the first lot, a reduced two-stage model M_2 was used. When only one lot is being processed, the initial offset θ_1 can be considered not as a realization from a random variable but simply as an unknown parameter. Hence, the three-stage model M_1 “collapses” into a two-stage model, describing the data distribution at the first stage and priors on unknown parameters at the second one. Therefore, the first stage in model M_2 is given by the first equation in (8), just as in model M_1 . The second stage reports priors on the two unknown parameters θ_1 and σ_v^2 . The prior on the last parameter is given again by equation (14), and the prior on θ_1 is given by:

$$\theta_1 | \mu, \sigma_\theta^2 \sim N(\mu, \sigma_\theta^2) \quad , \quad (19)$$

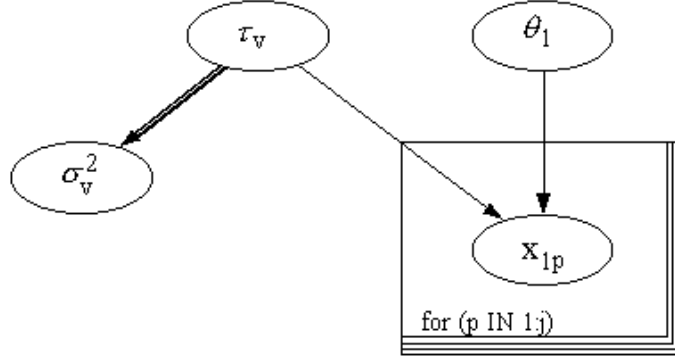


Figure 2: Graphical representation of the two-stage model M_2 where $\tau_v = 1/\sigma_v^2$ is the *precision* of the normal distribution in the model.

| Batch | U_{i0} | U_{ij} |
|---------|---------------|---|
| 1 | 0 | equation (21) for $j = 1$ and equation (20) for $j = 2, \dots, J$ |
| 2 | 0 | equation (9) for $j = 1, \dots, J$ |
| 3 ... I | equation (18) | equation (9) for $j = 1, \dots, J$ |

Table 1: Bayesian adjustment of initial offset over a set of lots.

where now μ and σ_θ^2 are supposed known and chosen to model vague prior information on θ_1 , i.e., $\mu = 0$ and $\sigma_\theta^2 = 1.0E + 10$. Therefore, the model for selecting adjustments in the first lot is given by:

Two-stage hierarchical model M_2

- *first stage*: data X_{1p} ($p = 1, \dots, j$) distribution given by the first equation in (8);
- *second stage*: parameters θ_1 and σ_v^2 with distributions given respectively by equations (19) and (14).

For the second model the adjustments ∇U_{1j} are obtained by computing the expected value of the posterior distribution of θ_1 , given data already observed and the model M_2 , i.e.:

$$U_{1j} = -E(\theta_1 | \mathbf{x}^{1j}, M_2), \quad j = 2, 3, \dots, J; \quad (20)$$

where j starts from 2 since at least two parts have to be processed to estimate the within-lot variance σ_v^2 . To start adjusting just after the first part in the first lot is processed and measured, the trivial estimator:

$$U_{11} = -x_{11} \quad (21)$$

can be used which coincides with Grubbs' harmonic rule. In summary, the proposed adjustment procedure consists of computing the adjustments as reported in Table 1.

5 Gibbs Sampling implementation and convergence issues

Although Bayesian models permit to overcome difficulties implied by sampling theory and to derive easy extensions to more complex problems, it is well-known that their main drawback is related to computational difficulties in the calculation of marginal posterior densities required for Bayesian inference. Problems exist related to the exact computation of functions of the marginal posterior densities since they require solving complicated high-dimensional integrals. In the last two decades, these difficulties have been greatly reduced due to the development of simulation-based approaches (Gelfand and Smith, 1990), which permit to numerically compute marginal posteriors. In particular, Markov Chain Monte Carlo methods are widely spread and consist of performing Monte Carlo integration using Markov Chains. One of the better known MCMC algorithms used to construct this chain is the Gibbs sampler. The basic ideas of the Gibbs sampler applied to the adjustment problem at hand are summarized in Appendix A1.

A complete implementation of a Gibbs sampler requires dealing with convergence issues. In particular, it is necessary to specify the variables t (truncation point) and m (stopping point) in equation (28) in Appendix A1. The first of these variables represents the number of iterations that are required to reach the steady-state (i.e., the approximately true) distribution. The t variable constitutes what is called the *burn in* period in which observations have to be discarded for computing moments of the posterior marginal densities. This represents the transient period during which the Markov Chain has not reached stability and is thus highly influenced by starting points (initial values of the parameters that have to be estimated).

The second variable, m , is the total number of iterations, and is required to determine the additional number of samples $m - t$ that have to be drawn after convergence has been reached to compute all relevant moments of the posterior distribution (equation 28). This variable is critical due to the Markov nature of the algorithm. If convergence has been reached, all the samples are identically distributed. However, since the samples are autocorrelated, the slower the simulation algorithm is in moving within the sample space, the higher the number of samples required to efficiently obtain the required estimates. Although typical choices in the literature ($t = 1000$ and $m = 10000$) have shown to be appropriate for many applications, we use convergence diagnostic algorithms proposed in the literature to evaluate if these values are satisfactory for the adjustment problem. These convergence diagnostic algorithms are briefly described next. This is followed by a discussion of the software implementation of the proposed approach to the setup adjustment problem.

5.1 Convergence diagnostic algorithms and their use in the adjustment problem

Although theoretical results oriented to define *ex-ante* the number of iterations t required to reach convergence appear to be a promising solution to deal with convergence issues in MCMC, results obtained in this field are far from being actually applied due to their mathematical complexity and excessive calculation involved (Cowles and Carlin, 1995). Thus,

most of the applied work on MCMC focuses on using a set of *diagnostic* tools for assessing convergence by analyzing output produced by MCMC simulations. Numerous studies (e.g., Brooks and Roberts, 1998; Cowles and Carlin, 1995) have compared the convergence diagnostics proposed in the literature over a set of problems in order to rank them. Since at this time, no diagnostic algorithm seems globally superior to all others, we follow the recommended approach of applying simultaneously more than one diagnostic method to assess convergence of the lot mean estimates needed for adjustment at each point in time (i, j) . In particular, we use two of the most popular methods adopted in the literature, proposed by Raftery and Lewis (1996) (RL) and Gelman and Rubin (1996) (GR).

The first algorithm (RL) is based on monitoring the autocorrelation within one single chain. It requires as input the quantile q of the posterior distribution that has to be estimated to within $\pm r$ with probability s . Default values suggested by Raftery and Lewis are $q = 0.025$, $r = 20\% \cdot q = 0.005$ and $s = 0.95$. Since the adjustment selection requires computing the posterior mean, the 0.5 quantile (median) has been selected as reference. Accuracy has been defined accordingly as ten percent of the desired quantile, thus obtaining $r = 10\% \cdot q = 0.05$. Concerning s , the value $s = 0.95$, suggested by Raftery and Lewis, has been utilized.

The RL algorithm has the main drawback of masking excessive slow convergence of Markov Chain simulation, a problem common to other methods based on a single simulated chain. It could happen that looking at one single chain, convergence seems achieved although the chain is stuck in one place of the target distribution. In order to overcome this problem, Gelman and Rubin (1992, see also Gelman, 1996) suggest to perform multiple simulations, starting from different initial values for parameters that have to be estimated. These values have to be chosen to be “overdispersed” with reference to the target density. Since the target density is not known in advance, sometime this technique has been criticized for the lack of guidelines about selecting the starting points. In our process adjustment application, however, this choice can be easily performed considering basic knowledge of the process producing the quality characteristic. For example, a rough idea of the process capability can give information on the range of variation (at least its magnitude) of the quality characteristic and this information can be easily used in determining overdispersed initial values for parameters derived through MCMC.

Once starting values are selected, the GR algorithm is based on mixing simulated chains and comparing the variance within each chain to the total variance of the mixture of chains. These (estimated) variances permit to derive the “estimated potential scale reduction” factor \hat{R} . As simulations converge, \hat{R} declines to 1, thus assessing that parallel chains are essentially overlapping. The rule of thumb suggested when performing the GR diagnostic is to continue simulation until \hat{R} is close to one, for example, lower than 1.2. In the setup adjustment application, diagnostics are run on the simulated lot mean (θ_i) corresponding to the current lot (i) .

Using both of the diagnostic algorithms briefly described, the proposed setup adjustment approach can be summarized as follows. After the first part is processed in the first lot and the first -trivial- value $U_{11} = -x_{11}$ is set on the machine, the second part in the first lot is processed and x_{12} is observed. Therefore, adjustment U_{12} has to be selected according to equation (20). The expected value required is thus computed through MCMC. Starting from the trial values suggested in the literature ($t = 1000$ and $m = 10000$) two MCMC chains are simulated, each one starting from different initial values, as previously mentioned. Then, the

RL diagnostic check is performed on each chain for θ_1 , the parameter of interest. If the RL algorithm suggests for one of the chains higher t or m values, (i.e., convergence diagnostic has failed thus far), MCMC is rerun for both chains using the new t, m parameters suggested by RL. To avoid an excessive number of reiterations, these values are rounded to the upper multiple of 500 for t and the upper multiple of 5000 for m (thus increasing t and m in steps of 500 and 5000, respectively). Hence, MCMC simulations and RL convergence checks are performed again until the diagnostic check is satisfied. Starting with the values eventually assessed for t and m , the GR diagnostic is then performed on both chains (for θ_1).

Similarly as for the RL algorithm, if the GR method concludes that convergence has been reached, i.e., the estimated potential scale reduction factor $\hat{R} \leq 1.2$, an adjustment is computed by calculating the expected value of the posterior distribution of θ_1 obtained from the two final MCMC chains. Otherwise, t and m are increased. Contrary to the RL diagnostic, the GR algorithm does not suggest new values for these parameters, so both are increased (in steps of 500 for t and 5000 for m). Both chains are thus simulated again and the GR check performed until the algorithm permits to conclude that convergence has been reached and setting U_{12} is possible. We then proceed to compute U_{13} and so on until we process all parts in all lots.

5.2 Computer implementation of the proposed approach

All the steps in the previous section are performed for each adjustment, computed following Table 1. A Visual Basic code has been written to simulate the determination of the adjustments over a set of lots. Each time an adjustment has to be computed at time (i, j) , the code launches the execution of both the MCMC simulations and the convergence checks. The MCMC simulation is coded in the *Bugs* (*Bayesian inference Using Gibbs Sampling*, Spiegelhalter *et al.*, 1994) language, while the RL and GR convergence diagnostic algorithms in the CODA library (Best *et al.*, 1995) are used running under R, the freely available version of S-plus. After MCMC convergence is assessed, the adjustment is computed and all input files required to rerun MCMC simulation for the next adjustment are automatically generated by the VB program.

With respect to computational time, the software requires few seconds to perform each MCMC simulation and around half a minute to perform each convergence check on a Pentium II 333MHz. However, due to the use of different available software packages, most of the computational time is due to writing and reading MCMC output, considering that each chain is constituted by thousands of values. Therefore, a specific code written to perform MCMC and convergence checks maintaining in memory all these values of the chains should significantly improve computational times. For most high-value added discrete part manufacturing processes we envision the MCMC adjustment approach to be applied, the aforementioned computing times should be adequate, that is, well below the time between consecutive parts are produced. The present version of the software performs the different MCMC simulation chains and the convergence checks sequentially, so computational times could be reduced by simulating different chains on parallel processors.

6 Two examples of sequential Bayesian setup offset adjustment over a set of lots

Bayesian adjustment of an initial or setup offset is useful when several lots have to be processed but no information is available on parameters characterizing offset and process distributions. Alternative feedback methods that can be used in this case are the harmonic rule due to Grubbs (1983):

$$\nabla U_{ij} = -\frac{1}{j}Y_{ij} \quad (22)$$

(with $U_{i0} = 0$) and the integral controller (Box and Luceño, 1997):

$$\nabla U_{ij} = -\lambda Y_{ij} \quad , \quad (23)$$

(with $U_{i0} = 0$) since both these rules do not require any information on offset or error distributions (Del Castillo *et al.* 2002b). The performance obtained with the MCMC-based adjustment method has been compared with the performance of these competitor rules considering two examples. Both examples were presented by Box and Tiao (1973) in the framework of Bayesian estimation in a one-way random effects model and are reproduced in Table 2. The first example refers to the yield of dyestuff measured from five samples taken from six batches of raw material while the second one relates to randomly generated data with $\mu = 4$, $\sigma_\theta = 2$ and $\sigma_v = 4$. This was used by the authors to study the Bayesian estimation of variance components in a case where sample based theory determines a negative estimate of $\hat{\sigma}_\theta^2$ obtained with equation (13) where $MS_b < MS_w$. For illustration purposes of the on-line MCMC adjustment approach, we pretend the data becomes available sequentially in time and apply the MCMC adjustment procedure at each time instant.

Two MCMC chains were simulated at each step of the adjustment procedure and different initial values of the parameters that have to be estimated (μ , σ_θ^2 , σ_v^2 and θ_i , where $i=1,\dots,6$ for the model M_1 ; σ_v^2 and θ_1 for the model M_2) were specified for each chain. Among these parameters, the θ_i 's in model M_1 and θ_1 in model M_2 can be randomly generated by the software, using equations (1) and (19), respectively. To make this random generation different for the two chains, a different seed must be specified for each chain. Initial values and seeds used for both examples are reported in Table 3. To see how the effect of these

| First example | | | | | | Second example | | | | | |
|---------------|------|------|------|------|------|----------------|-------|--------|--------|--------|--------|
| batch | | | | | | batch | | | | | |
| 1 | 1545 | 1440 | 1440 | 1520 | 1580 | 1 | 7.298 | 3.846 | 2.434 | 9.566 | 7.99 |
| 2 | 1540 | 1555 | 1490 | 1560 | 1495 | 2 | 5.22 | 6.556 | 0.608 | 11.788 | -0.892 |
| 3 | 1595 | 1550 | 1605 | 1510 | 1560 | 3 | 0.11 | 10.386 | 13.434 | 5.51 | 8.166 |
| 4 | 1445 | 1440 | 1595 | 1465 | 1545 | 4 | 2.212 | 4.852 | 7.092 | 9.288 | 4.98 |
| 5 | 1595 | 1630 | 1515 | 1635 | 1625 | 5 | 0.282 | 9.014 | 4.458 | 9.446 | 7.198 |
| 6 | 1520 | 1455 | 1450 | 1480 | 1445 | 6 | 1.722 | 4.782 | 8.106 | 0.758 | 3.758 |

Table 2: First example: dyestuff data Y_{ij} , for $i=1,\dots,6$ and $j=1,\dots,5$ (Yield of dyestuff in grams of standard color computed with respect to target value of 1400 g). Second example: data with $\mu = 4$, $\sigma_\theta = 2$ and $\sigma_v = 4$ (Box and Tiao, 1973).

| Chain | μ | σ_θ^2 | σ_v^2 | seed |
|--------|-------|-------------------|--------------|----------|
| First | 2000 | 0.001 | 0.001 | 1056126 |
| Second | -2000 | 1000 | 1000 | 56111235 |

Table 3: Initial values and seeds for the random generation adopted in the two MCMC chains simulated at each step of the adjustment procedure for both the examples.

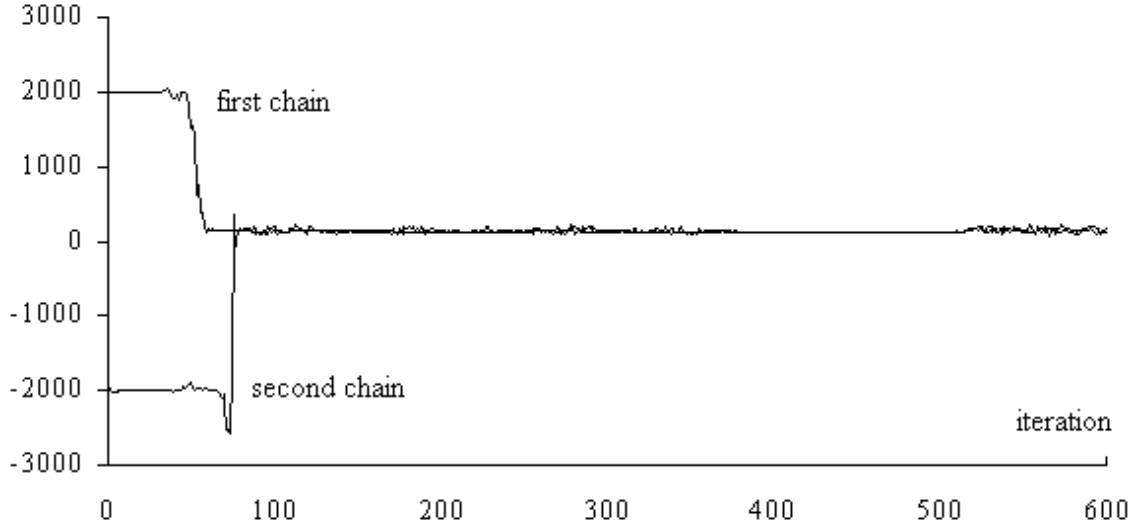


Figure 3: First 600 iterations of the two MCMC chains simulating θ_3 in adjusting the second part in the third lot for the first example.

starting values disappear as the MCMC simulation proceeds, consider the plot shown in Figure 3. It represents the first 600 iterations of the two MCMC chains estimating θ_3 in adjusting the second part of the third lot for the first example. As it can be observed, the two chains start respectively around 2000 and -2000 (their mean values) and have almost “forgotten” these starting points after 100 iterations. The trial value adopted, i.e., $t = 1000$, was observed to be adequate for all the simulations performed with the first data set while an increase to only $t = 1500$ was required for the second case studied. With respect to the total number of runs m , it was started from the trial value of 10000. Due to high autocorrelation of the chains, it was increased as suggested by convergence algorithms to the final values of 45000 and 20000 for the first and the second examples, respectively.

Adjustments were computed with the proposed procedure for all the lots processed in the two examples. After the thirty adjustments were computed, a quadratic symmetric loss function, given by:

$$C = \sum_{i=1}^6 \sum_{j=1}^5 Y_{ij}^2, \quad (24)$$

was used to compare the performance obtained with this approach vs. Grubbs’ harmonic rule and two discrete integral controllers (or EWMA controllers) with different weights $\lambda =$

| | Adjustment rule | Costs C | ΔC % |
|----------------|-------------------------|-----------|--------------|
| First example | MCMC | 160982.39 | |
| | GRUBBS | 221316.15 | 27% |
| | EWMA($\lambda = 0.4$) | 254072.75 | 37% |
| | EWMA($\lambda = 0.1$) | 437815.98 | 63% |
| Second example | MCMC | 533.27 | |
| | GRUBBS | 649.27 | 18% |
| | EWMA($\lambda = 0.4$) | 745.95 | 29% |
| | EWMA($\lambda = 0.1$) | 1072.52 | 50% |

Table 4: Costs C (equation 24) related to the different adjustment rules and percentage savings ΔC % induced by the MCMC-based approach over the other rules for the two examples.

0.1 and $\lambda = 0.4$. These weights were chosen as they cover the range of EWMA weights recommended in the process adjustment literature (Box and Luceño, 1997). Large values of λ (i.e., closer to 1) will bring the process back to target faster on average, but will increase the variance of the controlled process (Del Castillo, 2001), also increasing the cost function (24).

Figure 4 reports the pattern of deviations from target Y_{ij} for the dyestuff data (first example) observed in each lot using Grubbs' harmonic rule, the MCMC-based adjustment method, and the two EWMA controllers (with $\lambda = 0.1$ and 0.4 , respectively). As it can be observed, the deviations from target induced by the MCMC approach are smaller than the ones due to the other rules, especially after the first two lots have been machined, i.e., after information on the variance between lots becomes available.

The deviations from target Y_{ij} observed for the second example in each lot using the four competing rules are reported in Figure 5, which, as before, testifies the better performance of the MCMC approach compared to the other adjustment rules, especially after the first two lots have been processed.

Global savings in costs attained in each lot using the MCMC-based approach versus Grubbs' rule and the two EWMA controllers are reported in Figure 6 for both the examples studied. As it can be observed, savings in costs are almost always positive, showing that the MCMC approach outperforms the competitor adjustment rules considered. Furthermore, the three alternative methods can be ranked noting that savings induced by the MCMC approach decrease going from the EWMA controller ($\lambda = 0.1$) to the other EWMA controller ($\lambda = 0.4$) and eventually to Grubbs' adjustment rule. This ranking is even more clear considering the total costs C obtained applying the different adjustment rules to the two examples considered (Table 4). As it can be observed, percentage advantages determined by the MCMC Bayesian procedure vary from 27 % to 63 % for the first example and from 18 % to 50 % for the second one. The savings are mainly due to the ability to predict the initial offset before starting a new lot.

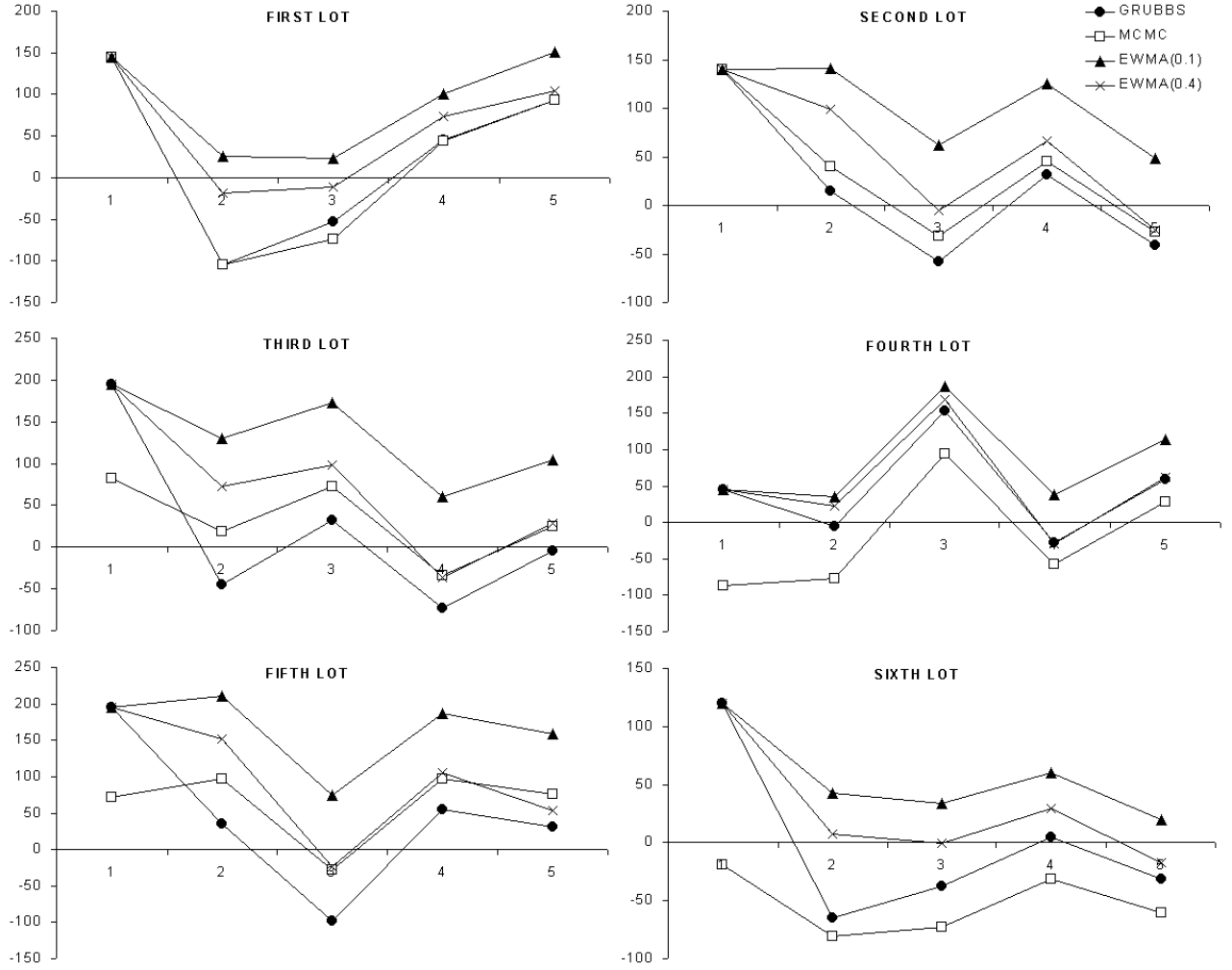


Figure 4: Deviations from target Y_{ij} obtained using Grubbs' harmonic rule, the MCMC approach and the two EWMA controllers (with $\lambda = 0.1$ and 0.4) to adjust an initial offset for the dyestuff data (Table 2).

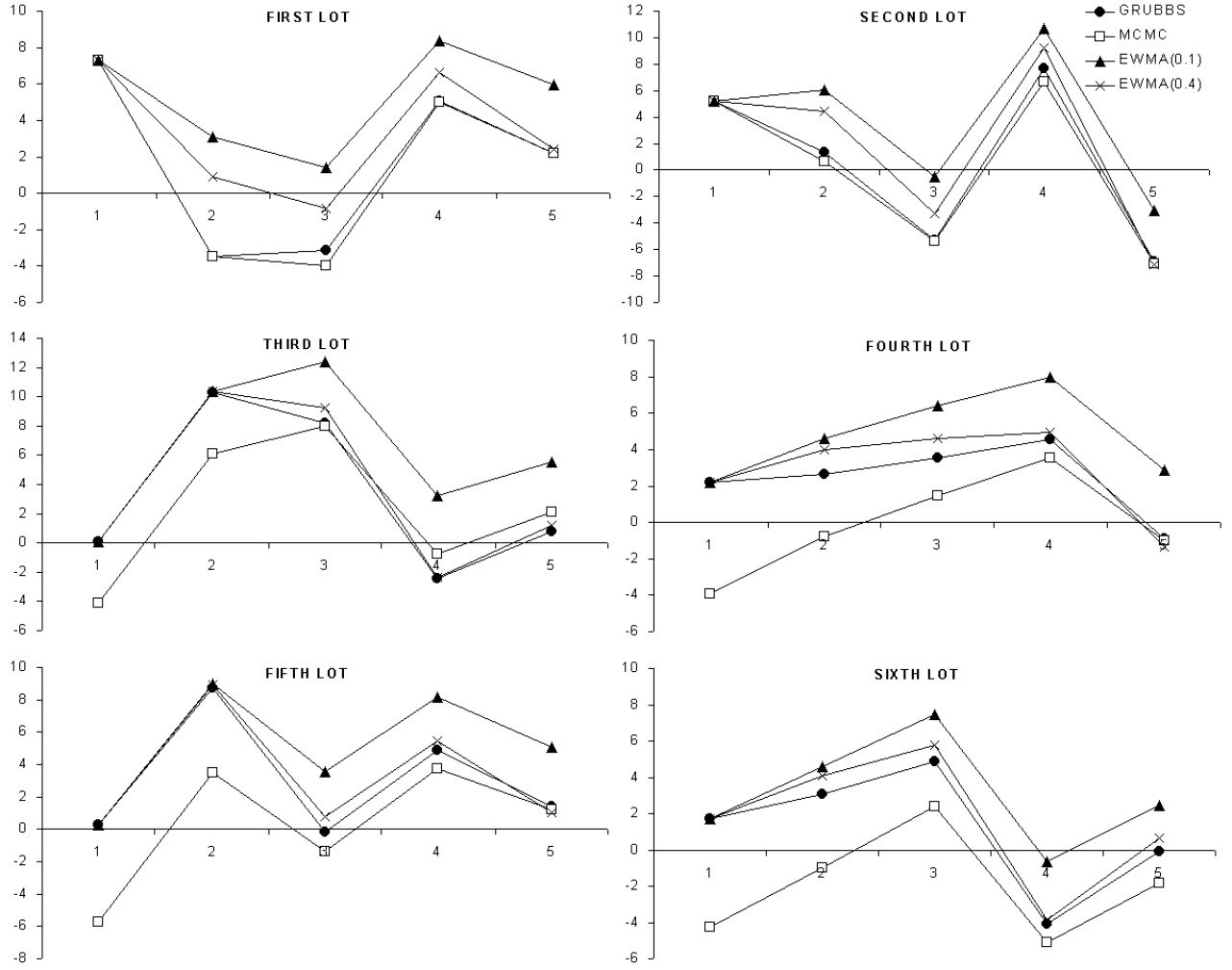


Figure 5: Deviations from target Y_{ij} obtained using Grubbs' harmonic rule, the MCMC approach and the two EWMA controllers (with $\lambda = 0.1$ and 0.4) to adjust an initial offset for the second example (Table 2).

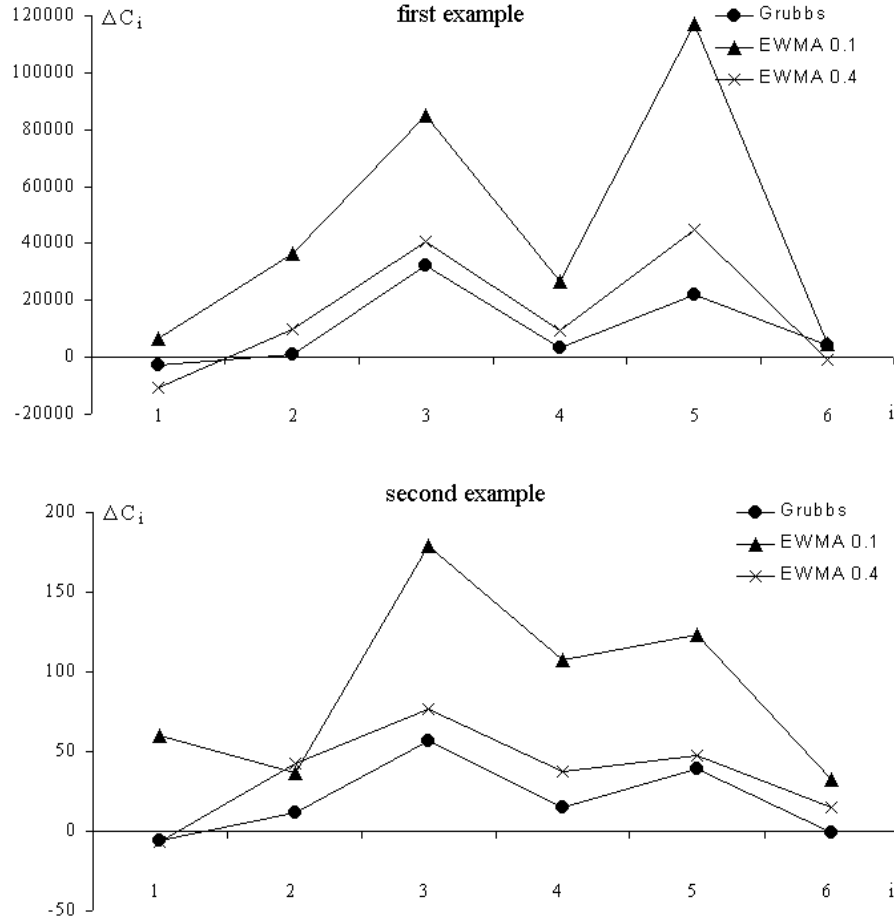


Figure 6: Savings in costs per each lot $\Delta C_i = C_i^{\text{alternative rule}} - C_i^{MCMC}$ obtained with the MCMC approach vs. alternative rules, i.e. Grubbs' harmonic rule and two EWMA controllers –with $\lambda = 0.1$ and 0.4 .

7 Conclusions and further research

This article presented a procedure to adjust the setup errors of a production process over a set of lots assuming a one-way random effects model (with unknown parameters) for the initial offset in each lot. The approach presented is based on repeatedly estimating the initial offset in each lot through Markov Chain Monte Carlo simulation of a Bayesian hierarchical random effects model. Assuming a quadratic off-target cost model, the approach presented was compared with adjustment rules that can be applied when parameters of the initial offset distribution are unknown, i.e. Grubbs' rule (based on stochastic approximation) and discrete integral controllers (or EWMA controllers). The procedure was illustrated with application to two examples presented in the literature (Box and Tiao, 1973) where the advantages attained with the approach presented here vary from 18 % to 27 % and from 29 % to 37 % when considering Grubbs' rule and the best EWMA controller utilized ($\lambda = 0.4$), respectively. The proposed approach for setup adjustment can be applied to high-value added, short run manufacturing processes, where the computational expense and relative complexity of the proposed adjustment rule compared to the simplicity of the alternative rules studied can be justified.

Starting from these encouraging results, a first direction for further research can be to extend the numerical comparisons to better understand the effect of different factors (such as the lot size, the number of lots, the parameters characterizing the initial offset distribution) on the advantages obtained with the Bayesian approach. As previously mentioned, a further direction concerns the situation where the number of (small) lots processed is relevant. In this case, it is possible to estimate the unknown parameters through MCMC and plug in these estimates in a simpler adjustment rule that assume known parameters, e.g., in Grubbs' extended rule.

Taking advantage of the flexibility of hierarchical models, a further direction of research could be related to different extensions of the approach presented to model situations that can arise in short-run manufacturing. As examples, measurement errors or tool wear can be easily included in the study adding a stage in the hierarchy and relaxing the assumption of i.i.d. random errors, respectively.

Acknowledgements

This research was partially funded by NSF grant DMI 0200056.

References

- Best, N.G., Cowles, M.K., and Vines, S.K., (1995) *CODA manual version 0.30*, MRC Biostatistics Unit, Cambridge, UK.
- Box, G.E.P. and Luceño, A. (1997) *Statistical Control by Monitoring and Feedback Adjustment*, John Wiley & Sons, New York, NY.

- Box, G.E.P. and Tiao, G.D. (1973) *Bayesian Inference in Statistical Analysis*, John Wiley and Sons, Inc.
- Brooks, S.P. and Roberts, G.O (1998) Assessing Convergence of Markov Chain Monte Carlo Algorithms. *Statistics and Computing*, **8**, 319-335.
- Carlin, B.P., and Louis, T.A. (1996) *Bayes and Empirical Bayes Methods for Data Analysis*, Chapman & Hall.
- Cowles, M.K. and Carlin, B.P. (1995) Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review. *Journal of the American Statistical Association*, **91**, 883-904.
- Del Castillo, E. (2001) Some Properties of EWMA Feedback Quality Adjustment Schemes for Drifting Processes. *Journal of Quality Technology*, **33**(2), 153-166.
- Del Castillo, E., Pan, Rong, and Colosimo, B.M. (2002a) An Unifying View of Some Process Adjustment Methods. To appear in *Journal of Quality Technology*. Technical Paper, Applied Statistics Laboratory, Department of Industrial & Manufacturing Engineering, Penn State University.
- Del Castillo, E., Pan, R., and Colosimo, B.M. (2002b) Small sample comparison of some process adjustment methods. To appear in *Communications in Statistics, Simulation and Computation*. Submitted for publication. Technical Paper, Applied Statistics Laboratory, Dept. of Industrial & Manufacturing Engineering, Penn State University.
- Gelfand, A.E. and Smith, A.F.M. (1990) Sampling-based Approaches to Calculating Marginal Densities. *Journal of the American Statistical Association*, **85**, 410, 398-409.
- Gelfand, A.E., Hills, S.E., Racine-Poon, A. and Smith, A.F.M. (1990) Illustration of Bayesian Inference in Normal Data Models Using Gibbs Sampling. *Journal of the American Statistical Association*, **85**, 412, 972-985.
- Gelman, A. (1996) Inference and monitoring convergence, in W.R. Gilks, S. Richardson and D.J. Spiegelhalter (eds.), *Markov Chain Monte Carlo in Practice* (pp. 131-140), London, UK: Chapman & Hall.
- Gelman, A., and Rubin, D. (1992) Inference from Iterative Simulation Using Multiple Sequences, *Statistical Science*, **7**, 457-511.
- Gelman, A., Carlin, J.B., Stern H.S. and Rubin, D.B. (1995) *Bayesian Data Analysis*, Chapman & Hall.
- Gilks, W.R., Richardson S. and Spiegelhalter, D.J. (edited by) (1996) *Markov Chain Monte Carlo in Practice*, Chapman & Hall.
- Raftery, A.E. and Lewis, S.M. (1996) Implementing MCMC, in W.R. Gilks, S. Richardson, and D.J. Spiegelhalter (eds.), *Markov Chain Monte Carlo in Practice* (pp. 115-130), London,

UK: Chapman & Hall.

Grubbs, F.E. (1954, 1983) An Optimum Procedure for Setting Machines or Adjusting Processes. *Industrial Quality Control*, July, reprinted in *Journal of Quality Technology*, 1983, **15**, 4, 186-189.

Spiegelhalter, D.J., Thomas, A., Best, N.G. and Gilks, W.R. (1994) *BUGS: Bayesian inference Using Gibbs Sampling, version 0.50*, Cambridge: Medical Research Council Biostatistics Unit.

Sullo, P. and Vandeven, M. (1999) Optimal adjustment strategies for a process with run-to-run variation and 0-1 quality loss. *IIE Transactions*, **31**, 1135-1145.

Trietsch, D., (1998) The Harmonic Rule for Process Setup Adjustment with Quadratic Loss. *Journal of Quality Technology*, **30**, 1, 75-84.

Appendix A1: Gibbs Sampling in the one-way random effects model

To estimate all the unknown parameters and hyperparameters in the random effects model, the Bayesian approach requires to assume a set of prior distributions. A common choice for the random effects model (Gelfand *et al.*, 1990; Gelfand and Smith, 1990) is adopting conjugacy at each step of the hierarchical model, thus assuming:

$$\begin{aligned}\mu|\mu_0, \sigma_0^2 &\sim N(\mu_0, \sigma_0^2) , \\ \sigma_\theta^2|a_1, b_1 &\sim IG(a_1, b_1) , \\ \sigma_v^2|a_2, b_2 &\sim IG(a_2, b_2) ,\end{aligned}\tag{25}$$

where $\mu_0, \sigma_0^2, a_1, b_1, a_2, b_2$ are assumed known. In particular, a_1, b_1 (and/or a_2, b_2) can be assumed equal to 0 to model the usual improper prior for σ_θ^2 (σ_v^2) and a vague prior for μ can be assumed by setting σ_0^2 sufficiently great, e.g., $\sigma_0^2 \geq 10^4$.

If all the observations from all lots $\mathbf{x}^{IJ} = (x_{ij}, i = 1, \dots, I; j = 1, \dots, J)$ have been collected, inference on the unknown parameters can be performed as follows. Due to conjugacy adopted at each step, all the full conditional posterior can be derived as follows (Gelfand *et*

al., 1990):

$$\begin{aligned}
\pi'(\mu|\mathbf{x}^{IJ}, \boldsymbol{\theta}, \sigma_\theta^2, \sigma_v^2) &= N\left(\frac{\sigma_\theta^2\mu_0 + \sigma_0^2\sum\theta_i}{\sigma_\theta^2 + I\sigma_0^2}, \frac{\sigma_\theta^2\sigma_0^2}{\sigma_\theta^2 + I\sigma_0^2}\right) \\
\pi'(\sigma_\theta^2|\mathbf{x}^{IJ}, \boldsymbol{\theta}, \mu, \sigma_v^2) &= IG\left(a_1 + \frac{1}{2}I, b_1 + \frac{1}{2}\sum(\theta_i - \mu)^2\right) \\
\pi'(\sigma_v^2|\mathbf{x}^{IJ}, \boldsymbol{\theta}, \mu, \sigma_\theta^2) &= IG\left(a_2 + \frac{1}{2}IJ, b_2 + \frac{1}{2}\sum\sum(x_{ij} - \theta_i)^2\right) \\
\pi'(\boldsymbol{\theta}|\mathbf{x}^{IJ}, \mu, \sigma_\theta^2, \sigma_v^2) &= N\left(\frac{J\sigma_\theta^2}{J\sigma_\theta^2 + \sigma_v^2}\bar{\mathbf{x}}^{I\cdot} + \frac{\sigma_v^2}{J\sigma_\theta^2 + \sigma_v^2}\mu\mathbf{1}, \frac{\sigma_\theta^2\sigma_v^2}{J\sigma_\theta^2 + \sigma_v^2}\mathbf{I}\right)
\end{aligned} \tag{26}$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_I)$, $\bar{x}_{i\cdot} = \sum_j \frac{x_{ij}}{J}$, $\bar{\mathbf{x}}^{I\cdot} = (\bar{x}_{1\cdot}, \dots, \bar{x}_{I\cdot})$, $\mathbf{1}$ is a $I \times 1$ column vector of 1s and \mathbf{I} is the $I \times I$ identity matrix. Although similar to the random effects model, the process adjustment problem focuses on the lot means θ_i as the parameters of interest, rather than on the variance components. The conditional posteriors can be all expressed in closed form as in (26), but the computation of the marginal posteriors requires solving complex high-dimensional integrals.

Introducing a more general notation, assume that the hierarchical model requires the estimation of K parameters $\mathbf{W} = (W_1, W_2, \dots, W_K)$, which are modelled as random variables in the Bayesian framework. In the random effects model, $K = I + 3$, i.e., the I components of $\boldsymbol{\theta}$, μ , σ_v^2 and σ_θ^2 . Due to conjugacy (although this is not a required hypothesis for the computational approach described) all the full conditional posteriors

$$p(w_k|\mathbf{x}^{ij}, w_1, w_2, \dots, w_{k-1}, w_{k+1}, \dots, w_K)$$

can be assumed known. To estimate each parameter W_k , the marginal posterior $p(w_k|\mathbf{x}^{ij})$ has to be computed.

Markov Chain Monte Carlo methods can be used to estimate these posterior marginals. In particular, one of the most powerful MCMC methods is the *Gibbs Sampler*, which has the relevant advantage of being almost independent of the number of parameters and stages in a hierarchical model. Consider the general problem of estimating K parameters $\mathbf{W} = W_1, W_2, \dots, W_K$ and assume that the full conditional distributions $p(w_k|w_{l \neq k}), k = 1, \dots, K$, are all given. Given an arbitrary set of starting values $(W_1^{(0)}, W_2^{(0)}, \dots, W_K^{(0)})$, the first iteration in Gibbs sampling is performed as follows:

$$\begin{aligned}
\text{Draw } W_1^{(1)} &\sim p(w_1|W_2^{(0)}, \dots, W_K^{(0)}) ; \\
\text{Draw } W_2^{(1)} &\sim p(w_2|W_1^{(0)}, W_3^{(0)}, \dots, W_K^{(0)}) ; \\
&\dots \\
\text{Draw } W_K^{(1)} &\sim p(w_K|W_1^{(0)}, \dots, W_{K-1}^{(0)}) ;
\end{aligned}$$

thus obtaining as a result a set of points $(W_1^{(1)}, W_2^{(1)}, \dots, W_K^{(1)})$ which represent the starting values for the next step. Iterating the process also in this case the convergence is proved. Indeed, denoting with l the generic iteration of the algorithm, it can be showed that:

$$(W_1^{(l)}, W_2^{(l)}, \dots, W_K^{(l)}) \xrightarrow{d} (W_1, W_2, \dots, W_K) \sim p(w_1, w_2, \dots, w_K) \tag{27}$$

Assume convergence has been reached after t steps, $m - t$ further samples can be drawn and the marginal posterior density for each variable W_k can be then computed as:

$$\hat{p}(W_k) = \frac{1}{m - t} \sum_{l=t+1}^m p_l(W_k | W_r = W_r^{(l)}, k \neq r) . \quad (28)$$

that can be viewed as a "Rao-Blackwellized" density estimator (Gelfand *et al.*, 1990).