Posterior Distribution Charts: A Bayesian Approach for Graphically Exploring a Process Mean

Daniel W. Apley

Department of Industrial Engineering and Management Sciences
Northwestern University
Evanston, IL 60208-3119
(*apley@northwestern.edu*)

We develop a Bayesian approach for monitoring and graphically exploring a process mean and informing decisions related to process adjustment. We assume a rather general model, in which the observations are represented as a process mean plus a random error term. In contrast to previous work on Bayesian methods for monitoring a mean, we allow any Markov model for the mean. This includes a mean that wanders slowly, that is constant over periods of time with occasional random jumps or combinations thereof. The approach also allows for any distribution for the random errors, although we focus on the normal error case. We use numerical integration to update relevant posterior distributions (e.g., for the current mean or for future observations), as each new observation is obtained, in a computationally inexpensive manner. Using an example from automobile body assembly, we illustrate how the approach can inform decisions regarding whether to adjust a process. Supplementary Materials for this article, including code for implementing the charts, are available online on the journal web site.

KEY WORDS: Bayesian monitoring; Control charts; Mean tracking; Process capability analysis; Statistical process control.

1. INTRODUCTION

Control charts, and in particular Shewhart charts, have been widely used in industry for decades. One shortcoming of Phase II control charting is that the primary piece of information that is produced—whether there is a statistically significant change from the in-control state—is largely black and white. In contrast, engineering decisions regarding whether one should adjust or intervene in the process are invariably driven by considerations that come in many shades of gray: How sure are we that process parameters have changed? How large was the change? How far off target is the current process mean? What fraction of the process distribution will fall out of specifications (or close to the specification limits), both before and after any potential process intervention? Moreover, this potentially quantifiable information must be balanced with other, less quantifiable considerations such as how costly (in time, physical resources, lost production if the process must be shut down to make a change, etc.) will the intervention be, whether key personnel are available or have higher-priority concerns at the moment, the negative consequences that the intervention might potentially have on other areas of the process, etc.

The following example from automobile body assembly illustrates these issues. To improve the dimensional control of the automobile bodies, laser-optical measurement stations are integrated into the assembly line. A number of critical dimensional features are automatically measured on every autobody. The desire is to keep the mean of each feature on target and control variation about the mean. A reduction in variation typically requires the replacement of worn tooling or a redesign of some portion of the assembly process. In contrast, a shift in the mean away from the target can usually be corrected by a "shim move," which entails inserting thin metal shims of appropriate thicknesses into various tooling elements that locate the sheet metal parts when they are joined. Inserting a shim changes the position of the locating elements in a fixture. Because the locating elements control the position of parts placed into the fixture, a shim move effectively changes the mean of various dimensional features. The exact size of the mean change brought about by a shim move is sometimes difficult to predict precisely, and the appropriate shim thickness is often found by trial and error. The process and measurement system are described in more detail in the article by Apley and Shi (2001).

Process engineers meet frequently to review the data and to decide whether shim moves or other interventions are required. As mentioned earlier, control charts are not ideally suited for informing decisions of this nature, because of their emphasis on signaling statistical significance, without regard to practical significance. Engineers must also assess whether a statistically significant shift is large enough to warrant a shim move and if so, how large a shim move to make, while taking into account the fact that there is uncertainty in the exact location of the mean. One might consider a decision-theoretic approach, in which one assumes a probabilistic model for the process, assigns costs to every aspect of action/inaction, and determines the optimal shim move. A problem with this approach is that

TECHNOMETRICS, AUGUST 2012, VOL. 54, NO. 3
DOI: 10.1080/00401706.2012.694722

costs are virtually impossible to accurately specify for many processes like autobody assembly. First, the true costs of an off-target dimension are difficult to specify for many reasons, one of which is that a great deal depends on what is happening downstream. Perhaps an off-target dimension will cause major assembly difficulties or quality issues at a downstream stage of the process; or perhaps the downstream operators have improvised a reasonable way to compensate for an off-target dimension and mitigate its adverse effects. Second, the costs of performing a shim move are difficult to specify. If process engineers currently have many other fires to put out elsewhere in the process, then the hidden cost of a shim move may be very high (because it takes resources away from other problems). On the other hand, if it is a slow day, and there currently are no other pressing problems, then devoting engineering manpower to a shim move may entail very little additional cost. For these and other reasons, process engineers usually do not want a fully automated decision-theoretic algorithm to dictate when to take action. More useful are easy-to-interpret graphical displays of information that will help engineers balance the severity of the problem with the myriad of other quantifiable and unquantifiable concerns, so that they can make a well-informed decision on whether to intervene or adjust the process.

In this article, we investigate an alternative to Phase II control charting that is better suited for these purposes. We assume that the process observations $\{x_t\colon t = 1, 2, \ldots\}$ are represented by the rather generic model

$$x_t = \mu_t + \varepsilon_t, \tag{1}$$

where ε_t is an iid random sequence, and μ_t represents the "mean" of the process at observation number t. For flexibility, we allow any Markov model for μ_t. Although we focus on the case of normally distributed ε_t, the procedures that we describe in Sections 2 and 3 apply directly for an arbitrary distribution for ε_t, providing one knows or can estimate the distribution. A special case that we use to illustrate the approach is $\varepsilon_t \sim N(0,\sigma^2)$, and μ_t constant over periods of time but with occasional jumps of random magnitude and at random points in time, according to the random jump (RJ) model

$$\mu_t = \begin{cases} \mu_{t-1}: & \text{with probability } 1-p \\ \mu_{t-1} + N\left(0, \eta^2\right): & \text{with probability } p \end{cases}, \tag{2}$$

where η and p are parameters.

As each new observation (say x_t at time period t) is obtained, we use a Bayesian analysis to update a set of relevant posterior distributions, given all of the available data $\mathbf{x}^t = \{x_j\colon j = 1, 2, \ldots, t\}$, and then we chart the posterior distributions as they evolve over time. Posterior distributions of particular interest are those of $\mu_t \mid \mathbf{x}^t$ and of $x_{t+1} \mid \mathbf{x}^t$. The distributions can be plotted together with upper and lower specification limits (denoted by USL and LSL) and a target value (denoted by T), which allows a graphical assessment of issues related to process capability and whether adjustments or interventions should be made. Numerical summaries of the posterior distributions, such as the predicted fraction that falls outside of specifications, can also be tabulated to aid in decision making. Figures 1–4 and the surrounding discussions in later sections illustrate the various graphical and numerical results that can be produced in the posterior distribution (PD) charts. The intent is to provide diagnostic information and graphical displays that can inform engineering decision making, in contrast to the Phase II control charting intent of providing an automated signal when a process parameter has changed.

There has been a substantial amount of previous work on Bayesian methods to update the posterior distribution of $\mu_t \mid \mathbf{x}^t$ in the context of control charting. Early works include those by Barnard (1959) and Chernoff and Zacks (1964), who developed Bayesian methods for tracking μ_t with a RJ model quite similar to Equations (1) and (2). More recent works include those by Crowder and Eshleman (2001), Tsiamyrtzis and Hawkins (2005), and Triantafyllopoulos (2007), who developed Bayesian methods for updating the posterior distribution of $\mu_t \mid \mathbf{x}^t$ under different models for μ_t. Crowder and Eshleman assumed a random walk (RW) μ_t. Tsiamyrtzis and Hawkins (2005) assumed a RW μ_t with, in addition, jumps of fixed, prespecified magnitude and direction, but at random times. Tsiamyrtzis and Hawkins (2007) extended their model by, among other things, allowing jumps of random size. Triantafyllopoulos (2007) assumed a RW μ_t, where the variances of ε_t and of the RW increments are unknown and must be tracked also. Woodward and Naylor (1993) assumed a piecewise constant mean in which the jump times are known (e.g., coinciding with when known disturbances occurred).

In addition, many authors have proposed tracking the process mean using Kalman filtering methods (Kalman 1960), which have Bayesian interpretations for certain linear Gaussian models, such as when μ_t is a Gaussian RW, and ε_t is Gaussian (see Meinhold and Singpurwalla 1983). For example, Downing, Pike, and Morrison (1980) compared a Kalman filter to a CUSUM control chart for an inventory control problem in which three parameters were tracked; Crowder (1986) investigated many aspects of Kalman filtering for statistical process control (SPC); Wasserman and Sudjianto (1993), del Castillo and Montgomery (1995), Sastri, Valdes, and Flores (1996), and others have developed Kalman filters for tracking the mean and other process parameters in the context of control charting. The edited volume by Colosimo and del Castillo (2007) includes a number of chapters on related methods and Bayesian analyses for process monitoring, control, and optimization.

Other notable Bayesian control charting works include those by Hamada (2002), Menzefricke (2002), and Bayarri and Garcia-Donato (2005), who developed methods for setting the control limits for various control charts, under consideration of uncertainty in various distribution parameters. The control limits are based on marginalizing the distribution of the charted statistic with respect to the posterior distribution of the unknown parameters, given $\mathbf{x}^t$ (or given a Phase I set of data from which in-control values for the distribution parameters must be inferred). There is also a large body of work on Bayesian methods for developing economically optimal inspection and decision policies in a quality control setting (e.g., Girshick and Rubin 1952; Taylor 1967; Ross 1971; White 1977; Calabrese 1995; Tagaras and Nikolaidis 2002).

This article is most closely aligned with the Bayesian methods for tracking μ_t, but our approach and presentation differ from the previous work in the following respects. Most of the previous work has placed a premium on having closed-form analytical expressions for updating the posterior distributions,

at the expense of imposing restrictive conditions on the model and/or using simplifying approximations when deriving the posteriors. Examples of restrictive conditions include disallowing a RJ model, which seems a quite reasonable model for many applications in which control charting is typically used, in favor of the more tractable RW model; requiring all variables to be Gaussian; and/or (for the RJ model) assuming the size of any jump is fixed and known. In contrast, our use of numerical integration allows any Markov model for μ_t with no restrictions on the distribution of ε_t. The only drawback of this added flexibility is that numerical integration is required to update the posteriors, but the overall computational expense is trivial considering modern desktop computing power.

This work also differs from previous work in many of the graphical implementation and interpretation issues on which we focus our discussion. After describing the general process model and the equations for updating the relevant posterior distributions in Section 2, we discuss various quantities that can be charted and how to interpret the charts in Section 3. In Section 4, we discuss Markov models for μ_t other than the RJ model of Equation (2). In Section 5, we present design guidelines for selecting or estimating various PD chart parameters. Section 6 presents Monte Carlo simulations that compare different Bayesian methods for tracking a mean and investigate robustness to model misspecification. Section 7 concludes the article.

2. UPDATING THE POSTERIOR DISTRIBUTIONS FOR THE GENERAL PROCESS MODEL

Let $h(\varepsilon)$ denote the distribution of ε_t, and let $p(\mu \mid \nu)$ denote the conditional distribution of μ_t given $\mu_{t-1} = \nu$. The following apply for arbitrary $h(\bullet)$ and $p(\bullet \mid \bullet)$. In general, for two random variables X and Y, let $f_{X \mid Y}(x \mid y)$ denote the conditional distribution of X given $Y = y$. For $\tau = 0, 1, 2, \ldots n$, and $t = \tau, \tau + 1, \tau + 2, \ldots \tau + n$, define $g^{\mu}_{t|\tau}(\mu) = f_{\mu_t \mid \mathbf{X}^{\tau}}(\mu \mid \mathbf{x}^{\tau})$. Our strategy for calculating $g^{\mu}_{t|t}(\bullet)$, the posterior distribution of $\mu_t \mid \mathbf{x}^t$, is recursive. That is, one begins with an initial distribution $g^{\mu}_{0|0}(\bullet)$ and calculates $g^{\mu}_{t|t}(\bullet)$ from $g^{\mu}_{t-1|t-1}(\bullet)$ for $t = 1, 2, \ldots n$, as each new observation x_t is obtained. A straightforward application of Bayes' rule gives

$$\begin{aligned} g^{\mu}_{t|t}(\mu) &= f_{\mu_t \mid X_t, \mathbf{X}^{t-1}}(\mu \mid x_t, \mathbf{x}^{t-1}) \\ &= \frac{f_{X_t \mid \mu_t, \mathbf{X}^{t-1}}(x_t \mid \mu, \mathbf{x}^{t-1})\, f_{\mu_t \mid \mathbf{X}^{t-1}}(\mu \mid \mathbf{x}^{t-1})}{\int f_{X_t \mid \mu_t, \mathbf{x}^{t-1}}(x_t \mid \nu, \mathbf{x}^{t-1})\, f_{\mu_t \mid \mathbf{X}^{t-1}}(\nu \mid \mathbf{x}^{t-1}) d\nu} \\ &= \frac{f_{X_t \mid \mu_t}(x_t \mid \mu)\, f_{\mu_t \mid \mathbf{X}^{t-1}}(\mu \mid \mathbf{x}^{t-1})}{\int f_{X_t \mid \mu_t}(x_t \mid \nu)\, f_{\mu_t \mid \mathbf{X}^{t-1}}(\nu \mid \mathbf{x}^{t-1}) d\nu} \\ &= \frac{h(x_t - \mu)\, g^{\mu}_{t|t-1}(\mu)}{\int h(x_t - \nu)\, g^{\mu}_{t|t-1}(\nu) d\nu}. \end{aligned} \tag{3}$$

Similarly, the function $g^{\mu}_{t|t-1}(\bullet)$ is obtained from $g^{\mu}_{t-1|t-1}(\bullet)$ via

$$\begin{aligned} g^{\mu}_{t|t-1}(\mu) &= \int f_{\mu_t \mid \mu_{t-1}, \mathbf{X}^{t-1}}(\mu \mid \nu, \mathbf{x}^{t-1})\, f_{\mu_{t-1} \mid \mathbf{X}^{t-1}}\left(\nu \mid \mathbf{x}^{t-1}\right) d\nu \\ &= \int p(\mu \mid \nu)\, g^{\mu}_{t-1|t-1}(\nu)\, d\nu, \end{aligned} \tag{4}$$

where the last equality follows from the Markov property of μ_t.

Together, Equations (3) and (4) are the recursions for calculating $g^{\mu}_{t|t}(\bullet)$ from $g^{\mu}_{t-1|t-1}(\bullet)$, based on $h(\bullet)$, $p(\bullet \mid \bullet)$, and the newest observation x_t. The main computational expense comes from Equation (4). For each value of μ in the support of $g^{\mu}_{t|t}(\bullet)$, one must perform the numerical integration in Equation (4). The numerical integration in the denominator of Equation (3) adds negligible computational expense, because it needs to be carried out only a single time (it does not depend on μ). In essence, it serves only to normalize the distribution $g^{\mu}_{t|t}(\bullet)$ so that it integrates to one. If K denotes the number of discrete points over the range of integration, the computational expense of updating $g^{\mu}_{t-1|t-1}(\bullet)$ to $g^{\mu}_{t|t}(\bullet)$ is proportional to K^2. In all of the examples discussed later, we used $K = 500$ and a uniform discretization. With $K = 500$, it took roughly 0.005 sec to update $g^{\mu}_{t-1|t-1}(\bullet)$ to $g^{\mu}_{t|t}(\bullet)$ in a Matlab™ environment on a notebook computer with a 1.73 GHz Intel Core™ i7 processor. This is far from prohibitive for almost any SPC application that we envision. With such modest computational expense, we recommend choosing K conservatively large to ensure that any inaccuracies due to numerical approximation of the integrals are negligible. One may wish to downsample $g^{\mu}_{t|t}(\bullet)$ when plotting the posterior distributions (it may take longer to plot $g^{\mu}_{t|t}(\bullet)$ than to update it). Regarding inaccuracies, we observed no appreciable difference in the posterior distributions when we increased K from 500 to 1000.

Based on $g^{\mu}_{t|t}(\bullet)$, one can calculate a number of other relevant posterior distributions, such as the predictive distribution of the future observation $x_{t+1} \mid \mathbf{x}^t$. Define $g^{x}_{t|\tau}(x) = f_{X_t \mid \mathbf{X}^{\tau}}(x \mid \mathbf{x}^{\tau})$. If we assume that the future mean is the same as the current one (i.e., $\mu_{t+1} = \mu_t$, the rationale for which we discuss in the next section), the posterior predictive distribution of $x_{t+1} \mid \mathbf{x}^t$ is

$$\begin{aligned} g^{x}_{t+1|t}(x) &= \int f_{X_{t+1} \mid \mu_{t+1}}(x \mid \mu)\, f_{\mu_{t+1} \mid \mathbf{X}^t}(\mu \mid \mathbf{x}^t)\, d\mu \\ &= \int h(x - \mu)\, g^{\mu}_{t|t}(\mu)\, d\mu. \end{aligned} \tag{5}$$

Note that for computation and plotting purposes, the posterior distributions (e.g., $g^{\mu}_{t|t}(\bullet)$) are discretized over their support and stored as vectors. This provides a completely nonparametric representation of the posterior distributions. One desirable consequence of this is that multiple modes in the posterior may emerge where the data imply they should, the importance of which we illustrate in subsequent examples. Matlab code for implementing the procedure and constructing the plots illustrated in Section 3 is available as online Supplementary Materials for this article.

3. SOME RELEVANT GRAPHICAL DISPLAYS AND NUMERICAL CRITERIA

The generality that results from numerical calculation of the posteriors in the recursions of Equations (3)–(5) allows a number of possible graphical displays of various relevant information. In this section, we discuss a few possibilities and illustrate with the autobody assembly example introduced in Section 1. The example data $\{x_t: t = 1, 2, \ldots, 30\}$ are measurements of one critical autobody dimension for a sample of 30 consecutively produced automobiles. The data, which are in units of mm,

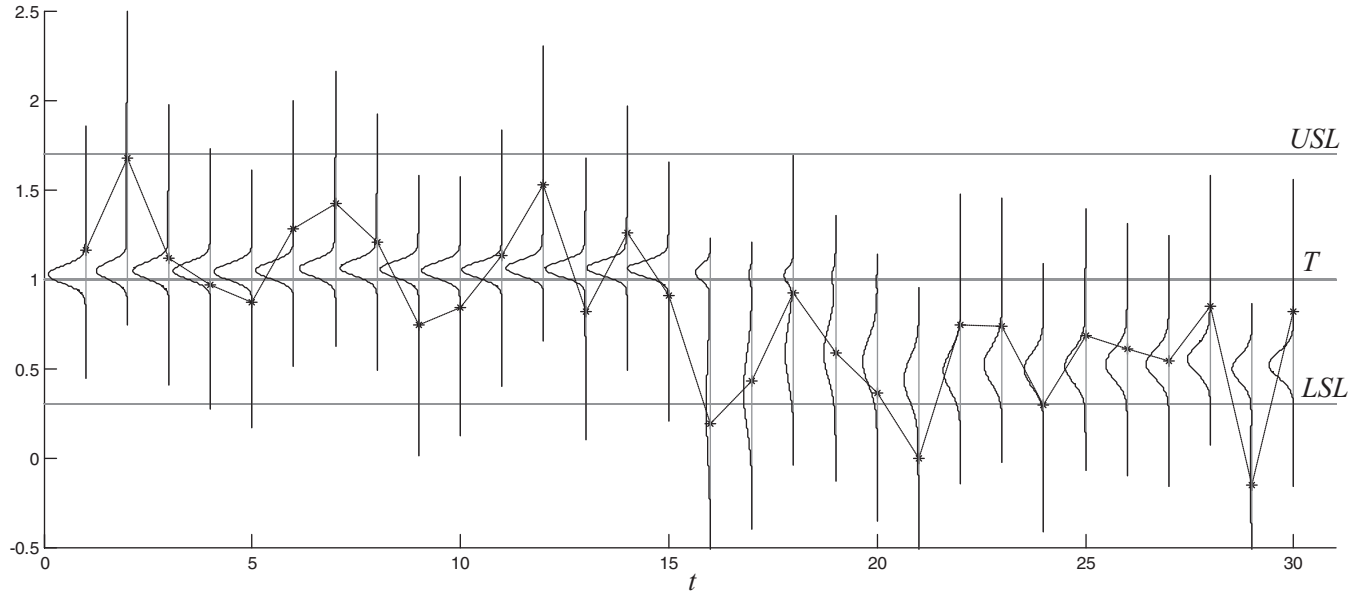

Figure 1. PD chart showing x_t (* connected by dashed lines) with upper and lower specification limits and a target, together with the posterior distributions of $\mu_t \mid \mathbf{x}^t$ for the autobody data. The horizontal scaling of the posteriors is for visual convenience.

are plotted in Figure 1. The target and specification limits are $T = 1.0$, $LSL = 0.3$, and $USL = 1.7$. Notice that some of the measurements in Figure 1 are negative. This is an artifact of the choice for the origin of the measurement coordinate system, which has little physical significance.

We assume the RJ model of Equation (2) for μ_t and a normal distribution for ε_t. We chose $p = 0.05$, $\eta = 4\sigma$, and $\sigma = 0.27$ as parameters of the model (Section 5 discusses choice/estimation of parameters), and we chose a starting distribution $g^{\mu}_{0|0}(\bullet) = N(1.02, 0.05^2)$. For the RJ model with normal ε_t, $h(\varepsilon) = \sigma^{-1}\phi(\varepsilon/\sigma)$ and $p(\mu \mid \nu) = \tilde{p}(\mu|\nu) + (1-p)\delta(\mu-\nu)$, where $\tilde{p}(\mu|\nu) = p\eta^{-1}\phi((\mu-\nu)/\eta)$, and $\phi(\bullet)$ and $\delta(\bullet)$ denote the standard normal density function and the Dirac delta function, respectively. The delta function component is a convenient notation for indicating that $p(\mu \mid \nu)$ has mass $1-p$ at $\mu = \nu$ and for allowing us to write the integral in Equation (4) without distinguishing between discrete and continuous random variables. Since the Dirac delta function is infinitesimally narrow, it cannot be directly included in $p(\mu \mid \nu)$ when performing the numerical integration in Equation (4). Instead, one should use $g^{\mu}_{t|t-1}(\mu) = \int \tilde{p}(\mu|\nu)\, g^{\mu}_{t-1|t-1}(\nu)\, d\nu + (1-p)\, g^{\mu}_{t-1|t-1}(\mu)$. The integral of Equation (4) should be computed in a like manner whenever there is a point mass in $p(\mu \mid \nu)$.

Figure 1 also shows the posteriors $g^{\mu}_{t|t}(\bullet)$ for the process mean at each of the 30 observations. It appears that for the first 15 observations, the mean is close to the target, but that a substantial downward shift occurred around observation 16. The behavior of $g^{\mu}_{t|t}(\bullet)$ immediately after the mean shift is quite interesting. Because x_{16} fell far below the support for the mean at the previous observation 15, $g^{\mu}_{16|16}(\bullet)$ is bimodal with one mode near the mode for the previous mean ($\mu \approx 1.05$) and the other mode near x_{16}. This is due to the nature of the RJ model. The same is true for $g^{\mu}_{17|17}(\bullet)$, except much more mass is allocated to the mode around x_{16} and x_{17}. At this point, the posterior probability that the mean has changed is quite high. The algorithm has "second thoughts" at observation 18 (i.e., the mode near $\mu \approx 1.0$ gains mass) because x_{18} happens to be quite high. After the next few observations are obtained, which are all well below 1.0, $g^{\mu}_{t|t}(\bullet)$ evolves into a relatively stable distribution with posterior mean around $\mu \approx 0.55$, which is closer to the *LSL* than to the target. This, together with Figure 4 discussed later, would support arguments for process engineers to make a shim move to bring the mean back on target. After adjusting the process, one would begin the PD chart anew with a new estimate of σ if the adjustment also caused the standard deviation to change.

Similarly, the quite low value for x_{29} also caused a smaller but discernible second mode (centered around x_{29}) to begin to develop in $g^{\mu}_{29|29}(\bullet)$. However, because this appears to be a single outlier observation, and x_{30} is back near the average of the observations between $16 \le t \le 28$, this second mode is not carried over to $g^{\mu}_{30|30}(\bullet)$.

When comparing $g^{\mu}_{t|t}(\bullet)$ to the specification limits, one should keep in mind that *LSL* and *USL* are specification limits on x, not on μ. However, we think that adding specification limits to the plot helps to put the location and spread of $g^{\mu}_{t|t}(\bullet)$ into perspective and also makes it easier to visually compare this chart to charts showing $g^{x}_{t+1|t}(\bullet)$ (e.g., Figure 4, discussed later).

Another word of caution is regarding the interpretation of the PD chart. Let N denote the time index for the most recent observation ($N = 30$ in Figure 1). Recall that the posterior $g^{\mu}_{t|t}(\bullet)$ for $t < N$ is conditioned on only the data $\mathbf{x}^t$ and not on the entire data $\mathbf{x}^N$. Hence, the last posterior $g^{\mu}_{N|N}(\bullet)$ is the only one that is conditioned on all the data. If one calculated and plotted the posterior of $\mu_t \mid \mathbf{x}^N$ for $t = 1, 2, \ldots, N$, then the posteriors around the middle of the data would almost certainly indicate an abrupt shift at $t = 16$ with much higher probability than is implied by $g^{\mu}_{16|16}(\bullet)$ in Figure 1. This would be preferable from the viewpoint of retrospectively analyzing the entire dataset. However, we avoid this for two reasons: First, the PD chart is meant to be used in Phase II and updated as each new observation is obtained, with primary focus on the state of the process

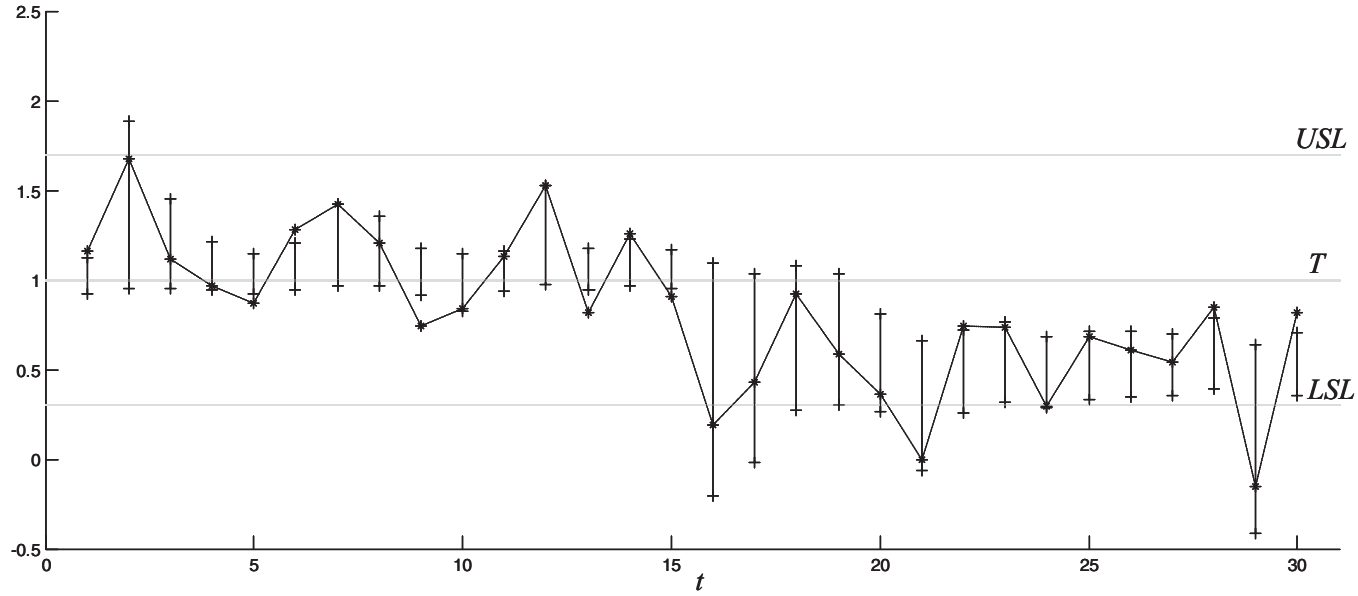

Figure 2. PD chart showing x_t (* connected by dashed lines) and the upper and lower 0.025 quantiles (+ marks connected by vertical lines) of the posterior distribution of $\mu_t \mid \mathbf{x}^t$ for the autobody data.

at current time N. Second, it would be substantially more complicated to update the posterior of $\mu_t \mid \mathbf{x}^N$ for $t = 1, 2, \ldots, N$, which is proportional to $f_{\mathbf{X}^{t,N} \mid \mu t}(\mathbf{x}^{t,N} \mid \mu) g^{\mu}_{t|t}(\mu)$, where $\mathbf{x}^{t,N} = \{x_{t+1}, x_{t+2}, \ldots, x_N\}$. The term $f_{\mathbf{X}^{t,N} \mid \mu_t}(\mathbf{x}^{t,N} \mid \mu)$ would generally be intractable. Crowder and Eshleman (2001) discussed updating $\mu_t \mid \mathbf{x}^N$ for the special case of a Gaussian RW model for μ_t, in which case $f_{\mathbf{X}^{t,N} \mid \mu_t}(\mathbf{x}^{t,N} \mid \mu)$ is jointly Gaussian.

If N is large, it can be visually cumbersome to plot each $g^{\mu}_{t|t}(\bullet)$ for $t = 1, 2, \ldots, N$. An alternative in this case is to plot certain quantiles of $g^{\mu}_{t|t}(\bullet)$ for most t and only plot the full posterior for a few t (e.g., $t = N$). Figure 2 shows the upper and lower 0.025 quantiles for the $g^{\mu}_{t|t}(\bullet)$ shown in Figure 1, and Figure 3 (discussed later) shows similar quantiles with the full distributions shown only for $t = 30$. The quantile interval can be viewed as a two-sided $1-\alpha$ Bayesian credible region (Bernardo and Smith 2002) for μ_t with $\alpha = 0.05$. Crowder and Eshleman (2001) also plotted quantiles of the posterior of μ_t without plotting the full distribution. Because they only considered the Gaussian RW model for μ_t, for which the posteriors of μ_t are also Gaussian, little information was lost by plotting only the quantiles. However, for the RJ model, plotting only the quantiles can be misleading. Consider the posteriors for time indices 16,

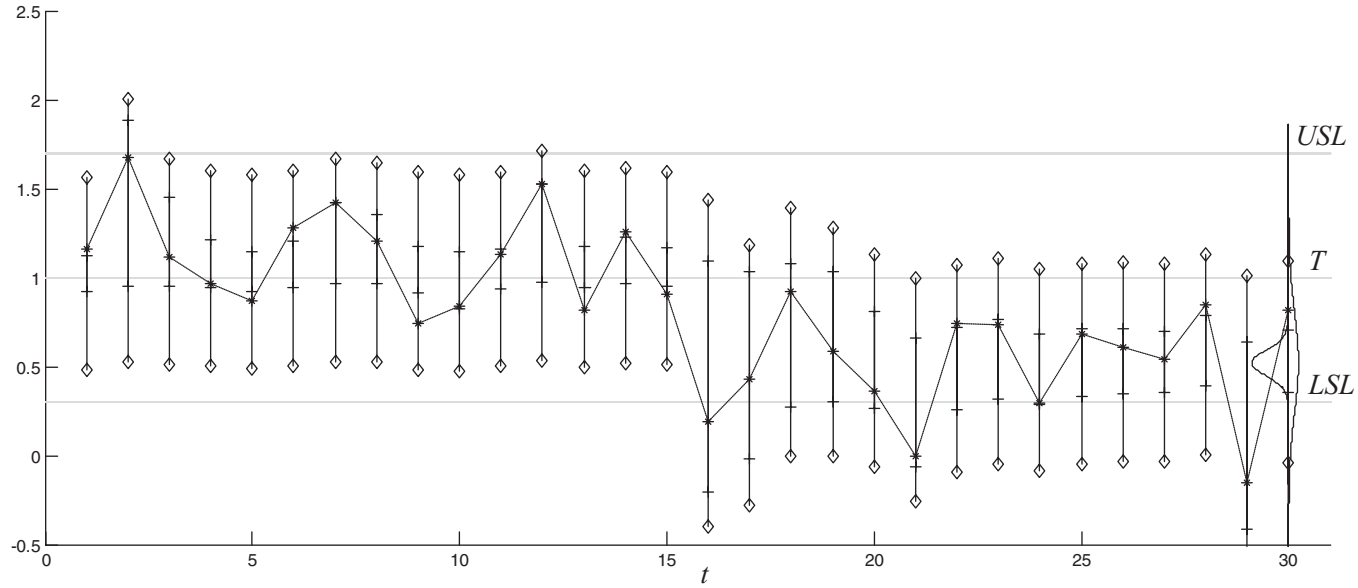

Figure 3. PD chart showing x_t (* connected by dashed lines) and the posterior distributions of $\mu_t \mid \mathbf{x}^t$ and of $x_{t+1} \mid \mathbf{x}^t$ for the autobody data. The full posteriors are shown for $t = 30$, and the upper and lower 0.025 quantiles (+ marks for $g^{\mu}_{t|t}(\bullet)$ and $\diamondsuit$ marks for $g^{x}_{t+1|t}(\bullet)$) are shown for each t. $g^{\mu}_{30|30}(\bullet)$ is plotted to the left of the vertical line at $t = 30$, and $g^{x}_{31|30}(\bullet)$ to the right.

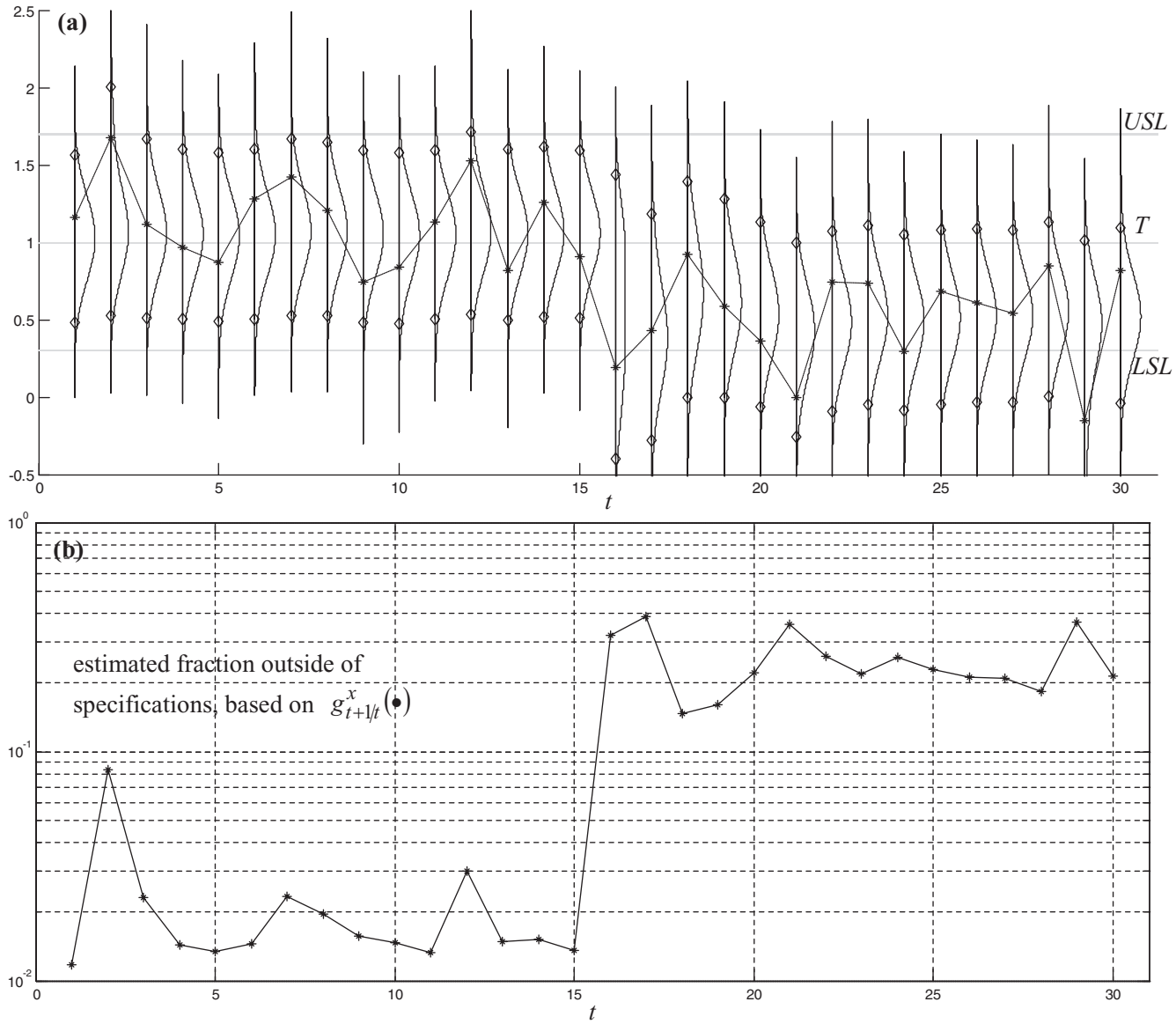

Figure 4. Estimated fraction outside of specifications for the autobody data: (a) PD chart showing x_t (* connected by dashed lines) and the posterior predictive distribution of $x_{t+1} \mid \mathbf{x}^t$ with its upper and lower 0.025 quantiles (◊ marks); (b) estimated fraction outside of specification limits, obtained by integrating $g^x_{t+1|t}(x)$ over values of x outside the specification limits.

17, and 18 plotted in Figure 1. Because these are bimodal, there is very small probability that the mean is a mid-range value (e.g., $\mu \approx 0.8$ or 0.9). This fact is obscured in Figure 2.

There are a number of alternatives that would less obscure multimodal characteristics, while avoiding the visual clutter of plotting entire posterior distributions. One possibility is to display box plots, which Hoadley (1981) considered for a related problem involving the parameter of Poisson data. Another possibility is to calculate a highest posterior density (HPD) credible region for μ_t, instead of using the interval between the upper and lower $\alpha/2$ quantiles. This would reveal multimodal characteristics if the posterior density between modes drops to sufficiently small values. Perhaps the most attractive approach for a commercial software implementation would be to plot credible regions for μ_t, but have the full posteriors pop up if the user drags the cursor over a particular quantile. For example, after inspecting Figure 2, a user might conclude that the posteriors at time indices 16–18 are so dispersed that they bear further scrutiny.

In general, it is useful to see the posterior distribution of $\mu_t \mid \mathbf{x}^t$ relative to the specification limits and the target. In terms of understanding the impact of a mean shift on process capability over the near future production, it is also useful to compare the posterior predictive distribution of $x_{N+k} \mid \mathbf{x}^N$ (denoted by $g^x_{N+k|N}(\bullet)$) to the specification limits for certain $k \geq 1$. Figure 3 shows this for $k = 1$ and under the condition that $\mu_{N+1} = \mu_N$. A visual inspection gives one an approximate idea of the fraction of production that would fall out-of-specifications if the process were to continue to operate with no further change in the mean. The condition that there is no further mean change is not in strict agreement with the Markov model for the mean, but it could be nonetheless useful for interpretation purposes. Based on $g^x_{31|30}(\bullet)$ in Figure 3, for example, the estimated fraction below the *LSL* is 22.5%, which would strengthen the argument

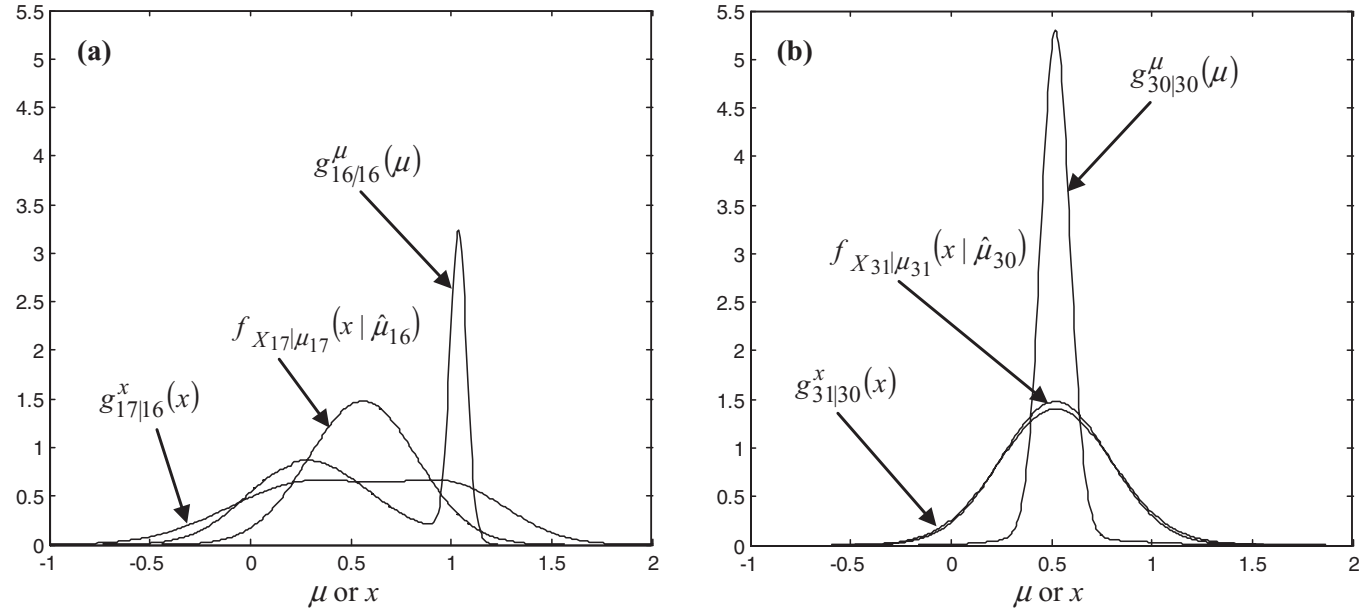

Figure 5. At time indices $t = 16$ (for panel (a)) and $t = 30$ (for panel (b)), plots of $g^{\mu}_{t|t}(\mu)$, $g^{x}_{t+1|t}(x)$, and $f_{X_{t+1}\,|\,\mu_{t+1}}(x\,|\,\hat{\mu}_t)$. At $t = 16$, the large uncertainty in μ_{16} inflates the dispersion of $g^{x}_{t+1|t}(x)$.

for process engineers to make a shim move to bring the mean back on target.

One could also chart any of a number of statistics related to process capability, based on the distribution $g^{x}_{t+1|t}(\bullet)$. Because $g^{x}_{t+1|t}(\bullet)$ may be quite far from being normally distributed (see Figure 5(a), discussed below), process capability indices like C_p and C_{pk} could be misleading. Instead, it might be preferable to directly chart the estimated fraction that falls outside of the specification limits. This is obtained by integrating $g^{x}_{t+1|t}(x)$ over the range of x values outside the specification limits. Figure 4 charts the estimated fraction outside of specifications in conjunction with $g^{x}_{t+1|t}(x)$, for the autobody data.

When interpreting $g^{x}_{t+1|t}(\bullet)$, one should keep in mind that its dispersion reflects any uncertainty in the mean, as well as true process variation. By "true process variation," we mean the dispersion of $h(\varepsilon)$, the conditional distribution of $x_{t+1} \mid \mu_{t+1}$. This widening of $g^{x}_{t+1|t}(\bullet)$ due to uncertainty in the mean is most extremely following a sudden mean shift, such as at time index 16 in Figures 1–4, at which time there is much uncertainty in where μ_t lies. Figure 5(a) better illustrates this by plotting the following for $t = 16$: $g^{\mu}_{t|t}(\mu)$, $g^{x}_{t+1|t}(x)$ (again under the assumption that $\mu_{t+1} = \mu_t$), and $f_{X_{t+1}\,|\,\mu_{t+1}}(x\,|\,\hat{\mu}_t) = h(x-\hat{\mu}_t)$. Here, $\hat{\mu}_t$ denotes the mean of $g^{\mu}_{t|t}(\mu)$, which can be viewed as a point estimate of μ_t. Figure 5(b) shows the same, except at time index $t = 30$. Note that $\hat{\mu}_{16} = 0.56$, and $\hat{\mu}_{30} = 0.52$. From Figure 5(a) it appears that at time index 16, as much of the dispersion of $g^{x}_{17|16}(x)$ stems from the uncertainty in μ_{16} as from the conditional variance of $x_{17} \mid \mu_{17}$. In contrast, Figure 5(b) shows that at time index 30, most of the dispersion of $g^{x}_{31|30}(x)$ comes from the conditional variance of $x_{30} \mid \mu_{30}$, because the uncertainty in μ_{30} is relatively small. If one plots $g^{x}_{t+1|t}(\bullet)$ in the PD chart, one could also plot the conditional distribution of $x_{t+1} \mid \mu_{t+1}$ as a reference, to help convey how much of the dispersion of $g^{x}_{t+1|t}(\bullet)$ is due to uncertainty in the mean.

Another numerical summary of the posterior distributions at each time t that might be relevant is the posterior probability that the mean differs from the target by more than some specified value c, where c is chosen to represent the smallest mean difference that is of practical interest. In the autobody example, the smallest shim thickness is 0.1 mm. Hence, one might choose $c = 0.1$ in this case and adjust the process mean via a shim move only if the posterior probability that $\mu_t \notin [T - c, T + c]$ is sufficiently large.

The posterior probability can be obtained by numerically integrating $g^{\mu}_{t|t}(\bullet)$ outside the interval $[T - c, T + c]$. A procedure such as this is reminiscent of algorithmic process adjustment schemes. However, we do not envision this being used as part of a formal algorithm for dictating when to make adjustments. As discussed in the introduction, there are often many other considerations—quantitative and qualitative—that must be weighed when deciding whether to make an adjustment. For example, even if there is a high probability that $\mu_t \notin [T - c, T + c]$, one might assign very low priority to making an adjustment if the estimated fraction outside of specification limits (assessed via something like Figure 4) is negligibly small.

4. OTHER MARKOV MODELS FOR μ_t

In the examples up to this point, we have focused on processes that are iid with occasional abrupt shifts in the mean, which can be represented by the RJ model. However, the general Markov model for μ_t assumed in Equations (3)–(5) can accommodate many types of wandering or autocorrelated μ_t, with or without random jumps. For example, a first-order autoregressive model for μ_t is a Markov model. Likewise, if μ_t is modeled as the sum of a first-order autoregressive process plus a random jump component, the result is still a Markov model. A proper treatment

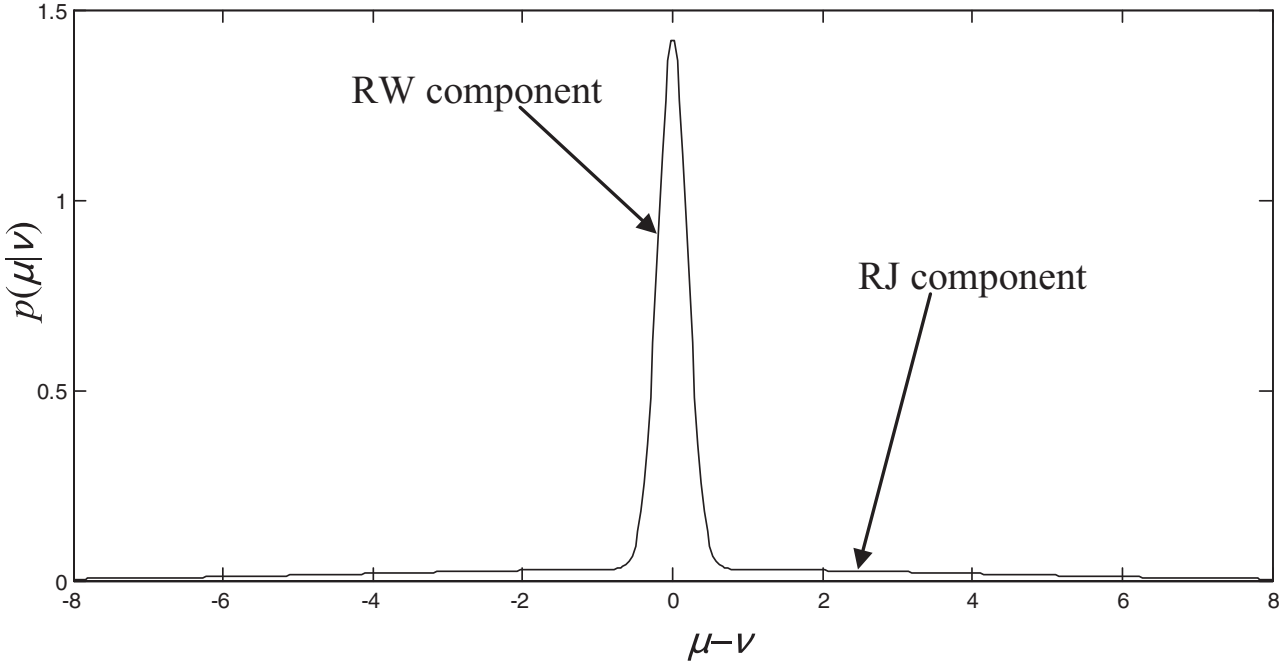

Figure 6. Plot of $p(\mu \mid \nu) = p(\mu - \nu)$ for the RJ+RW model.

of autocorrelated processes is beyond the scope of this article. Instead, in this section, we focus on some simple models related to the RJ and RW models for μ_t.

4.1 Random Jump Plus Random Walk (RJ+RW) Model

In the random jump plus random walk (RJ+RW) model, we represent μ_t as the sum of a RJ component (with small p) and a RW component via

$$\mu_t = \begin{cases} \mu_{t-1} + N\left(0, \beta^2\right): & \text{with probability } 1-p \\ \mu_{t-1} + N\left(0, \beta^2\right) + N\left(0, \eta^2\right): & \text{with probability } p \end{cases}, \tag{6}$$

where β is typically small (e.g., $\beta/\sigma < 0.2$), and η is typically large (e.g., $\eta/\sigma > 3$). The interpretation is as follows: At any fixed point in time, with some small probability p the mean will experience a random jump that is perhaps quite large (η represents the standard deviation of the jump magnitude). If no random jump occurs at that point in time, then the mean takes a small step as would a RW (β represents the standard deviation of the RW increments). The transition probability density for the RJ+RW model is

$$p(\mu|\nu) = \frac{p}{\sqrt{\beta^2+\eta^2}}\phi\left(\frac{\mu-\nu}{\sqrt{\beta^2+\eta^2}}\right) + \frac{1-p}{\beta}\phi\left(\frac{\mu-\nu}{\beta}\right). \tag{7}$$

Figure 6 is a plot of $p(\mu \mid \nu)$ for the RJ+RW model with $p = 0.3$, $\eta = 4$, and $\beta = 0.2$. Notice that for the RJ+RW model, $p(\mu \mid \nu)$ depends on μ and ν only via their difference $\mu - \nu$. We chose an overly large value of $p = 0.3$ for Figure 6 so that the RJ component of the density would be discernible in the plot. The RJ and RW components of $p(\mu \mid \nu)$ are defined as the first and second terms of Equation (7), respectively. For the RJ model, $p(\mu \mid \nu)$ is the same as shown in Figure 6, except that the RW component is replaced by a point mass at $\mu - \nu = 0$.

4.2 Fixed-Size RJ+RW Model of T&H (2005)

We refer to the model by Tsiamyrtzis and Hawkins (2005) as the T&H model. It is similar to the RJ+RW model, except that the size(s) γ of the random jump is assumed known and is a prespecified parameter of the algorithm. Mathematically, the T&H model is $x_t = \mu_t + \varepsilon_t$ with $\varepsilon_t \sim N(0,\sigma^2)$ and iid and a Markov model

$$\mu_t = \begin{cases} \mu_{t-1} + N\left(0, \beta^2\right): & \text{with probability } 1-p \\ \mu_{t-1} + N\left(0, \beta^2\right) + \gamma: & \text{with probability } p \end{cases},$$

for the mean. To allow for more than one size jump, one can specify a set of γ values and a corresponding set of p-values. For example, using $p = \{0.005, 0.005\}$ and $\gamma = \{1, -3\}$ means that the probability of experiencing a random jump at any time t is 0.01, and this probability is equally divided between a positive jump of magnitude 1.0 and a negative jump of magnitude 3.0.

The T&H method evaluates the posterior distributions analytically, as opposed to the numerical evaluation of this article. For a single specified jump size γ, the posterior $g^{\mu}_{t|t}(\bullet)$ is a mixture of 2^t normal distributions. For two or more specified jump sizes, the number of component distributions in the mixture further explodes. T&H recommended using one or two jump sizes and truncating the number of component distributions in the mixture to only the most probable. For any number of jump sizes, however, the T&H model is a special case of the Markov model that we consider. Consequently, the T&H method can be implemented exactly (aside from any inaccuracies due to the discretization used in the numerical integration, which can easily be rendered negligible) with quite moderate computational expense using the method of this article. Retaining the weights for the mixture components in the analytical implementation of

T&H can be useful for diagnosing when the shift occurred. The plots discussed in Section 3 can also serve this purpose.

4.3 RJ+RW Model of Tsiamyrtzis and Hawkins (2007)

Tsiamyrtzis and Hawkins (2007) extended the T&H model in a number of ways. First, they removed the restriction of a fixed jump size. Instead, they represented the unknown jump size γ using an $N(\gamma_0,\eta^2)$ normal distribution as in the RJ and RJ+RW models of Equations (2) and (6), but with a nonzero prior mean γ_0 allowed in case one knows that shifts in one direction are more likely. Second, instead of choosing a fixed p, they used a beta prior distribution for p (independent of γ) and then marginalized all posterior distributions with respect to the prior for p. Third, instead of updating the posterior of $\mu_t \mid \mathbf{x}^t$, they update the joint posterior of $(\mu_t, \sigma^2) \mid \mathbf{x}^t$. They assume the ratios η/σ and β/σ are known. Although they update the posterior of σ^2, it is not for the same tracking purposes as is the updating of the posterior of μ_t, because they assume σ remains constant over time (in contrast to the time-varying μ_t). Their primary purpose for updating the posterior of $\sigma^2 \mid \mathbf{x}^t$ is to provide a type of self-starting feature that avoids having to estimate σ from a set of Phase I data. In contrast, our objective is not a self-starting procedure, and we assume that one can estimate σ with sufficient precision from a set of Phase I data, as described in Section 5.

Although Tsiamyrtzis and Hawkins (2007) assumed a "prior" for p, they did not use this to produce an updated posterior for $p \mid \mathbf{x}^t$. Instead, they marginalized what is their equivalent of $p(\mu \mid \nu)$ with respect to the prior of p in the same way that $p(\mu \mid \nu)$ is marginalized with respect to the normal distribution for γ. For a fixed, specified σ (i.e., a point mass posterior for σ^2, instead of the actual posterior calculated by Tsiamyrtzis and Hawkins (2007)) and $\gamma_0 = 0$, it is straightforward to show that this is equivalent to using the RJ+RW $p(\mu \mid \nu)$ in Equation (7) with p replaced by its marginal prior mean. Consequently, we refer to the Tsiamyrtzis and Hawkins (2007) approach as using the RJ+RW model. However, we reiterate that the Tsiamyrtzis and Hawkins (2007) algorithm is somewhat different than the RJ+RW algorithm in this article in that they also updated the posterior of σ^2 for the self-starting (as opposed to adaptive tracking) purposes discussed in the preceding paragraph and then marginalized other quantities with respect to the posterior of σ^2.

4.4 Marginalized RJ Model

The RJ model is an attractive model for many manufacturing processes, considering the large body of traditional control charting work that has assumed process data are iid with occasional abrupt mean shifts. However, one must specify p and η, and performance depends on the choices for these parameters, as we illustrate with the following example. Figure 7 shows 60 simulated observations that follow a normal distribution with $\sigma = 1$. The mean is zero for the first 20 observations. A small positive mean shift (of size 1σ) was added at $t = 21$, and a larger negative mean shift (of size 3σ) was added at $t = 41$. Figure 8(a) and 8(b) compare the PD chart behavior for RJ models with small p and large η ($p = 0.01$ and $\eta = 4\sigma$) versus for large p and small η ($p = 1$ and $\eta = 0.15\sigma$, which reduces to a RW model), respectively. The PD chart with small p and large η in Figure 8(a) reacts quickly to the large mean shift at $t = 41$. By observation 43, the distribution of $\mu_t \mid \mathbf{x}^t$ is nearly centered on the new mean. In contrast, the PD chart with large p and small η in Figure 8(b) reacts very slowly to the large mean shift. Even by observation 55, the distribution of $\mu_t \mid \mathbf{x}^t$ is still not centered on the new mean. The relative performance is reversed for tracking the small mean shift that was introduced at $t = 21$: The PD chart with large p and small η does a better job of tracking the small shift. One can see in Figure 8(b) that the distribution of $\mu_t \mid \mathbf{x}^t$ is

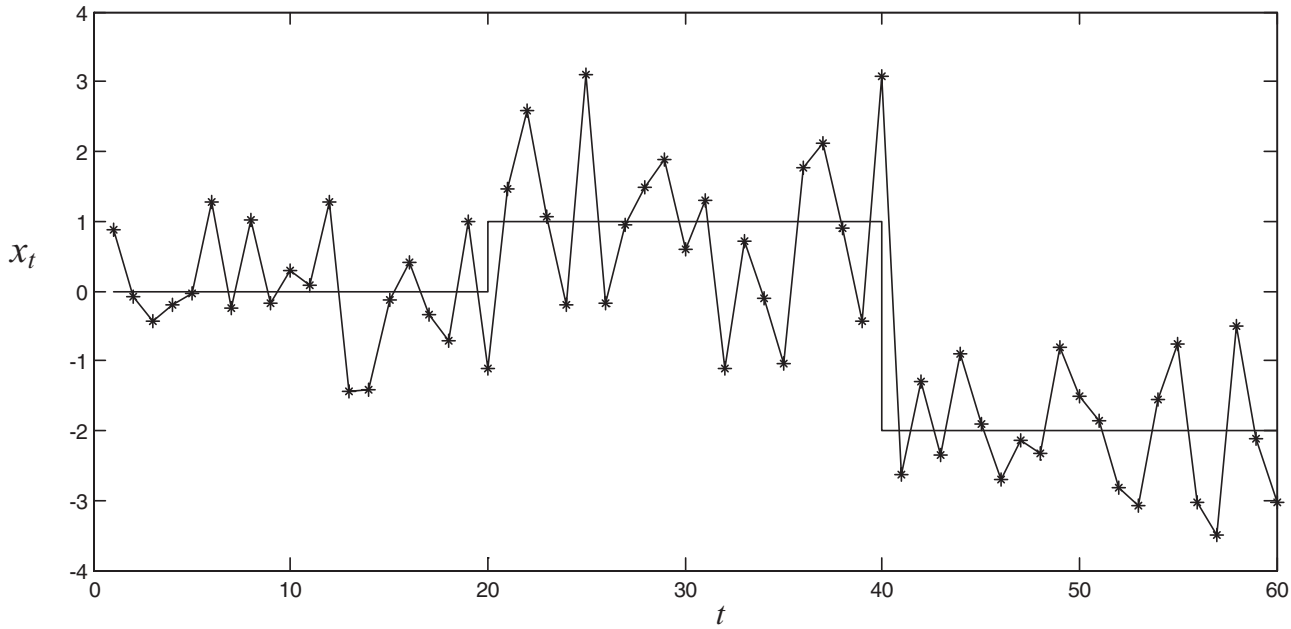

Figure 7. A set of simulated observations with a small 1σ mean shift at $t = 21$ and a large -3σ mean shift at $t = 41$. The solid horizontal lines indicate the true process mean. PD charts for these data for different values of p and η are compared in Figure 8.

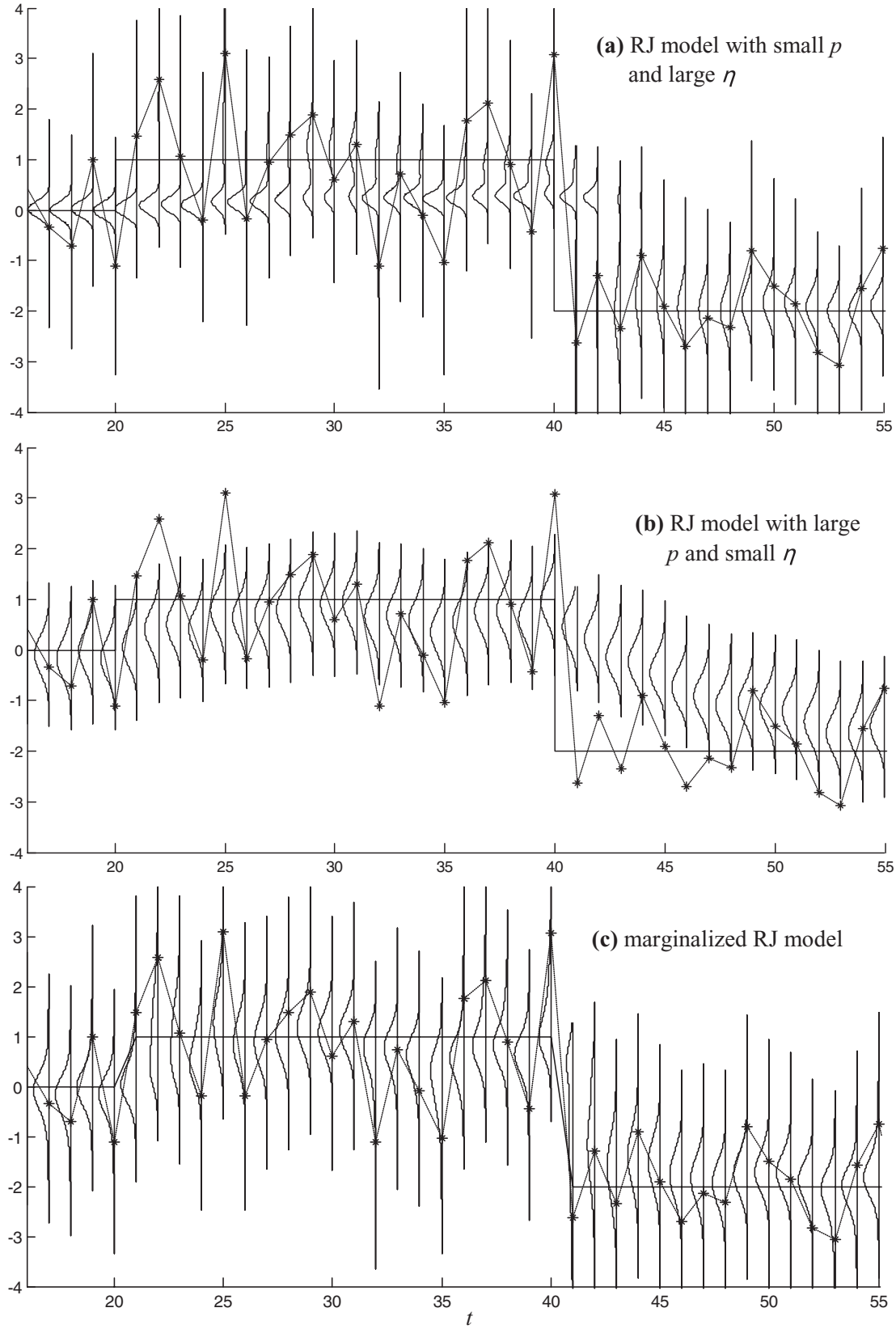

Figure 8. PD charts showing x_t (* connected by dashed lines) from Figure 7 and the posterior distributions of $\mu_t \mid \mathbf{x}^t$ for time indices 17–55. Three different Markov mean models were used: (a) RJ model with $p = 0.01$, $\eta = 4\sigma$; (b) RJ model with $p = 1.0$, $\eta = 0.15\sigma$; and (c) marginalized RJ model with $\{(\alpha_j, \eta_j)\colon j = 1,2,3\} = \{(0.01, 4), (0.1, 1), (0.25, 0.2)\}$. The solid horizontal lines indicate the true process mean.

nearly centered on the new mean by observation 28. In contrast, even at observation 40, the distribution of $\mu_t \mid \mathbf{x}^t$ for small p and large η in Figure 8(a) remains centered well below the new mean.

An alternative to specifying a single (p, η) is to assume some joint distribution for (p, η) that has mass both at (small p, large η) values and at (large p, small η) values. For simplicity, suppose the distribution is discrete with mass w_j at (p_j, η_j) for J distinct combinations $\{(p_j, \eta_j): j = 1, 2, \ldots, J\}$. Marginalizing $p(\mu \mid \nu,p,\eta) = p\eta^{-1}\phi((\mu - \nu)/\eta) + (1 - p)\delta(\mu - \nu)$ for the RJ model with respect to the joint distribution of (p_j, η_j) gives

$$p(\mu|\nu) = \sum_{j=1}^{J} w_j \, p(\mu|\nu, p_j, \eta_j) = \sum_{j=1}^{J} \alpha_j \frac{1}{\eta_j}\phi\left(\frac{\mu - \nu}{\eta_j}\right) + \alpha_{J+1}\,\delta(\mu - \nu), \quad (8)$$

which is a mixture of J zero-mean normal distributions, the jth of which has standard deviation η_j and weight $\alpha_j = w_j p_j$, plus a point mass $\alpha_{J+1} = 1 - \sum_{j=1}^{J} \alpha_j$ at $\mu - \nu = 0$ (which corresponds to the no-jump situation $\mu_t = \mu_{t-1}$). This is equivalent to modifying the RJ model of Equation (2) so that $\mu_t = \mu_{t-1} + N(0,\eta_j^2)$ with probability α_j $(j = 1, 2, \ldots, J+1)$, where $\eta_{J+1} = 0$ corresponds to no jump. For $J = 2$ with $\eta_1^2 = \eta^2 + \beta^2$, $\eta_2^2 = \beta^2$, $\alpha_1 = p$, and $\alpha_2 = 1 - p$, this reduces to the RJ+RW model of Equation (7).

The marginalization in Equation (8) has a different effect than the marginalization with respect to p in the Tsiamyrtzis and Hawkins (2007) approach. In the latter, it is straightforward to show that the marginalization with respect to p serves only to replace p by $E[p]$ in Equation (7). In addition, the further level of marginalization in Equation (8) with respect to the standard deviation η of the random jump is not present in the Tsiamyrtzis and Hawkins (2007) approach.

Figure 8(c) shows the results for the marginalized RJ model with $J = 3$ and $\{(\alpha_j, \eta_j): j = 1,2,3\} = \{(0.01, 4), (0.1, 1), (0.25, 0.2)\}$, which represent large $(\eta = 4\sigma)$, medium $(\eta = \sigma)$, and small $(\eta = 0.2\sigma)$ sized shifts, respectively. In this case, the PD chart does an excellent job of tracking both the small shift and the large shift. We further demonstrate this using Monte Carlo simulation in Section 6.

5. SELECTION OF PARAMETERS AND PHASE I ANALYSIS

To implement PD charts, one must specify a number of quantities: The Markov transition function $p(\bullet \mid \bullet)$ that defines the behavior of μ_t; the standard deviation of ε_t in the normal case (more generally, the distribution $h(\bullet)$ of ε_t); and an initial distribution $g^{\mu}_{0|0}(\bullet)$ for the mean. Unless one requires accurate estimation of $g^{\mu}_{t|t}(\bullet)$ at very early t, choice of $g^{\mu}_{0|0}(\bullet)$ is not critical. As t increases, the choice of $g^{\mu}_{0|0}(\bullet)$ has little influence on the posterior $g^{\mu}_{t|t}(\bullet)$ because of the sequential nature of the approach. Consequently, we recommend a relatively diffused distribution for $g^{\mu}_{0|0}(\bullet)$. Regarding $p(\bullet \mid \bullet)$, if one has reasonable knowledge of the distribution of the jump sizes (reflected by η) and jump frequency (reflected by p), then the RJ model can be used. In the absence of such knowledge, we recommend using the marginalized RJ model with $J = 3$ and $\{(\alpha_j, \eta_j): j = 1, 2, \ldots, J\}$ chosen to represent large shifts that occur with low probability (e.g., $\alpha = 0.01$, $\eta = 4\sigma$), moderate shifts that occur with moderate probability (e.g., $\alpha = 0.1$, $\eta = \sigma$), and small shifts that occur more frequently (e.g., $\alpha = 0.25$, $\eta = 0.2\sigma$). As discussed in the previous section, this is equivalent to using the RJ model but marginalizing with respect to a discrete prior distribution on the unknown (p, η) pair. Based on extensive simulations (some of which are shown in Section 6 and in the online Supplementary Materials for this article), we have observed that the preceding choices result in quite good tracking performance for a range of shift sizes, and the performance is also robust to moderate variations in the α_j values.

For the preceding, we recommend against using large α_j in conjunction with large η_j. This would only be appropriate if one expects a drastically unstable process that experiences very frequent and very large jumps, but in this case it would seem more advisable to identify and remove the source of the instability than to track a mean that is known to change by large amounts at nearly every observation. If one did use large α_j in conjunction with large η_j, the posterior distribution of $\mu_t \mid \mathbf{x}^t$ would jump around dramatically from observation to observation.

For certain parameterizations of $h(\bullet)$ and $p(\bullet \mid \bullet)$, one might consider using Bayesian or maximum likelihood methods to estimate all quantities from a set of Phase I data. For example, for Gaussian $h(\bullet)$ and a RJ $p(\bullet \mid \bullet)$, the Markov chain Monte Carlo (MCMC) approach by McCulloch and Tsay (1993) can be used to calculate Bayesian estimates of p, σ, and η, as well as the jump times and magnitudes over the Phase I data. For more general parameterizations, this would be substantially more difficult, and we also caution against it for other reasons. First, given that $x_t = \mu_t + \varepsilon_t$, there are obvious identifiability issues, unless one places restrictions on $p(\bullet \mid \bullet)$ that discourage large, frequent changes in μ_t. With a RJ $p(\bullet \mid \bullet)$, for example, a Bayesian estimation algorithm might have tendency to overfit by estimating $p \approx 1$ and η large, which attributes most of the variability in x_t to μ_t. To avoid this, McCulloch and Tsay (1993) recommended choosing a prior for p that places most of its mass around small values, which assumes that the random jumps in μ_t are infrequent. Alternatively, one could choose a prior that allowed for large values of p, but restrict η to small values. This would allow for a slowly varying μ_t that behaves similarly to the RW model. The second reason we caution against estimating $p(\bullet \mid \bullet)$ from Phase I data is related to the first: Either of the assumptions for ensuring better identifiability (either infrequent jumps in μ_t or slowly varying μ_t), which are natural ones to make in control charting applications, imply that very long segments of data are required to estimate $p(\bullet \mid \bullet)$. Moreover, $p(\bullet \mid \bullet)$ represents variability in the process mean, which, from the standard control charting perspective, is reflective of out-of-control process behavior. If the Phase I data are "in-control" (e.g., within the framework of the RJ model, if μ_t experiences no jumps over the Phase I data), then there is no information available with which to estimate the parameters of $p(\bullet \mid \bullet)$. This difficulty does not exist in standard control charting, unless one wishes to optimize the chart for a particular type of anticipated mean shift, because standard control charts are designed around in-control process behavior.

In light of this, we recommend the marginalized RJ model with the preceding guidelines for $\{(\alpha_j, \eta_j): j = 1, 2, \ldots,$

$J\}$, which takes into account uncertainty in (p, η). We also recommend the following procedure for estimating the error standard deviation σ for the normal $h(\bullet)$. Assume that one has available a set of Phase I data, and let M denote the number of Phase I observations. The procedure is summarized as:

Step 1. Calculate suitable estimates $\{\hat{\mu}_t: t = 1, 2, \ldots, M\}$ of the mean over the Phase I data.
Step 2. Calculate $\{e_t = x_t - \hat{\mu}_t: t = 1, 2, \ldots, M\}$ as estimates of ε_t over the Phase I data.
Step 3. Estimate σ from $\{e_t: t = 1, 2, \ldots, M\}$ using a simple procedure that borrows from Shewhart control charting.

More specifically, if one applies Shewhart $\bar{X}$ and S or R charts (or individual and moving range charts for data collected as individual observations) to the Phase I data, and the data are in-control, then the procedure is straightforward. Because an in-control process implies that the mean is stable, one can set $\hat{\mu}_t$ $(t = 1, 2, \ldots, M)$ equal to the grand average of the Phase I data and take $\{e_t: t = 1, 2, \ldots, M\}$ to be the Phase I observations minus the grand average. Notice that in the context of the marginalized RJ model, an in-control dataset means that either there was no random jump over the duration of the data or the mean has wandered by only a negligible amount (e.g., due to an (α_j, η_j) pair with small η_j but larger α_j) that was insufficient to cause an alarm in the Shewhart chart.

Continuing the previous autobody example, Figure 9 shows Shewhart individual and moving range charts for a set of $M = 92$ individual Phase I observations. The upper and lower control limits (UCL and LCL) and center line (CL) were based on a grand average of 1.02 and an average moving range of 0.309 (see Section 6.4 of ~~by~~ Montgomery 2009, for details on control limit calculation). Because both charts indicate that the process is in-control, we will take $\hat{\mu}_t = 1.02$ for all t. Subtracting this from the data gives $\{e_t: t = 1, 2, \ldots, 92\}$, a histogram and autocorrelation plot of which (not shown for brevity) indicate that the data are close to normal with little autocorrelation. Hence, we have used the normal, iid error model with standard deviation $\sigma = 0.27$, which was estimated as the average moving range divided by the constant $d_2 = 1.128$.

The situation is less straightforward if the Phase I data are out-of-control. First, consider the situation in which the Phase I data are collected according to rational subgrouping principles, in m subgroups, each of size n observations, with an attempt to ensure that μ_t does not change appreciably within most of the subgroups. This is reasonable in the marginalized RJ model if n is small and each subgroup consists of consecutive observations. This assumption is also the basis for the standard Phase I approach for estimating the in-control parameters in Shewhart control charting. The $\bar{X}$ and S (or R) charts constructed for the Phase I data can be used to verify whether the mean was indeed constant over each subgroup, because large mean shifts within a subgroup tend to cause points on the S chart to plot out-of-control. In the approach described in the following paragraph, any mean shifts that occur between subgroups will not adversely affect the estimation of σ. However, a mean shift within a subgroup will inflate the estimate of σ. Hence, any subgroups that plot out-of-control on an S chart (or are otherwise suspected of having a mean shift within the subgroup) should be removed from the analysis.

After removing all subgroups within which the mean may have changed, one can calculate the errors over the Phase I data

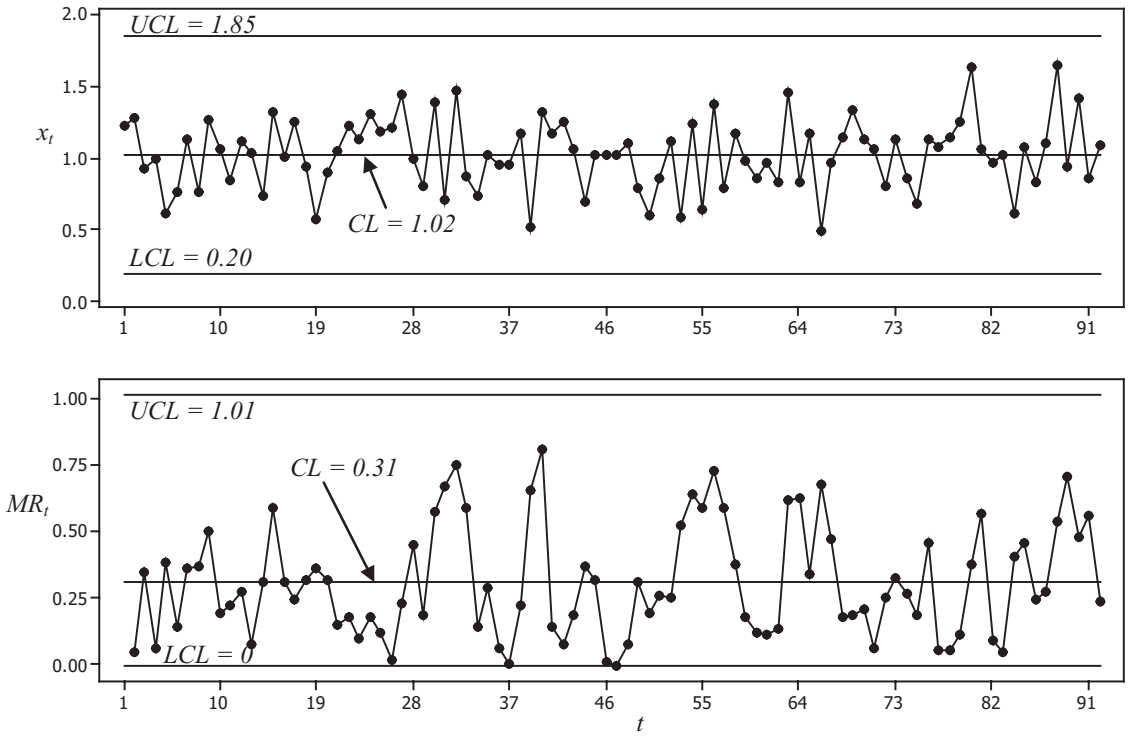

Figure 9. Shewhart individual (top panel) and moving range (bottom panel) charts for a set of Phase I observations for the autobody example. The process is in-control with an estimated mean of 1.02.

via $\{e_{i,j} = x_{i,j} - \hat{\mu}_i : i = 1, 2, \ldots, m; j = 1, 2, \ldots, n\}$, where $x_{i,j}$ and $\hat{\mu}_i$ denote the jth observation from the ith subgroup and the average of the ith subgroup, respectively. An unbiased estimate of the variance of ε_t is

$$\hat{\sigma}^2 = \frac{1}{m\,(n-1)} \sum_{i=1}^{m} \sum_{j=1}^{n} e_{i,j}^2.$$

If the Phase I data were originally collected as individual observations without rational subgrouping, and Shewhart individual and moving range charts indicate that the data were out-of-control, the most appropriate approach for estimating σ is less clear. In this case, we recommend artificially subgrouping the data into m subgroups of n consecutive observations. As in the rational subgrouping situation, n should be small to help ensure that the mean is constant within each subgroup. The remainder of the procedure reduces to that described previously for the rational subgrouping situation.

6. COMPARISON OF BAYESIAN TRACKING METHODS

In this section we use Monte Carlo simulation to compare the performance of various Bayesian methods for tracking the mean. Figure 10 shows box plots of $\hat{\mu}_t$ (the mean of $g^{\mu}_{t|t}(\bullet)$) for five different methods for tracking the same step mean function used in the example of Figure 8 with $\sigma = 1$. The solid line is the true mean $\{\mu_t : t = 1, 2, \ldots, N\}$ over the $N = 60$ time points in the simulation. Comparing the box plots to the true mean provides an indication of the tracking effectiveness of the point estimator $\hat{\mu}_t$ for each method.

The five methods considered are the posterior mean estimators for five different models: (a) the RJ+RW model with $p = 0.01$, $\eta = 4$, $\beta = 0.1$; (b) the RJ model with $p = 0.01$, $\eta = 4$; (c) the marginalized RJ model with $\{(\alpha_j, \eta_j) : j = 1, 2, 3\} = \{(0.01, 4), (0.1, 1), (0.25, 0.2)\}$; (d) the T&H model with $p = \{0.005, 0.005\}$, $\gamma = \{1, -3\}$, $\beta = 0.1$; and (e) the T&H model with $p = \{0.005, 0.005\}$, $\gamma = \{1, -1\}$, $\beta = 0.1$. Recall that the RJ+RW model is nearly identical to the Tsiamyrtzis and Hawkins (2007) model but with a fixed value for σ that is not updated over the Phase II data (see Section 4).

Comparing Figure 10(a)–10(d), the RJ+RW model and the marginalized RJ model perform comparably to the T&H model with $\gamma = \{1, -3\}$. The latter can be viewed as a benchmark for this example, because the true jump sizes were also 1 and -3, giving the T&H model with $\gamma = \{1, -3\}$ somewhat of an unfair advantage for this example. Figure 10(e), which is the T&H model with $\gamma = \{1, -1\}$, illustrates the consequences of incorrectly specifying the jump size(s) in the T&H method. The large negative jump at $t = 41$ is tracked very poorly in this case. The marginalized RJ model reacts more quickly than the RJ+RW model for tracking the smaller mean shift, but slightly less quickly for tracking the large shift. This is not surprising when one considers that the RJ+RW model is a special case of the marginalized RJ model with $J = 2$, $(\alpha_1, \eta_1) = (0.01, 4.01)$, and $(\alpha_2, \eta_1) = (0.99, 0.1)$. Hence, the RJ+RW model is marginalized for very large shifts ($\eta \approx 4\sigma$) and very small shifts ($\eta \approx 0.1\sigma$), but not for moderate shifts.

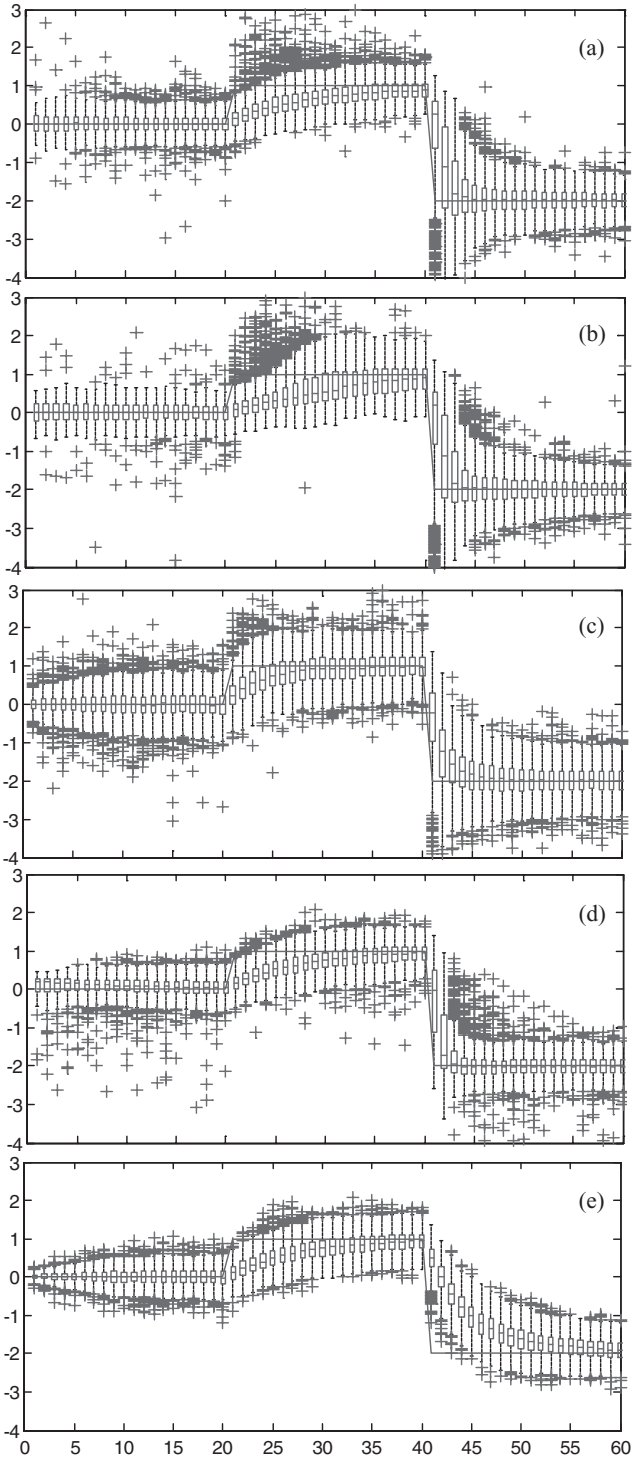

Figure 10. Box plots of $\hat{\mu}_t$ (vertical axis) versus t (horizontal axis) for tracking a step mean (solid line) over 1000 replicates using: (a) RJ+RW model with $p = 0.01$, $\eta = 4$, $\beta = 0.1$; (b) RJ model with $p = 0.01$, $\eta = 4$; (c) marginalized RJ model with $\{(\alpha_j, \eta_j) : j = 1,2,3\} = \{(0.01, 4), (0.1, 1), (0.25, 0.2)\}$; (d) T&H model with $p = \{0.005, 0.005\}$, $\gamma = \{1,-3\}$, $\beta = 0.1$; and (e) T&H model with $p = \{0.005, 0.005\}$, $\gamma = \{1,-1\}$, $\beta = 0.1$. The online version of this figure is in color.

7. CONCLUSIONS

In this article, we have developed a Bayesian approach for graphically monitoring a process mean that is better suited for

informing many engineering decisions than Phase II control charting. The PD chart approach is based on a rather general model (observations = Markov mean + iid errors) and allows one to chart a number of relevant quantities derived from the posterior distributions of the mean ($\mu_t \mid \mathbf{x}^t$) or the future data ($x_{t+k} \mid \mathbf{x}^t$), such as credible regions for the mean, posterior quantiles for future data, the predicted fraction of future data that will fall outside of specifications, etc.

One attractive feature of PD charts is that a target value and process specification limits can be included naturally, allowing one to assess *statistical significance* and *practical significance* in a cohesive manner, using the same chart. By comparing the posteriors of $\mu_t \mid \mathbf{x}^t$ and $x_{t+k} \mid \mathbf{x}^t$ to a target value and specification limits, one can assess whether there is statistically significant evidence that the mean differs substantially from the target and also assess the impact of this on the fraction of production that will fall outside of specifications.

We have advocated numerical integration, rather than closed-form analytical expressions, to update the posteriors. This allows one to choose a completely general Markov model for μ_t and a completely general distribution for ε_t, although we have focused on normally distributed ε_t. Requiring closed-form expressions largely limits one to joint Gaussian distributions. Among other detrimental side effects, this would preclude a RJ model, which is a very attractive one in the context of monitoring a process mean and is perhaps essential for quickly tracking large mean shifts. Resorting to numerical integration provides the flexibility to choose a model such as the marginalized RJ model, which we have demonstrated is quite effective at tracking both large and small shifts.

The flexibility afforded by numerical integration opens up a number of potentially useful avenues that we think warrant further research. One is treatment of a more general error distribution $h(\bullet)$. For highly-skewed error distributions, a PD chart that assumes a Gaussian $h(\bullet)$ would likely produce biased posterior distributions. In addition, the use of a heavy-tailed $h(\bullet)$ would likely provide some level of robustness to occasional outliers, but likely at the expense of lessening the chance of immediate detection of large jumps in the mean. Although the approach presented for updating the posteriors can be used directly with a general $h(\bullet)$ with no increase in computational complexity, the primary difficulty may lie in estimating a nonparametric $h(\bullet)$ from a set of Phase I data.

A second interesting avenue is adaptation of the PD chart for use in the standard Phase II control charting paradigm of signaling an alarm when a mean shift is detected. Perhaps the most common implicitly assumed model in the standard Phase II paradigm is that μ_t is constant until it experiences a jump (of known or unknown size) at some random time, after which it remains constant at the new value. There are many variations of this that could be represented by the Markov mean model with appropriate choice for $p(\bullet \mid \bullet)$. It is possible that with certain choice of $p(\bullet \mid \bullet)$ and of the rule for signaling (based on some characteristic of $g^{\mu}_{t|t}(\bullet)$), the resulting chart could be effective in the standard Phase II control charting paradigm.

A third potentially fruitful avenue is extending the PD chart for use in detecting patterned mean shifts in the standard Phase II control charting paradigm. Certain patterned mean shifts, such as a ramp or a sinusoid, can be represented as a Markov process with a two-dimensional state space. It is conceptually straightforward to extend Equations (3) and (4) from a one-dimensional to a two-dimensional state space. However, the computational expense of numerical integration may be prohibitive in the two-dimensional case. For these and other potential extensions to more complex situations (including tracking other parameters, in addition to the mean), sequential Monte Carlo methods for nonlinear Bayesian filtering (Doucet, Godsill, Andrieu 2000; Doucet and Johansen 2011) may be quite useful.

SUPPLEMENTARY MATERIALS

Source Code: This Supplementary Materials file contains Matlab source code for implementing the PD charts in this article. Code for three different routines are included: (1) for calculating the posterior distributions for the marginalized RJ model; (2) for calculating the posterior distributions for the RJ+RW model; and (3) for plotting the posterior distributions that were calculated using one of the other two routines.

Robustness Discussion: This Supplementary Materials file provides some additional discussion on the robustness of the PD chart tracking performance with respect to choice of hyperparameters $\{(\alpha_j, \eta_j): j = 1,2, \ldots, J\}$.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation under grant #CMMI-0758557. The author gratefully acknowledges the careful reviews and helpful comments from two anonymous Referees, the Associate Editor, and the Editor.

[Received August 2009. Revised December 2011.]

REFERENCES

Apley, D. W., and Shi, J. (2001), "A Factor Analysis Method for Diagnosing Variability in Multivariate Manufacturing Processes," *Technometrics*, 43, 84–95. [293]

Barnard, G. A. (1959), "Control Charts and Stochastic Processes," *Journal of the Royal Statistical Society,* Series B, 21, 239–271. [294]

Bayarri, M. J., and Garcia-Donato, G. (2005), "A Bayesian Sequential Look at u-Control Charts," *Technometrics*, 47, 142–151. [294]

Bernardo, J. M., and Smith, A. F. M. (2002), *Bayesian Theory*, New York: Wiley. [297]

Calabrese, J. M. (1995), "Bayesian Process Control for Attributes," *Management Science*, 41, 637–645. [294]

Chernoff, H., and Zacks, S. (1964), "Estimating the Current Mean of a Normal Distribution Which is Subjected to Changes in Time," *Annals of Mathematical Statistics*, 35, 999–1018. [294]

Colosimo, B. M., and del Castillo, E. (eds.) (2007), *Bayesian Process Monitoring, Control and Optimization*, Boca Raton, FL: Chapman & Hall/CRC. [294]

Crowder, S. V. (1986), "Kalman Filtering and Statistical Process Control," unpublished Ph.D. dissertation, Iowa State University, Ames, IA. [294]

Crowder, S. V., and Eshleman, L. (2001), "Small Sample Properties of an Adaptive Filter Applied to Low Volume SPC," *Journal of Quality Technology*, 33, 29–46. [294,296,297]

del Castillo, E., and Montgomery, D. C. (1995), "Kalman Filtering Process Control Scheme With an Application in Semiconductor Short Run Manufacturing," *Quality and Reliability Engineering International*, 11, 101–105. [294]

Doucet, A., Godsill, S., and Andrieu, C. (2000), "On Sequential Monte Carlo Methods for Bayesian Filtering," *Statistics and Computing*, 10, 197–208. [306]

Doucet, A., and Johansen, A. M. (2011), "A Tutorial on Particle Filtering and Smoothing: Fifteen Years Later," in *The Oxford Handbook of Nonlinear*

Filtering, eds. D. Crisan and B. Rozovsky, New York: Oxford University Press, pp. 656–704. [306]

Downing, D. J., Pike, D. H., and Morrison, G. W. (1980), "Application of the Kalman Filter to Inventory Control," *Technometrics*, 22, 17–22. [294]

Girshick, M. A., and Rubin, H. (1952), "A Bayes Approach to a Quality Control Model," *Annals of Mathematical Statistics*, 23, 114–125. [294]

Hamada, M. (2002), "Bayesian Tolerance Interval Control Limits for Attributes," *Quality and Reliability Engineering International*, 18, 45–52. [294]

Hoadley, B. (1981), "Quality Measurement Plan (QMP)," *Bell System Technical Journal*, 60, 215–274. [298]

Kalman, R. E. (1960), "A New Approach to Linear Filtering and Prediction Problems," *Journal of Basic Engineering*, 82, 35–45. [294]

McCulloch, R. E., and Tsay, R. S. (1993), "Bayesian Inference and Prediction for Mean and Variance Shifts in Autoregressive Time Series," *Journal of the American Statistical Association*, 88, 968–978. [303]

Meinhold, R. J., and Singpurwalla, N. D. (1983), "Understanding the Kalman Filter," *The American Statistician*, 37, 123–127. [294]

Menzefricke, U. (2002), "On the Evaluation of Control Chart Limits Based on Predictive Distributions," *Communications in Statistics – Theory and Methods*, 31, 1423–1440. [294]

Montgomery, D. C. (2009), *Introduction to Statistical Quality Control* (6th ed.), New York: Wiley. [304]

Ross, S. (1971), "Quality Control Under Markovian Deterioration," *Management Science*, 17, 587–596. [294]

Sastri, T., Valdes, J. B., and Flores, B. (1996), "Nonlinear Control Charts for Jump Detection," *International Journal of Production Research*, 34, 1023–1044. [294]

Tagaras, G., and Nikolaidis, Y. (2002), "Comparing the Effectiveness of Various Bayesian Control Charts," *Operations Research*, 50, 878–888. [294]

Taylor, H. M. (1967), "Statistical Control of a Gaussian Process," *Technometrics*, 9, 29–41. [294]

Triantafyllopoulos, K. (2007), "Feedback Quality Adjustment With Bayesian State-Space Models," *Applied Stochastic Models in Business and Industry*, 23, 145–156. [294]

Tsiamyrtzis, P., and Hawkins, D. M. (2005), "A Bayesian Scheme to Detect Changes in the Mean of a Short-Run Process," *Technometrics*, 47, 446–456. [294,300]

——— (2007), "A Bayesian Approach to Statistical Process Control," in *Bayesian Process Monitoring, Control and Optimization*, eds. B. M. Colosimo and E. del Castillo, Boca Raton, FL: Chapman & Hall/CRC Press, pp. 87–107. [294,301,303,305]

Wasserman, G. S., and Sudjianto, A. (1993), "Short-Run SPC Based Upon the Second Order Dynamic Linear Model for Trend Detection," *Communication in Statistics – Simulation and Computation*, 22, 1011–1036. [294]

White, C. C. (1977), "A Markov Quality Control Process Subject to Partial Observation," *Management Science*, 23, 843–852. [294]

Woodward, P. W., and Naylor, J. C. (1993), "An Application to Bayesian Methods in SPC," *The Statistician*, 42, 461–469. [294]