Designing Bayesian EWMA Monitors

Using Gage R & R and Reliability Data

by

Spencer Graves, Søren Bisgaard, and Murat Kulachi

Affiliations

Spencer Graves
Productive Systems Engineering
751 Emerson Ct.
San José, CA 95126
sgraves@prodsyse.com

Søren Bisgaard
Søren Bisgaard and Associates
PO Box 101
2315 B Forest Dr.
Annapolis, MD 21401
s.bisgaard@monitoring-intl.com

Murat Kulahci
Arizona State University
Dept. of Industrial Engineering
P. O. Box 875906
Tempe, AZ 85287
Murat.Kulahci@asu.edu

Designing Bayesian EWMA Monitors Using Gage R & R and Reliability Data

ABSTRACT

We derive an exponentially weighted moving average (EWMA) as the Bayesian prior mean for a random walk observed with normal error. This derivation shows how the weight on the last observation varies over time. This weight depends on the migration rate of the random walk and the noise variance, which can be estimated from reliability data and studies of gage repeatability and reproducibility, respectively. The variations in the weight on the last observation provide a solution to the "fast initial response" problem that we believe is intuitively more satisfying than the current standard solution.

KEY WORDS: Exponentially weighted moving average; Fast initial response (FIR); Bayesian Sequential Updating; Gage repeatability and reproducibility; Hazard rate; On-board diagnostics (OBDs).

1. INTRODUCTION

Our productivity, comfort and health in modern society require the proper functioning of increasing complex systems subject to a growing variety of malfunctions. The increasing computerization of such systems (which we call "plants" for consistency with the control theory literature) provides opportunities for increasingly sophisticated monitors designed to detect and

isolate faults, thereby limiting potential damage and facilitating repairs. Current law requires that all new automobiles sold in the US, Canada and Europe have "on-board diagnostics" (OBDs) to detect and isolate situations that might compromise emission controls. These concepts could also be applied in manufacturing, public health, clinical trials, and monitoring financial figures. In this article, we apply the principle of Bayesian sequential updating (Figure 1) to a random walk observed with error, obtaining thereby a Bayesian exponentially weighted moving average (EWMA) with parameters determined from reliability / hazard rate data and gage repeatability and reproducibility studies. For an application of Bayesian sequential updating to detect an abrupt jump, see Graves, Bisgaard and Kulahci (2002); for more general applications, see Graves et al. (2001).

(Figure 1 about here)

Consider a physical system with a condition x_t that is not directly observable but that is assumed to follow a random walk, as

$$x_{t+1} = m_t + x_t + w_t, \quad w_t \sim N(0, s^2_{w,t}), \tag{1}$$

where m_t represents a potential deterministic drift, and w_t represents an unpredictable portion of system reliability. In Section 2, we show how m_t and $s_{w,t}$ relate to the reliability hazard rate. In particular, we show that any reliability distribution can be modeled in terms of $s_{w,t}$ or (m_t, $s_{w,t}$).

The analyses of this report assume we can model adequately the distribution of x_t at first use, $t = 1$. No monitor is designed for a completely unprecedented application: The time and money to design and use a monitor is justified from experience with other applications, which

can be used to estimate a distribution at first use. For a manufactured product, this could be obtained from control chart data collected at the end of the production line. If this is a new product, the distribution at first use could be estimated from the history of similar products, adjusted if appropriate considering design objectives, prototype test data, and previous new product introductions. In biostatistics, it could be obtained from previous clinical trials of roughly comparable therapies. In portfolio management, one could consider the behavior of similar financial instruments. In this article, we shall assume that $x_1 \sim N(x_{1|0},\ s^2_{1|0})$.

In Section 3, we consider a process of the form (1) (with $m_t = 0$ and $s_{w,t} = s_w$ constant) observed with error,

$$y_t = x_t + v_t, \quad v_t \sim N(0,\ s^2_v). \tag{2}$$

In many situations, s_v can be estimated from a study of gage repeatability and reproducibility (NIST 2001, ch. 2).

With adequate estimates of the distribution at first use, the hazard rate and s_v, we can estimate the relative frequency distribution of the condition of plants at any future point in time, among all plants with comparable observed histories. Subjective probabilities can be used, but objective probabilities are also available, within the limits of estimation precision and the comparability of the initial reference set.

The conceptual framework is outlined in Figure 1, with mathematical details summarized in Table 1. The result is an exponentially weighted moving average (EWMA), except that the weight on the last observation varies with time. When the migration variance is a constant fraction (or multiple) of the observation noise variance, this weight converges to a constant;

except in Section 2, we assume that in (1), $m_t = 0$ and $s^2_{w,t} = s^2_w$, a constant. Expression numbers in Table 1 are keyed to the discussion below.

(Table 1 about here)

This procedure provides a fast initial response (FIR) approach that is different from the FIR technique in the literature and is clearly tied to information obtainable for virtually any monitoring application: the distribution of the condition of the plant at first use, x_1, the reliability, coded in s_w, and the measurement noise s_v. If the condition at first use is non-informative, this theory sets $K_1 = 1$ and has K_t declining monotonically to an asymptote as information about x_t accumulates. We believe this has much greater intuitive appeal than the traditional FIR approach, which essentially asserts that the prior at $t = 1$ is as informative as the prior at any $t > 1$.

This FIR approach is discussed in Section 6 after considering an example in Section 4 and discussing further the asymptotic behavior of the algorithm in Section 5. Questions of robustness are considered in Section 7. Section 8 considers when to declare a malfunction, and a summary discussion of this "normal random walk observed with error" appears in Section 9. The Bayesian EWMA is in fact the simplest Kalman filter, although a Bayesian development differs from the minimum mean square prediction error principle used by Kalman (1960).

2. HAZARD AND MIGRATION RATES

We show here how the migration parameters (m_t, $s_{w,t}$) in (1) determine the reliability distribution, expressed in the hazard rate h_t, and conversely how the hazard rate constrains m_t

and determines $s_{w,t}$ given m_t. Of course, complex systems, whether products, production processes or humans, can be impacted by many different kinds of problems. We assume that h_t is the hazard rate relevant to a process x_t observed indirectly via y_t. Standard techniques in biostatistics and reliability support cause-specific estimation of hazard rates.

To understand the relationship between $(m_t, s_{w,t})$ and h_t, we start by assuming that x_t is good as long as $L \le x_t \le U$. If the distribution at first use is $x_1 \sim N(x_{1|0}, s_{1|0}^2)$, then $h_1 = h_{1,0} + h_{1,1}$, where

$$h_{1,0} = \Phi\left(\frac{L - x_{1|0}}{s_{1|0}}\right),$$

and

$$h_{1,1} = \left[1 - \Phi\left(\frac{U - x_{1|0}}{s_{1|0}}\right)\right]. \tag{3}$$

For a manufactured product, h_1 is the proportion of units with x_1 outside (L, U) or that fail for this reason when the customer first attempts to use them.

Let $F_t(x_t) = F_t(x_t \mid L \le x_t \le U, \text{ for all } t \le t)$ be the cumulative distribution function (cdf) for x_t given that it is not bad and has not previously been bad. Then

$$F_1(x_1) = \begin{cases} 0 & \text{if } x_1 < L \\ \dfrac{\Phi\left(\dfrac{x_1 - x_{1|0}}{s_{1|0}}\right) - h_{1,0}}{1 - h_1} & \text{if } L \le x_1 < U \\ 1 & \text{if } U \le x_1 \end{cases} \tag{4}$$

Starting from (3) and (4), we derive the hazard rate h_t and the cdf for x_t good $F_t(x_t)$ recursive in pieces as follows. First, let $F_{t,0}(x_t)$ be the cdf for x_t good or bad at time t given that it has not previously been bad as

$$F_{t,0}(x_t) = \int \Phi\left(\frac{x_t - x_{t-1} - m_{t-1}}{s_{w,t-1}} \right) dF_{t-1}(x_{t-1}) , \tag{5}$$

t = 2, 3, ... Then the proportions of units too small and too large at time t among those good at $t - 1$ are

$$h_{t,0} = F_{t,0}(L),$$

and

$$h_{t,1} = 1 - F_{t,0}(U).$$

The hazard rate at time t is the sum of those failing both small and large, as

$$h_t = h_{t,0} + h_{t,1}. \tag{6}$$

The distribution of those still good is the truncated distribution from $F_{t,0}$, as

$$F_t(x_t) = \begin{cases} 0 & \text{if } x_t < L \\ \left[\dfrac{F_{t,0}(x_t) - h_{t,0}}{1 - h_t} \right] & \text{if } L \le x_t < U \\ 1 & \text{if } U \le x_t \end{cases}. \tag{7}$$

From (3) - (7), we see that the sequence (m_t, $s_{w,t}$) uniquely determines the hazard rate.

Conversely, if $L \le x_{1|0} < U$ and $m_t = 0$ for all t, then h_t is monotonically increasing in $s_{w,t-1}$ and is 0 when $s_{w,t-1} = 0$. Thus, in this case, h_t uniquely determines $s_{w,t}$. [Expressions (1) and (3) - (7) can be generalized to a multivariate state space by assuming that $\mathbf{x}_t$ is bad if it is outside an acceptance region A and defining F_t in the obvious way for all Borel sets. In this more general setting, the parameters of the transition distributions uniquely determine the hazard

rate. With suitable additional restrictions, the hazard rate can uniquely determine some univariate aspect of the migration distribution.]

Therefore, given data on product reliability or time to onset of an adverse reaction in clinical trials, a reasonably parsimonious model can be built for $(m_t, s_{w,t})$ consistent with available data. This does not require data on a new product or therapy never used before; it only requires data on previously tested products or therapies believed to be comparable.

3. UNIVARIATE BAYESIAN UPDATING AND AN EWMA

Now suppose we have y_t being a noisy observation of the unknowable state of the plant x_t, per (2). We shall apply Bayesian sequential updating to this example with the added simplifications of assuming $m_t = 0$ and $s_{w,t} = s_w$ = constant. We shall find that this gives us an exponentially weighted moving average (EWMA) in the limit for large t with an intuitively satisfying Bayesian answer to the fast initial response (FIR) problem. This approach will be compared to the traditional FIR in Section 6 below, after discussing sample computations in Section 4.

As outlined in Figure 1, Bayesian sequential updating is a two-step process: "1. Observation" and "2. Transition". We find it convenient here to divide step 1 into substeps "1.1. Preparing" and "1.2. Updating". This distinction highlights the fact that "1.1. Preparing" can take place between the previous execution of step 2 and the current "1.2. Updating" step. With stationary systems most of "1.1. Preparing" can be computed offline in advance of the application. With traditional Kalman filtering (and traditional EWMAs), the Kalman gain is often replaced by an asymptotic value, and its transients are ignored. This can help reduce

demands on a real-time microprocessor, allowing in some cases the use of a cheaper microprocessor.

We will preface this development with a brief comment about notation: As indicated with (1) and (2), observations y_t provide information about an unknown state of nature x_t. Just before each observation arrives, our knowledge of x_t is summarized in the prior $(x_t \mid D_{t-1}) \sim N(x_{t|t-1},\ s^2_{t|t-1})$, where $D_{t-1} = \{y_{t-1}, y_{t-2}, \ldots y_1, x_{1|0},\ s^2_{1|0}\}$; at time $t = 1$, this is the distribution at first use. Step 1 in Figure 1 transforms this prior into the posterior $(x_t \mid D_t)$, which we show is normal, $N(x_{t|t},\ s^2_{t|t})$, say. Step 2 then models a transition from x_t to x_{t+1}, and our knowledge then degrades accordingly to $(x_{t+1} \mid D_t) \sim N(x_{t+1|t},\ s^2_{t+1|t})$, which becomes the prior at the next point in time. We now consider specifics of these steps.

Step 1.1. Preparing. We divide step 1.1 further into three substeps: (1.1a) Predictive Distribution, (1.1b) Posterior Variance, and (1.1c) Kalman Gain, as we now explain.

Step 1.1a. Predictive Distribution. We begin by combining $(x_t \mid D_{t-1}) \sim N(x_{t|t-1},\ s^2_{t|t-1})$ with the observation process (2) and integrating out the unknowable x_t to get the predictive distribution as follows:

$$(y_t \mid D_{t-1}) \sim N(f_t,\ s^2_{y|t-1}),$$

where

$$f_t = x_{t|t-1}, \tag{8}$$

since the expected value of the sum in (2) is the sum of the expectations, and

$$s^2_{y|t-1} = s^2_{t|t-1} + s^2_v, \tag{9}$$

since the variance of a sum of uncorrelated random variables is the sum of the variances. We maintain the distinction between f_t and $x_{t|t-1}$ because they serve different functions, as witnessed

by the fact that their associated confidence intervals are different, even though the quantities themselves are in this application numerically identical. This notational distinction makes it easier for us to understand the EWMA and explain its different properties. It also makes it easier for us to generalize this procedure to situations where f_t and $x_{t|t-1}$ may be different.

Step 1.1b. Posterior Variance. Fisher's efficient score [derivative of the log density, $l(...)$] of the posterior is the prior score plus the score from the data:

$$\frac{\partial l(x_t \mid D_t)}{\partial x_t} = \frac{\partial l(x_t \mid D_{t-1})}{\partial x_t} + \frac{\partial l(y_t \mid x_t)}{\partial x_t} = \left[-\frac{x_t - x_{t|t-1}}{s^2_{t|t-1}} \right] + \left[-\frac{x_t - y_t}{s^2_v} \right] \qquad (10)$$

Graves (2002) calls this "Keeping score with Bayes' theorem." From this, we see that the posterior score is linear in x_t and can therefore be written $\left[-\left(x_t - x_{t|t}\right) \Big/ s^2_{t|t} \right]$ for appropriately chosen $x_{t|t}$ and $s^2_{t|t}$. The integral of this gives us the logarithm of the posterior density as a parabola up to an additive constant. Since the support of x_t runs over the entire real line, this proves that the posterior is $N(x_{t|t},\ s^2_{t|t})$.

To determine $s^2_{t|t}$, we take another derivative of (10). This gives us "Bayes' Rule of Information" (Graves 2002) that the posterior (observed) information is the sum of the information from the prior and the data, which in this case is

$$s^{-2}_{t|t} = s^{-2}_{t|t-1} + s^{-2}_v . \qquad (11)$$

As noted by Graves (2002), this holds for the observed information [negative second derivative of log(density)] quite generally, even with non-regular likelihood where it does not hold for the Fisher (expected) information. DeGroot (1970) calls a reciprocal squared scale factor a "precision". However, reciprocal squared scale factors do not necessarily add with Bayes'

theorem except with normal distributions. Therefore, to emphasize the generality of (11), we shall use the term "information" for the normal reciprocal variance in this article.

Step 1.1c. Kalman Gain. We now let $x_t = 0$ in (10) and solve for $x_{t|t}$ as follows:

$$x_{t|t} = s_{t|t}^2 \left\{ s_{t|t-1}^{-2} x_{t|t-1} + s_v^{-2} y_t \right\}, \tag{12}$$

The weight on the last observation y_t in (12) is called the Kalman gain and will be denoted as follows:

$$K_t = s_{t|t}^2 s_v^{-2}. \tag{13}$$

We solve (11) for $s_{t|t-1}^{-2}$ and substitute the result with (13) into (12) to get

$$\begin{aligned} x_{t|t} &= s_{t|t}^2 \left\{ \left(s_{t|t}^{-2} - s_v^{-2} \right) x_{t|t-1} + s_v^{-2} y_t \right\} \\ &= x_{t|t-1} + K_t \left(y_t - x_{t|t-1} \right). \end{aligned} \tag{14}$$

For plants with stationary transitions and constant observation and transition variances (which we assume here), all the computations of substep 1.1 can be done offline except for the mean of the predictive distribution. With or without those offline computations, if these "preparations" are done prior to the arrival of the latest observation y_t it can shorten slightly the time required to update our knowledge of the state of the plant.

Step 1.2. Updating. In "updating", we compute the prediction error and use that to update the "posterior mean", our point estimate of the state of the plant.

Step 1.2a. Prediction Error. When the observation y_t arrives, we compute the prediction error as,

$$e_t = y_t - f_t. \tag{15}$$

Step 1.2b. Posterior Mean. With the prediction error in hand, we multiply it by the Kalman gain and add the product to the prior mean to obtain the posterior mean per (14), as

$$x_{t|t} = x_{t|t-1} + K_t e_t . \tag{16}$$

This completes step 1, observation, in Bayesian sequential updating as outlined in Figure 1. Next, we permit the plant to transition in preparation for the next observation, per step 2.

Step 2. Transition and Prior for the Next Observation. Given the posterior mean and variance from step 1, we can easily compute using (1) the prior mean and variance for the next observations, as follows:

Step 2.1. Prior Mean.

$$x_{t+1|t} = x_{t|t}, \tag{17}$$

and

Step 2.2. Prior Variance.

$$s^2_{t+1|t} = s^2_{t|t} + s^2_w . \tag{18}$$

This completes step 2. The resulting prior distribution at one point in time $N(x_{t+1|t}, s^2_{t+1|t})$ becomes an input for step 1.1, $N(x_{t|t-1}, s^2_{t|t-1})$, at the next point in time. In this way, observations are processed sequentially as they arrive. If the model (1) - (2) is correct, then the prior $N(x_{t+1|t}, s^2_{t+1|t})$ summarizes all the information in $D_t = \{y_t, y_{t-1}, ..., y_1, x_{1|0}, s_{1|0}\}$ about the state of the plant just prior to observation y_{t+1}.

The most important expressions in this section are summarized in Figure 2. We next apply this iteration to an example (Section 4) before deriving some general properties of this case. These properties include the convergence of the Kalman gain to an asymptote (Section 5). This convergence turns out to be monotonic, which helps to establish it as a natural Bayesian answer to the fast initial response (FIR) problem. This is followed by discussions of

robustness (Section 7), threshold selection (Section 8), and concluding remarks on EWMAs (Section 9).

(Figure 2 about here)

4. SAMPLE COMPUTATIONS FOR A BAYESIAN EWMA

Sample computations using this procedure are given in Table 2 and Figure 3. As suggested in the summary box at the bottom of Figure 2, we really only need three columns from Table 2.2: the prior mean and variance and the Kalman gain. The remaining columns of Table 2.2 are provided to describe more clearly the machinery of Bayesian updating as discussed with Figures 2 and 1.

(Table 2 about here)

(Figure 3 about here)

Numbers for simulated "true state" and "observation" are given in the second and third columns of Table 2.2. These were obtained as pseudo-random numbers generated according to the "manufacturing distribution", "observation error", and "migration" described in Table 2.1. The manufacturing distribution is assumed to be normal with mean, variance, and information as given in Table 2.1; these define the prior at time $t = 1$. We begin computing the "posterior information" for observation 1 using expression (11), as $s_{1|1}^{-2} = s_{1|0}^{-2} + s_v^{-2} = 10 + 100 = 110$. The posterior variance $s_{1|1}^{2} = 1/110 = 0.00909$. The Kalman gain is obtained from (13) as $K_t = s_{1|1}^{2} s_v^{-2} = 0.00909 \times 100 = 0.909$. The prediction error (–0.063) is obtained as usual as the observation (–0.063) minus the forecast 0, per (15) and (8). The posterior mean, per (16),

is then the prior mean plus $K_t e_t = 0 + (0.909) \times (-0.063) = (-0.057)$; this appears in Table 2.2 as the prior for time $t = 2$, per (17).

Note that the prior and posterior variances and information terms, $s^2_{t|t-1}$, $s^2_{t|t}$, $s^{-2}_{t|t-1}$, and $s^{-2}_{t|t}$, and the Kalman gain, K_t, all converge to constants to three significant digits by observation $t = 20$. This occurs here because s_v and s_w are constant. This is a special case of a more general result that for "completely observable" models (Gelb 1999, p. 142; Kalman and Bucy 1961) with constant, linear transitions and constant observation and transition covariance matrices, the Kalman gain and the prior and posterior covariance matrices all converge to constants. We next consider more carefully the behavior of K_t in this EWMA case.

5. KALMAN GAIN FOR A BAYESIAN EWMA

In this section, we study the behavior over time of the Kalman gain of (13) and tie more carefully the above model to a traditional EWMA. First, we combine (14) with (17) to obtain the following:

$$\begin{aligned} x_{t+1|t} &= x_{t|t-1} + K_t(y_t - x_{t|t-1}) \\ &= (1 - K_t)x_{t|t-1} + K_t y_t . \end{aligned} \tag{19}$$

This shows more clearly than (12) that the posterior mean $x_{t+1|t}$ is a weighted average of $x_{t|t-1}$ and y_t. By recursively substituting (19) into itself, we can show that $x_{t+1|t}$ is a weighted average of y_{t-j}, $j = 0, 1, ...$, with weights declining exponentially, provided $0 < K_t < 1$ and K_t is bounded away from 0 as $t \to \infty$. [Box and Luceño (1997, p. 69, 91) show how (19) implies that $x_{t+1|t}$ is

a weighted average of y_{t-j}, $j = 0, 1, ...$, with weights declining exponentially assuming $K_t = K_\infty$ is constant.]

To confirm that $0 < K_t < 1$, substitute (11) into (13) to obtain the following:

$$K_t = s_{t|t}^2 s_v^{-2} = \left(s_{t|t-1}^{-2} + s_v^{-2}\right)^{-1} s_v^{-2} = \left(1 + s_v^2 s_{t|t-1}^{-2}\right)^{-1}. \tag{20}$$

But s_v^2 and $s_{t|t-1}^2$ are both variances and strictly positive, from which we conclude that $0 < K_t < 1$. We shall develop a recursion for K_t, which becomes $K_t = 1/t$ when $K_1 = 1$ and the migration rate $s_w = 0$. This will establish that $x_{t+1|t}$ in (19) is in this case a simple average updated one observation at a time. For a rigorous proof of the asymptotic behavior of K_t, we will refer the reader to the more general result of Kalman and Bucy (1961, theorem 5), though we will derive expressions for that asymptote and discuss its nature.

To derive a recursion for K_t, we first substitute (18) into (11) to obtain the following:

$$\begin{aligned} s_{t|t}^{-2} &= \left(s_{t-1|t-1}^2 + s_w^2\right)^{-1} + s_v^{-2} \\ &= s_v^{-2}\left\{\left(r^2 + s_{t-1|t-1}^2 / s_v^2\right)^{-1} + 1\right\}, \end{aligned}$$

where $r = s_w / s_v$. We multiply both sides of this equation by s_v^2 and recall the definition of K_t, (13), to get the following:

$$K_t^{-1} = \left\{\left(r^2 + K_{t-1}\right)^{-1} + 1\right\}. \tag{21}$$

In Figure 4, we present the behavior of K_t over time for different levels of r assuming $s_{1|0}^2 = \infty$. A similar analysis would establish that $s_{t+1|t}^2$ and $s_{t|t}^2$ also approach asymptotes, which we denote by $s_{\infty+1|\infty}^2$ and $s_{\infty|\infty}^2$ with an obvious abuse of notation. These latter two asymptotes must satisfy (11) and (18), which means that $s_{\infty+1|\infty}^2 > s_{\infty|\infty}^2$.

(Figure 4 about here)

This figure suggests that K_t goes to an asymptote, K_∞, say, that depends on r. Moreover, for $r \geq 0.3$, this asymptote is essentially achieved by observation 10. For $r \geq 3$, the asymptote is essentially achieved by $t = 2$. For $r \leq 0.1$, the asymptote is not achieved 10 observations. We shall see below that for $r = 0$, this asymptote is 0; otherwise, this suggests that K_t is bounded away from 0, which with (19) establishes that $x_{t+1|t}$ is a moving average of all previous observations with weights declining exponentially with age. For a rigorous proof of this, see the more general work of Kalman and Bucy (1961). For a similar discussion of the EWMA, see Kirkendall (1989) and Harvey (1989, pp. 119, 124).

To obtain a formula for this asymptote, we substitute K_∞ for both K_t and K_{t-1} in (21) and solve for K_∞. We get the following:

$$\begin{aligned} K_\infty &= \frac{1}{2}\left\{\sqrt{r^4 + 4r^2} - r^2\right\} \\ &= \frac{r^2}{2}\left\{\sqrt{1 + \left(4/r^2\right)} - 1\right\} \\ &= r\left\{\sqrt{1 + r^2/4} - \left(r/2\right)\right\}. \end{aligned} \tag{22}$$

To study the asymptotic behavior of K_∞ as r gets large or small, we use the binomial theorem as $\sqrt{1+x} = 1 + (x/2) - (x^2/8) + O(x^3)$ in these last two expressions to get the following:

$$\begin{aligned} K_\infty &= 1 - r^{-2} + O\left(r^{-4}\right) \\ &= r\left\{1 - \left(r/2\right) + \left(r^2/8\right) - \left(r^4/128\right) + O\left(r^6\right)\right\}. \end{aligned} \tag{23}$$

The asymptote (22) is plotted vs. r in Figure 5. The most obvious conclusion from (22) and Figures 4 and 5 is that the choice of weight on the last observation is equivalent to specifying r = the square root of the migration variance, s_w^2, as a proportion of the

measurement noise, s_v^2. This relationship quantifies what we would qualitatively expect: With processes that change slowly relatively to the measurement noise, history is more informative that the last observation. On the other hand, rapidly changing processes with relatively informative observations find recent history more relevant than the past for predicting the future. The asymptotic expansions in (23) quantify the behavior we see in Figure 5 for large and small r.

(Figure 5 about here)

For future reference, we multiply K_∞ in (22) by s_v^2 to obtain the asymptotic posterior variance $s_{\infty|\infty}^2$ per (13) as follows:

$$s_{\infty|\infty}^2 = \frac{1}{2} s_w^2 \left\{ \sqrt{1 + \left(4/r^2\right)} - 1 \right\}.$$

Similarly, we use (18) to get the asymptotic prior variance $s_{\infty+1|\infty}^2$ as follows:

$$s_{\infty+1|\infty}^2 = \frac{1}{2} s_w^2 \left\{ \sqrt{1 + \left(4/r^2\right)} + 1 \right\}. \tag{24}$$

If the initial prior variance $s_{1|0}^2$ exactly equals the asymptotic prior variance $s_{\infty+1|\infty}^2$, the weight on the last observation K_t will be constant, giving us a traditional EWMA. If, as is usually the case, $s_{1|0}^2 > s_{\infty+1|\infty}^2$, the Kalman gain will start larger than K_∞ and decline monotonically towards it. If $s_{1|0}^2 < s_{\infty+1|\infty}^2$, the Kalman gain will start smaller than K_∞ and increase monotonically towards it, reflecting the fact that we now lose more information from migration than we gain from each observation.

6. BAYESIAN AND TRADITIONAL APPROACHES TO FIR FOR EWMA

Lucas and Saccucci (1990) observed that in many applications of a traditional EWMA [with a constant weight on the last observations $K_t = K_\infty$ in (19)], the resulting EWMA may require too many observations to cross a threshold if the plant is bad, starting with $x_{1|0} = m_0$. Their solution is to start with 25, 50 or 75 percent "head start", i.e., with $x_{1|0} = m_0 + p(m_1 - m_0)$, where $p = 0.25$, 0.5, or 0.75.

Bayesian sequential updating, as outlined in Figures 1 and 2, seems to provide a more comprehensive and understandable approach to this important problem: In terms of the theory developed in Sections 3 - 5 above, Lucas and Saccucci essentially assume that $s_{1|0} = s_{\infty+1|\infty}$ [see (24)], but that $x_{1|0}$ is misspecified and is better given as $x_{1|0} = m_0 + p(m_1 - m_0)$, where $p =$ 0.25, 0.5, or 0.75. A constant weight on the last observation K_t implies that the starting value for the EWMA $x_{1|0}$ tells us as much about the condition of the plant as $x_{101|100}$, the value of the EWMA after 100 observations. While this would very rarely be true, there might be many situations where the benefits from using a varying K_t might not justify the work of estimating migration and noise variances and computing K_t with each observation.

The data required to determine a changing K_t is virtually always available. A monitor is almost never designed for a situation (plant), that is totally unique. Someone suspects that a fault of a certain type may occur. This suspicion is based on somebody's experience with other applications that bear some resemblance to the problem at hand. This experience provides access to an external reference distribution and data on the reliability of the plant from which estimates for the initial prior ($x_{1|0}$, $s^2_{1|0}$) and migration variance s^2_w can be derived. We

combine this with a gage repeatability and reproducibility study (e.g., NIST 2001, ch. 2) to estimate the noise variance s_v^2. In this way, Bayesian sequential updating shows a person designing a monitor precisely how to use this relevant information; the previously existing theory seems silent about the relevance of this information and how to use it.

For the FIR problem, this recommends adjusting $s_{1|0}^2$, not $x_{1|0}$, in the initial prior. The disadvantage is that the weight on the last observation is not constant but must be updated with each observation per (21) until convergence to K_∞ is essentially achieved.

Steiner (1999) suggested varying the control limits, using the variance of an EWMA assuming the first observation is known exactly. This produces initial detection limits that are tighter than the asymptotic limits. In our judgment, this is exactly the opposite of what is needed in most FIR applications, where a fast initial response is needed precisely because of initial uncertainty regarding the state of the plant. We suspect that detection limits should perhaps also vary with time, as Steiner suggests. However, the "best" limits should be driven by an appropriate cost structure. As noted in Section 8 below, this is beyond the scope of the present article.

Finally, we believe that Bayesian sequential updating, as exemplified by the current work, provides a comprehensive theoretic foundation for work on the short run process control problem, recently discussed, e.g., by Nembhard and Mastrangelo (1998).

7. ROBUSTNESS

Box has noted that robustness is often more important than optimality, since a theoretically optimal solution may be so non-robust that it performs miserably under common discrepancies between reality and standard assumptions. Box and Luceño (1997, pp. 117-127) find that the EWMA provides a quite robust procedure for tracking a drifting process, even if the migration mechanism differs substantially from the random walk of (1). This is consistent with other work, e.g, Srivistava and Wu (1993) and Roberts (1966), that finds that the EWMA performs reasonably well under a broad variety of circumstances, though not as well as a cumulative sum in reacting to certain abrupt jumps.

However, a procedure may be robust to one kind of model inadequacy but quite sensitive, nonrobust, to another. For example, an EWMA may follow a process average reasonably well even if the transitions differ substantially from the random walk of (1) and the weight on the last observation differs from the optimal. However, we would expect that confidence intervals using the predictive or the prior variance, (9) or (18), might *not* perform very well if either the measurement or the migration variance, s_v^2 or s_w^2, were poorly estimated. Traditional methods for estimating these parameters and evaluating the applicability of this model are discussed by Box and Luceño (1997, pp. 117-127).

There is at present another practical disadvantage to the use of our Bayesian EWMA, (19) with (21): We do not currently have a simple method for estimating the run length characteristics for the Bayesian EWMA, other than suggesting that it probably does not differ

substantially from the traditional FIR technique proposed by Lucas and Saccucci. However, this is not conceptually a difficult problem and can be addressed, e.g., by Monte Carlo.

8. WHEN TO DECLARE A MALFUNCTION?

It is not as easy here to decide when to set an alarm as it is when applying Bayesian sequential updating to the problem of detecting an abrupt jump from a simple null hypothesis to a simple alternative (Graves, Bisgaard and Kulahci 2002). There, the posterior consisted of one number; here, it is a distribution with two parameters that migrate over time. One approach might be to develop (*a*) an appropriate cost structure and (*b*) a decision procedure to minimize the cost per unit time or total discounted cost over an indefinite future. Related problems have discussed by Berger (1985, ch. 7) and West and Harrison (1999, sec. 11.6).

We have not done that here. Instead, we divided the real line into "acceptable", "unacceptable", and "undefined" regions: The plant is "good" if $L \leq x_t \leq U$, "bad" if $\{x_t \leq L_1 < L$ or $x_t \geq U_1 > U\}$, and "undefined" if $\{L_1 < x_t < L$ or $U < x_t < U_1\}$; L and U are "worst acceptable", while L_1 and U_1 are "best unacceptable" (Box et al. 1999).

We further simplified the problem by selecting decision limits L^* and U^* and indicating a malfunction when the prior variance $s_{t+1|t}$ is sufficiently small *and* $x_{t+1|t}$ is outside the L^*-U^* limits. The engineering design criteria for this decision procedure (or on-board diagnostic, OBD) are typically expressed in terms of an acceptably small probability of an excessive delay (to detection of x_t being bad) and simultaneously a small probability of a false alarm in the design life of the plant (Box et al. 1999; 2002).

This procedure is, apart from the changes in the weight on the last observation, a standard EWMA. Run length distributions of EWMAs have been studied, for example, by Crowder (1987) and Lucas and Saccucci (1990), though the effect of the Bayesian non-constant weights (13) and (21) seem not to have been described in the literature. Decision limits L^* and U^* could be obtained by Monte Carlo simulation if the work of Crowder and others does not seem adequate.

Steiner (1999) proposed an FIR approach using varying control limits whose width was a constant multiple of the standard deviation of a finite duration EWMA assuming the initial state is known exactly. This produces detection limits that are initially narrow and expand to the standard asymptotic limits. Our work suggests that decision limits of this nature would only be appropriate if the initial condition of the plant were known quite precisely. In our experience, the opposite is more common: Initial data collection is often much more informative about the condition of the plant than the distribution at first use. In this case, we suspect that optimal detection limits would start begin wider and converge to constants asymptotically, depending on the assumed cost structure.

9. DISCUSSION

In this article, we have derived Bayesian sequential updating for indirect observation of a univariate normal random walk. The result is a Bayesian EWMA, previously discussed by Kirkendall (1989) and Harvey (1989). Our derivation is, we believe, more methodical and more easily understood and generalized than previous discussions of this case. Neither

Kirkendall nor Harvey discussed the objective Bayesian possibilities of using reliability / hazard rate data, gage repeatability and reproducibility studies, and data on the distribution at first use.

Part of this development established that the selection of a weight for the last observation in an EWMA is equivalent to specifying the migration variance s_w^2 relative to the noise variance s_v^2, as discussed with (21) above. Both of these quantities are generally available from external sources: The migration variance s_w^2 is tied to reliability. The noise variance s_v^2 can be estimated from a metrology study. Moreover, the distribution at the initiation of monitoring $N(x_{1|0}, s_{1|0}^2)$ is obtainable from data typically collected at the end of the production line for manufactured products or from other sources for monitoring in clinical trials or other applications. This places at the disposal of a person designing a monitor relevant information whose use in this context has not been previously discussed in the literature that we have seen. This also provides an alternative to traditional Fast Initial Response strategies, which are equivalent to assuming that the starting value used for the EWMA tells us as much about the condition of the plant as later values; see Section 5 above.

Our proposed use of gage R & R and reliability data provides an alternative approach for determining K_t to the integrated moving average IMA(0, 1, 1) estimation procedure recommended by Box and Luceño (1997, sec. 4.8). We recommend that users do estimation and model criticism for an IMA(0, 1, 1) as discussed by Box and Luceño as a check on the assumption of the indirectly observed random walk model of (1) and (2) and the parameters estimated using gage R & R and reliability data. Apart from the non-constant nature of K_t, the model (1) - (2) is virtually equivalent to an IMA(0, 1, 1). Therefore, if a substantive

discrepancy is found either in an IMA(0, 1, 1) model or with the estimated weight K_t, it suggests a problem that warrants further investigation and corrective action.

As noted in Section 7, our Bayesian EWMA is essentially as robust as traditional EWMA procedures, being computationally almost identical to them. The Bayesian EWMA provides an additional interpretation as the prior and posterior at each step of objective distributions of units or plants among all with comparable histories. We would not expect this probability interpretation to be robust to departures from serial independence or normality or from poor estimation of s_w^2 or s_v^2. However, more casual usage of this theory, consistent with current EWMA usage, should be quite robust.

Box and Luceño (1997) comment extensively about an EWMA as a forecast for an integrated moving average IMA(0, 1, 1) process. The present development is asymptotically equivalent to this, and we recommend traditional EWMAs for applications where the transients are unimportant and the situation does not justify the effort of attempting to access other data such as the distribution at first use, gage R & R studies, and reliability data.

The present study was conducted as part of a larger project to demonstrate the value of Bayesian sequential updating as a foundational principle for designing monitors (Graves et al. 2001). To detect an abrupt jump, this principle produces a "Bayes-adjusted Cusum" (Graves, Bisgaard and Kulahci 2002a). If both mean and variability migrate over time, the prior for the next observation involves EWMAs for both mean and variance (Graves, Bisgaard and Kulachi 2002b). For noisy observations linearly related to a multivariate state vector, Bayesian sequential updating produces Kalman filters that can isolate as well as detect faults (Graves et

al. 2001). In applications where the output of different sensors should be related, this can allow one sensor to check another, supporting fault isolation without duplicating sensors and increasing per-unit costs. If the migration rate is zero, this algorithm degenerates to ordinary least squares regression. This work generalizes Pole, West and Harrison (1994), West (1986), West and Harrison (1986), Gelb (1999), Gordon and Smith (1988, 1990), Harrison and Lai (1999), Lindley and Smith (1972), and others.

10. ACKNOWLEDGEMENTS

The work reported here was funded in part by the Low Emissions Research and Development Partnership of DaimlerChrysler, Ford and General Motors. The authors wish to thank John Van Gilder and Tom Ting of General Motors, Ken Marko and John James of Ford, and Hal Zatorski and Cuiping Wu of DaimlerChrysler for many stimulating discussions about this and related monitoring issues.

REFERENCES

Berger, J. O. (1985) *Statistical Decision Theory and Bayesian Analysis*, 2nd ed. (NY: Springer).

Box, G., Graves, S., Bisgaard, S., Van Gilder, J., Marko, K., James, J., Seifer, M., Poublon, M., and Fodale, F. (2000) "Detecting Malfunctions in Dynamic Systems", *Proceedings of the 2000 SAE World Congress & Exposition* (SAE Technical Paper Series number 2000-01-0363).

Box, G., Bisgaard, S., Graves, S., Kulahci, M., Marko, K., James, J., Van Gilder, J., Ting, T., Zatorski, H., and Wu, C. (2002). "The Waterfall Chart", *Quality Engineering*. To appear.

Box, G., and Luceño, A. (1997) *Statistical Control by Monitoring and Feedback Adjustment* (NY: Wiley).

Crowder, S. V. (1987) "A Simple Method for Studying Run-Length Distributions of Exponentially Weighted Moving Average Charts", *Technometrics*, 29, 401-407.

DeGroot, M. H. (1970) *Optimal Statistical Decisions* (NY: McGraw-Hill).

Gelb, A. (1999) *Optimal Applied Estimation* (Cambridge, MA: MIT Press).

Gordon, K., and Smith, A. F. M. (1988) "Modeling and Monitoring Discontinuous Changes in Time Series" in *Bayesian Analysis of Time Series and Dynamic Models*, ed. J. Spall, N.Y.: Marcel Dekker, 359-391.

Gordon, K., and Smith, A. F. M. (1990) "Modeling and Monitoring Biomedical Time Series", *Journal of the American Statistical Association*, 85, 328-337.

Graves, S. B. (2002) "Bayes' Rule of Information" (unpublished technical report downloadable from www.prodsyse.com).

Graves, S. B., Bisgaard, S., and Kulahci, M. (2002a) "A Bayes-Adjusted Cusum" (unpublished technical report downloadable from www.prodsyse.com).

Graves, S. B., Bisgaard, S., and Kulahci, M. (2002b) "A Bayesian EMWA for Mean and Variance" (unpublished technical report downloadable from www.prodsyse.com).

Graves, S. B., Bisgaard, S., Kulahci, M., Van Gilder, J., Ting, T., Marko, K., James, J., Zatorski, H., and Wu, C. (2001) "Foundations of Monitoring Dynamic Systems" (unpublished technical report downloadable from www.prodsyse.com).

Harrison, P. J., and Lai, I. C. H. (1999) "Statistical Process Control and Model Monitoring", *Journal of Applied Statistics*, 26: 273-292.

Harvey, A. C. (1989) *Forecasting, Structural Time Series Models and the Kalman Filter* (NY: Cambridge University Press).

Kalman, R. E. (1960) "A New Approach to Linear Filtering and Prediction Problems", *Journal of Basic Engineering*, 340-345.

Kalman, R. E., and Bucy, R. (1961) "New Results in Linear Filtering and Prediction Theory", *Journal of Basic Engineering* (Transactions of the ASME), vol. 83D, pp. 95-108.

Kirkendall, Nancy J. (1989) "The Relationship Between Certain Kalman Filter Models and Exponential Smoothing Models", pp. 89-107 in J. B. Keats and N. F. Hubele, *Statistical Process Control in Automated Manufacturing*, NY: Marcel Dekker.

Lindley, D. V., and Smith, A. F. M. (1972) "Bayes Estimates for the Linear Model", *Journal of the Royal Statistical Society, Series B*, 34, 1-41.

Lucas, J. M., and Saccucci, M. S. (1990) "Exponentially Weighted Moving Average Control Schemes: Properties and Enhancements" (with discussion), *Technometrics*, 32(1), pp. 1-29.

Nembhard, H. B., and Mastrangelo, C. M. (1998) "Integrated Process Control for Startup Operations", *Journal of Quality Technology*, 30(3): 201-211.

NIST (2001) *Engineering Statistics Handbook* (Washington, DC: National Institute of Standards and Technology web-based handbook: http://www.itl.nist.gov/div898/handbook)

Pole, A., West, M., and Harrison, H. (1994) *Applied Bayesian Forecasting and Time Series Analysis* (NY: Chapman & Hall)

Roberts, R. W. (1966) "A Comparison of Some Control Chart Procedures", *Technometrics*, 8: 411-430.

Srivistava, M. S., and Wu, Y. (1993) "Comparison of EWMA Cusum, and Shyryayev-Roberts Procedures for Detecting a Shift in the Mean", *Annals of Statistics*, 21: 645-670.

Steiner, S. (1999) "EWMA Control Charts with Time-Varying Control Limits and Fast Initial Response", *Journal of Quality Technology*, 31: 75-86.

West, M. (1986) "Bayesian Model Monitoring", *Journal of the Royal Statistical Association* B, 48: 70-78.

West, M. and Harrison, P. J. (1986) "Monitoring and Adaptation in Bayesian Forecasting Models", *Journal of the American Statistical Association*, 81, 741-750.

_____ (1999) *Bayesian Forecasting and Dynamic Models*, 2nd ed. (NY: Springer).

Figure 1. Bayesian Sequential Updating with a Random Walk

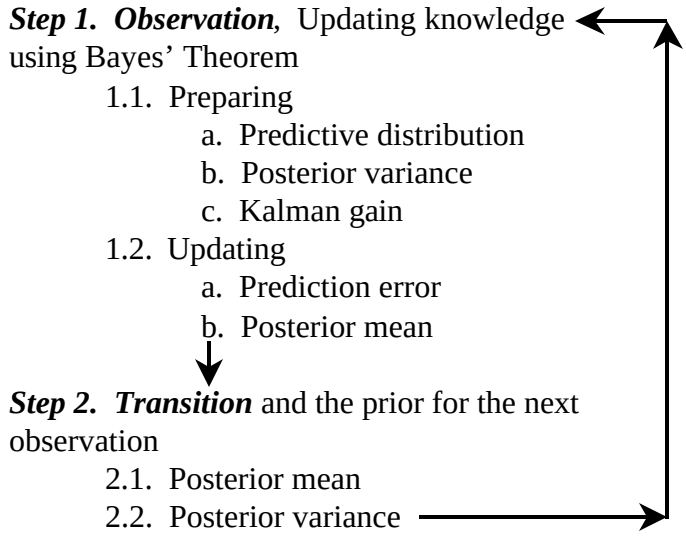

Figure 2. Bayesian EWMA Iteration

Step 1. Observation, updating knowledge using Bayes' theorem

1.0. Observation model

a. Prior $(x_t \mid D_{t-1}) \sim N(x_{t|t-1},\ s^2_{t|t-1})$ (from step 2)

b. Observation $(y_t \mid x_t) \sim N(x_t,\ s^2_v)$ (2)

1.1. Preparing

a. Predictive distribution $(y_t \mid D_{t-1}) \sim N(f_t,\ s^2_{y|t-1})$,

$f_t = x_{t|t-1},\ s^2_{y|t-1} = s^2_{t|t-1} + s^2_v$ (9)

b. Posterior information and variance $s^{-2}_{t|t} = s^{-2}_{t|t-1} + s^{-2}_v$ (11)

c. Kalman gain $K_t = s^2_{t|t} s^{-2}_v$ (13)

1.2. Updating

a. Prediction error $e_t = y_t - f_t$ (15)

b. Posterior mean $x_{t|t} = x_{t|t-1} + K_t e_t$ (16)

Step 2. Transition and the prior for the next observation

2.0. Model $(x_{t+1} \mid x_t) \sim N(x_t,\ s^2_w)$ (1)

2.1. Posterior mean $x_{t+1|t} = x_{t|t}$ (17)

2.2. Posterior variance $s^2_{t+1|t} = s^2_{t|t} + s^2_w$ (18)

In sum: Combining (14) - (16):

$$x_{t+1|t} = x_{t|t-1} + K_t\left(y_t - x_{t|t-1}\right) = \left(1 - K_t\right)x_{t|t-1} + K_t y_t \quad (19)$$

where

$$K_t = 1\Big/\left\{1 + \left[1\Big/\left(r^2 + K_{t-1}\right)\right]\right\},\ \ r = s_w / s_v \quad (21)$$

$$\rightarrow K_\infty = \left(r^2/2\right)\left\{\sqrt{1 + \left(4/r^2\right)} - 1\right\} \quad (22)$$

For confidence limits, combine (11) and (18):

$$s^2_{t+1|t} = \left(s^{-2}_{t|t-1} + s^{-2}_v\right)^{-1} + s^2_w$$

Figure 3. A Bayesian EWMA

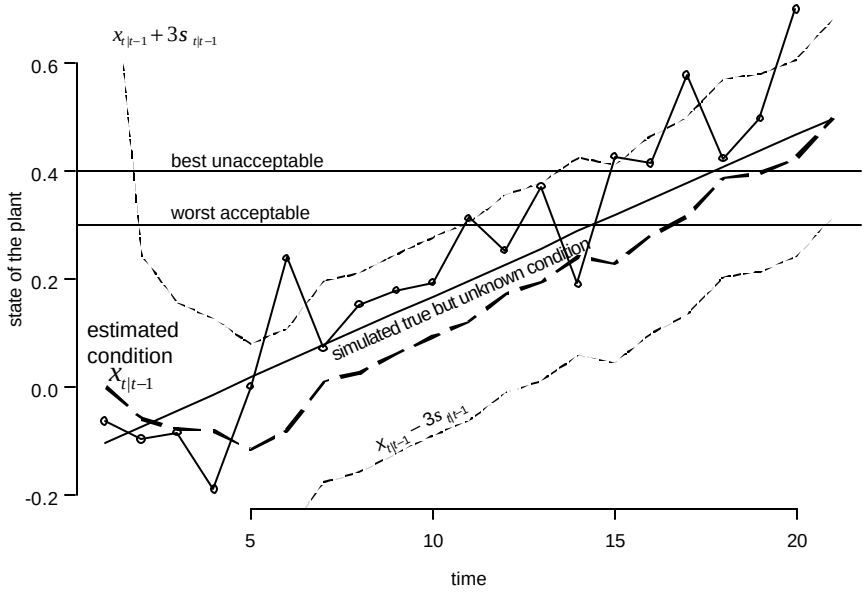

Figure 4. Kalman Gain for EWMA vs. Time

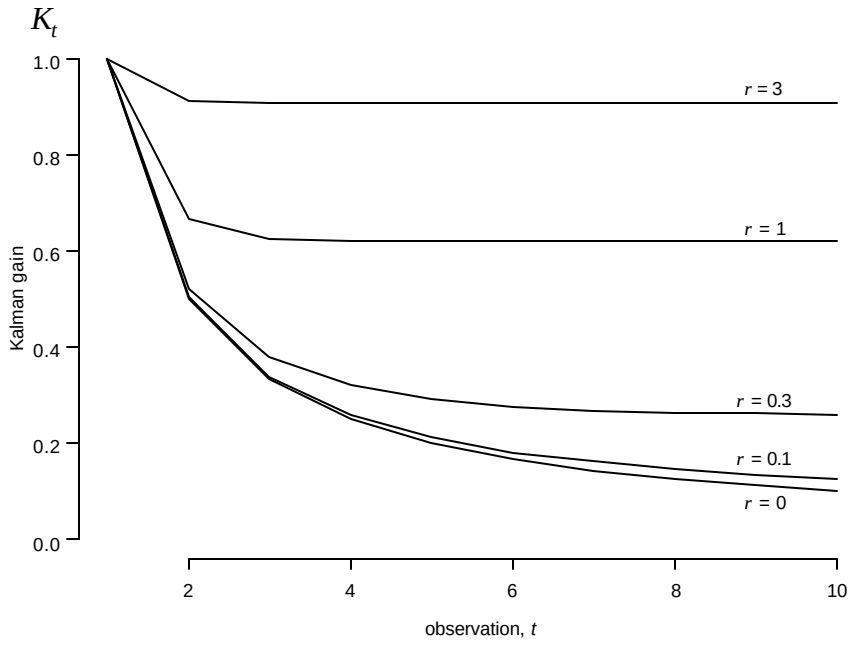

Figure 5. Equivalence between EWMA Weight and Relative Migration Rate

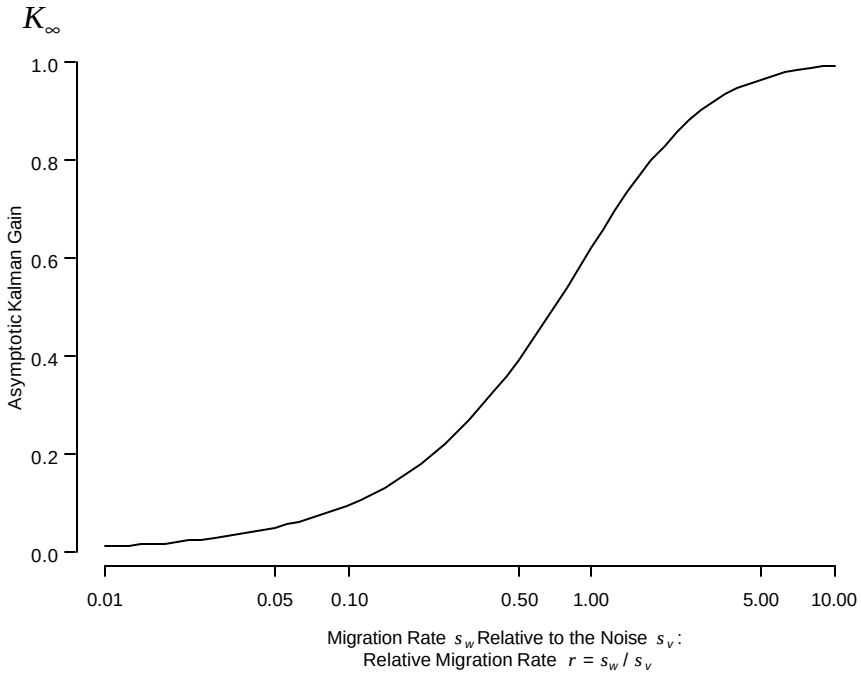

Table 1. Bayesian EWMA Computations

Prediction $x_{t+1|t}$ for time $(t + 1)$ given information $D_t = \{y_t, y_{t-1}, ...\}$, available at time t:

$$x_{t+1|t} = (1 - K_t)x_{t|t-1} + K_t y_t = x_{t|t-1} + K_t e_t, \quad e_t = (y_t - x_{t|t-1}) \tag{19}$$

Weight on the last observation (Kalman gain):

$$K_t = 1 / \{ 1 + [1/(r^2 + K_{t-1})] \} \tag{21}$$

where

$r^2 = s_w^2 / s_v^2$ = (migration variance) / (measurement variance)

and

$K_1 = 1$ with no prior knowledge of the initial condition of the plant

Confidence bounds on x_t are obtained from $(x_t \mid D_{t-1}) \sim N(x_{t|t-1}, s_{t|t-1}^2)$, where

$$s_{t+1|t}^2 = \left(s_{t|t-1}^{-2} + s_v^{-2}\right)^{-1} + s_w^2 \tag{11) \& (18}$$

Evaluate prediction error e_t relative to

$$\text{var}(y_t \mid D_{t-1}) = s_{t|t-1}^2 + s_v^2 \tag{9}$$

Table 2. Univariate Bayesian Sequential Updating: Illustrative Calculations

Table 2.1. Scenario Simulated

Manufacturing distribution

Mean	Variance	Standard Deviation	Information
$x_{1\|0}$	$s^2_{1\|0}$	$s_{1\|0}$	$s^{-2}_{1\|0}$
0	0.1	0.316	10

Observation error	Variance	Standard Deviation	Information
	s^2_v	s_v	s^{-2}_v
	0.01	0.1	100

Migration	Mean	Variance	Standard Deviation	Information
	m_t	s^2_w	s_w	s^{-2}_w
"Actual"	0.03	0	0	∞
Assumed	0	0.001	0.0316	1,000

Table 2.2. Illustrative Calculations

	Simulated		Prior from Previous Step 2			Intermediate Computations in Step 1			
						Posterior		Kalman	Prediction
Time	True State x_t	Observation $y_t = x_t + v_t$	Mean $x_{t\|t-1}$	Variance $s^2_{t\|t-1}$	Information $s^{-2}_{t\|t-1}$	Information $s^{-2}_{t\|t}$	Variance $s^2_{t\|t}$	Gain K_t	Error e_t
eq'n→	(1)	(2)	(17) [+(16)]	(18)		(11)		(13)	(15)+ (8)
1	– 0.103	– 0.063	0	0.1000	10.0	110.0	0.00909	0.909	– 0.063
2	– 0.073	– 0.097	– 0.057	0.0101	99.1	199.1	0.00502	0.502	– 0.040
3	– 0.043	– 0.084	– 0.077	0.0060	166.0	266.0	0.00376	0.376	– 0.007
...									...
19	0.437	0.497	0.396	0.0037	270.2	370.2	0.00270	0.270	0.101
20	0.467	0.698	0.423	0.0037	270.2	370.2	0.00270	0.270	0.275

21	0.497	0.497	0.0037	270.2

Figure Captions

Figure 1. Bayesian Sequential Updating with a Random Walk

Figure 2. Bayesian EWMA Iteration

Figure 3. A Bayesian EWMA

Figure 4. Kalman Gain for EWMA vs. Time

Figure 5. Equivalence between EWMA Weight and Relative Migration Rate