



## An alternative mean estimator for processes monitored by SPC charts

ARGON CHEN<sup>†\*</sup> and E. A. ELSAYED<sup>‡</sup>

Statistical Process Control (SPC) chart is a graphical tool that helps to identify a possible process shift. Though the SPC chart is, in theory, a monitoring tool that reveals only the result of a statistical hypothesis testing, in practice the chart's signaling is often used for making process adjustment. In this paper we develop a robust estimator of the process mean for processes monitored by SPC charts. This mean estimate can serve as an important reference for investigating assignable causes and taking appropriate corrective actions. When the control chart is implemented as a control device, the mean estimate can also provide a more accurate assessment of the amount of adjustment to be made to the process. We also demonstrate in this paper that when the process shift is relatively small, the proposed mean estimate outperforms the mean estimate usually used for the CUSUM scheme.

### 1. Introduction

SPC charts are very widely used in practice (Montgomery 1996). In particular, the Shewhart control chart (Shewhart 1931) developed more than 60 years ago remains the most popular, if not the most effective, monitoring tool. Though the statistical theories and assumptions underlying the SPC chart have been fully discussed, its employment draws continuous debates and discussions. Different types of industries have different needs in monitoring or controlling production processes. Therefore, distinct employment strategies are developed to accommodate the SPC charts. In some industries the chart is used as a straightforward monitoring device that warns workers of a possible out-of-control process. Other industries use it as a control device that alerts workers and provides a basis for process adjustment.

SPC charts are designed to detect shifts among natural fluctuations caused by chance noises. For example, the Shewhart chart utilizes the standard deviation (SD) statistic to measure the size of the in-control process variability. By graphically contrasting the observed deviations against a multiple (usually, triple) of SDs, the control chart is intended to identify unusual departures of the process from its normal state (controlled state). Under certain assumptions, when the observed deviation from the mean exceeds three SDs, it is said that the process is out of control since there is only a probability of 0.0026 for the observation to fall outside the three SD limits given an unshifted mean chance the process mean is shifted. This Shewhart chart scheme is in effect a statistical hypothesis testing that reveals only whether the process is still in-control. This is certainly not enough for practical applications

---

Revision received February 2000.

<sup>†</sup> Graduate Institute of Industrial Engineering, National Taiwan University, 1, Sec. 4, Roosevelt Rd., Taipei, Taiwan, 106.

<sup>‡</sup> Department of Industrial Engineering, Rutgers, The State University of New Jersey, 96 Frelinghuysen Rd., Piscataway, NJ 08854, USA.

\* To whom correspondence should be addressed. e-mail: achen@ccms.ntu.edu.tw

where the causes of the problem should be found and removed to restore the production process to its satisfactory operational state.

It is the search for the out-of-control causes and the reinstatement of the process that complete the statistical process control loop. To reinstate the process, necessary corrective actions or process adjustment are performed following the chart's signaling. Before corrective actions are taken, it is critical to search for the cause of the out-of-control occurrence. During this search process, a robust process mean estimate can serve as a valuable reference that helps to determine the size of the process deviation and thus limit the search to a selective set of possible causes. Sometimes, process adjustment is made following the out-of-control signal to bring the departed process back to the in-control state. The size of the adjustment is usually based on the runs pattern of the data points and most frequently on the judgment of experienced process engineers. An accurate estimate of the mean will certainly provide a more precise adjustment of the process. In this paper, a robust mean estimate for processes monitored by SPC charts is developed to fulfill these needs.

In the literature, process mean estimates are usually developed for processes governed by a linear stochastic system and can be described using state-space models or autoregressive integrated moving average (ARIMA) models. Harrison and Stevens (1971, 1976), apply state-space models to situations where shifts are exhibited both in the process level and the slope. As the authors point out, their model for shifts in the mean is, in a way, analogous to the ARIMA (0, 1, 1) model in Box and Jenkins (1968, 1970). This integrated moving average (IMA or ARIMA (0, 1, 1)) model can be transformed into an exponentially weighted moving average (EWMA) estimator, which in turn can be realized as the popular PID controller. Recently, the EWMA estimator attracted the attention of both researchers (MacGregor 1988, Box and Kramer 1992, Ingolfsson and Sachs 1993, Yashchin 1995) and practitioners (Sachs *et al.* 1995, Boning *et al.* 1995).

Process adjustment based on process mean estimation is a feedback control scheme (Åström 1970, Box and Jenkins 1968, 1970, Davis and Vinter 1985) that is generally referred to as the Automatic Process Control (APC) approach. Most of the researchers in the APC area focus on identifying the dynamic nature of the process and disturbance, and constructing a linear system to describe them. The SPC chart is intended to identify the process departure from its normal state. This is referred to as the SPC approach. Extensive discussions about APC and SPC can be found in MacGregor (1987, 1988), Box and Kramer (1992), Vander Wiel *et al.* (1992), Tucker *et al.* (1993) and Sachs *et al.* (1995). While we agree that these two approaches have their own roles and should not be obscured with their seemingly related names, we also think that a plausible mean estimate should be developed to provide a basis for further actions when a process is signaled out-of-control by a statistical control chart.

Barnard (1959) was the first to point out the possibility of using control chart as an element in a feedback control loop. He suggests that the control chart should not only be used as a monitoring device but also as an estimator. The work, however, received limited attention from industrial practitioners. This is possibly because the cumulative sum (CUSUM) control chart (Page 1954, 1955, Chernoff and Zacks 1964) used by Barnard is relatively complex and is less popular than the Shewhart control chart. The Shewhart chart remains the dominant control chart in practice because of its simplicity and the development of its supplementary runs rules (Hoerl and Palm 1992). Nevertheless, Barnard's original idea provides a new perspective for

applications of control charts. Kelton *et al.* (1990) developed certain adjustment rules based on quality control charts but only looked on observations after the out-of-control point. Nishina (1992) investigated the problem of change-point estimation for control charts using accumulated data (CUSUM, EWMA and MA charts). In this paper we follow Barnard's idea but focus on the popular Shewhart control chart first. We then show how the approach can be applied to other types of control chart.

We start with the introduction in this section. The rest of this paper is organized into four sections. The process model is described and the basic assumptions behind SPC charts are discussed in the following section. In Section 3 we develop the mean estimate for the Shewhart chart with a single  $3\text{-}\sigma$  rule and a constant shift occurrence probability followed by an example and discussions. The mean estimate is then extended for Shewhart charts with supplementary runs rules and for SPC charts in general. An example will be given for the CUSUM scheme. Finally, we study a more realistic shift occurrence mechanism where the shift's inter-occurrence time follows a renewal process. The idea is also illustrated using an example.

## 2. Process model

In this paper we will use the simple Shewhart control chart and a semiconductor fabrication process as examples to illustrate the SPC ideas and to demonstrate the use of the proposed mean estimate. One of the very important processes in semiconductor manufacturing is to grow a layer of silicon dioxide ( $\text{SiO}_2$ ) by performing oxidation in a furnace. The critical output characteristic of the oxidation process is the thickness of the  $\text{SiO}_2$  layer. After oxidation, the thickness is measured in Angstrom ( $\text{\AA}$ ) and plotted on a Shewhart control chart to track the process condition. Very often, the process engineers use the control chart not only as a monitor of the furnace conditions but also a trigger for making adjustment to the furnace temperature.

Let the random variables  $X_1, X_2, X_3, \dots$  denote the 1st, 2nd, 3rd,  $\dots$  successive sample statistics of the quality characteristic observations, say the thickness of the  $\text{SiO}_2$  layer. The sample statistic  $X_i$  can be a single observed value or a sample mean of several observed values. It is assumed that  $X_i$  is normally distributed with mean  $\mu_i$  and variance  $\sigma^2$ . In the Shewhart control scheme, the process mean  $\mu$  is initially set to coincide with the target ( $T$ ). For instance, let the target of the oxide thickness be  $250 \text{ \AA}$ . When the process is in-control,  $X_i$  are said to have the same mean  $\mu (= 250)$  unless the following condition is violated:

$$\mu - 3\sigma \leq X_i \leq \mu + 3\sigma,$$

where  $\mu - 3\sigma$  and  $\mu + 3\sigma$  are the lower and upper control limits, respectively. Let the standard deviation ( $\sigma$ ) of the thickness be  $1.5 \text{ \AA}$ . When any of the control limits is exceeded, the process is said to be probably out-of-control; i.e. the process mean has probably departed from the target  $250 \text{ \AA}$ . The chance for either one of the control limits to be exceeded at observation  $i$  given that the mean of  $X_i$  remains at  $250$  is

$$\begin{aligned} \alpha &= P(X_i > 254.5 | \mu_i = 250) + P(X_i < 245.5 | \mu_i = 250) \\ &= [1 - \Phi(3)] + \Phi(-3) = 2\Phi(-3) = 0.0026, \end{aligned}$$

where  $\Phi$  is the cumulative distribution function of a standard normal random variable. However, there is still a probability of  $0.0026$  that the observation falls

outside the control limits given an unchanged process mean. This probability is referred to as the probability of Type I error and is denoted by  $\alpha$ .

When one of the control limits is exceeded at the  $i$ th observation, this does not mean that the process is out-of-control exactly at this sample point. The process could have been out-of-control long before  $i$  but the control limit was not exceeded until observation  $i$ . When the process is actually out-of-control but control limits are not exceeded, this misinterpretation of the process state is called Type II error. The probability of this Type II error's occurrence can be calculated given the magnitude of the deviation in the mean. For instance, if the mean changes from 250 to 253 (that is, the mean of the thickness deviates from the target by  $2\sigma$ ), then the Type II error probability, denoted by  $\beta$ , using the Shewhart chart scheme is

$$\begin{aligned} \beta &= P(X_i \in [245.5, 254.5] | \mu_i = 253) \\ &= \Phi\left(3 - \frac{253 - 250}{1.5}\right) - \Phi\left(-3 - \frac{253 - 250}{1.5}\right) = 0.841. \end{aligned}$$

As mentioned earlier, this Shewhart scheme reveals only a hypothesis testing result. In the following sections, we will estimate the process mean when the Shewhart chart signals an out-of-control.

### 3. Basic estimation procedure for Shewhart chart

Before estimating the process mean based on the Shewhart chart's signaling, we first examine the case where the point of the mean change is known. Given this change point information, two conditional mean estimators can be easily derived. Suppose that the current observation is  $j$  and the most recent mean change occurred at observation  $k$ . A straightforward estimate of the process mean is the sample mean; i.e. the arithmetic average of observations  $k$  to  $j$ :

$$\bar{x}_{j|k} = \frac{\sum_{i=k}^j x_i}{j - k + 1}, \tag{1}$$

where  $x_i$  is the observed value of  $X_i$ . Since the shift occurred at  $k$ , all observations from observation  $k$  forward have the same mean. This justifies (1). However, this estimate suffers from high instability especially when the sample size  $j - k + 1$  is small. After enough experience has been accumulated and information about the amount of mean changes has been gathered, the past experience can actually serve as the prior knowledge on how the process mean is likely to shift. Baye's estimation technique can then be applied to stabilize the mean estimate. Since the SPC chart's assumption about the process is normal distribution, a proper conjugate prior distribution is, therefore, also a normal distribution. Let the prior distribution's mean be  $\xi$  and variance be  $\tau^2$ . Bayes estimator under a quadratic loss function can be obtained (Lehmann 1991, pp 236–249) as follows:

$$\hat{\mu}_{j|k} = \frac{\tau^2 \bar{x}_{j|k} + \frac{\sigma^2}{j - k + 1} \xi}{\frac{\sigma^2}{j - k + 1} + \tau^2}. \tag{2}$$

By examination of (2), one can find that when the number of observations  $(j - k + 1)$  between the current point and the change point is large, then more weight is assigned to  $\bar{x}_{j|k}$ . On the other hand, when  $j - k + 1$  is small,  $\xi$  is assigned a higher weight; that is, the past experience becomes more important in the estimate. Also, the mean estimate approaches  $\xi$  when  $\tau^2$  is relatively small. This is because a small  $\tau^2$  means that we learn from experience that the process mean moves to a deterministic value  $\xi$  whenever a shift occurs. In addition, the Bayes estimator has been elegantly crafted to prevent the estimator from diverging for the case of a large  $\tau^2$  since the larger the  $\tau^2$  the less the weight assigned to the prior knowledge on the shift size ( $\xi$ ). In fact, when  $\tau^2$  is large, which means a very unreliable prior knowledge, one should really use the estimator in (1) instead of the Bayes estimator in (2).

We have estimated the process mean given that the change point is known. The change point is, however, unlikely to be known. The Shewhart chart is thus devised to detect the process change. The control chart does not reveal the actual location of the change point either. In other words, with the control chart detection we only know that a mean change may have occurred at or before this detection point. To more accurately estimate the process mean we need to know where the change point is likely to reside. A lag-window of observation points prior to the control chart's detection is defined to include all possible points at which the shift might occur.

Suppose that the current control chart's detection point is at  $S_2$  and the one preceding it is at  $S_1$ . In practice, corrective actions are taken after  $S_1$ , and when the process is in operation again the process mean is immediately subject to possible new shifts. As shown in figure 1, the lag-window is defined as the time period containing all observation points between  $S_1$  and  $S_2$  including  $S_2$ . The lag-window's definition depends greatly on the out-of-control practice. In any case, the lag-window's starting point should lie at the point where the process is deemed back to its mean condition.

Given the current detection point at  $S_2$ , it is plausible to say that a shift has occurred in the lag-window. Let the random variable  $K$  represent the exact location at which the shift occurs. We would like to know the probability distribution of  $K$ :

$$P_{k|S_2} = \Pr(K = k | \text{detection at } S_2, K \in (S_1, S_2]).$$

It should be noted that when the control chart gives an alarm and the process mean needs to be re-estimated it is assumed that one and only one process shift indeed occurs in the lag-window. That is, it is assumed that with preliminary investigation of the alarm, the false alarm possibility has been first ruled out. This assumption, however, is not unrealistic since in most cases alarms will be investigated and the shifted mean is only estimated for the out-of-control occurrence. There also exists a slim probability for more than one process shift to occur in the lag-window. The only

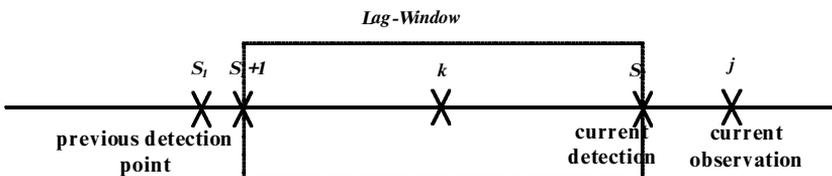


Figure 1. Shift occurrence lag window.

one assumed shift occurrence actually accounts for the very last shift that determines the latest process condition.

Applying Bayes' theorem, the probability that a shift occurs at a specific observation point, say  $k$ , in the lag-window given the control chart detection at  $S_2$  can be expressed as:

$$\begin{aligned}
 P_{k|S_2} &= \frac{P(K = k, \text{ detection at } S_2 | K \in (S_1, S_2])}{P(\text{detection at } S_2 | K \in (S_1, S_2])} \\
 &= \frac{P(\text{detection at } S_2 | K = k, K \in (S_1, S_2])P(K = k | K \in (S_1, S_2])}{\sum_{l=S_1+1}^{S_2} P(\text{detection at } S_2 | K = l, K \in (S_1, S_2])P(K = l | K \in (S_1, S_2])}. \tag{3}
 \end{aligned}$$

To compute  $P_{k|S_2}$ , two probabilities  $P(\text{detection at } S_2 | K = k, K \in (S_1, S_2])$  and  $P(K = k | K \in (S_1, S_2])$  need to be calculated.

$P(\text{detection at } S_2 | K = k, K \in (S_1, S_2])$  is the probability for the shift to be discovered at  $S_2$  given that the shift is known to occur at  $k$ ; that is,  $K = k$  and  $K \in (S_1, S_2]$ . This probability can be seen as the probability of taking  $S_2 - k + 1$  observations to detect the mean change.  $S_2 - k + 1$  is also called the run length, i.e. the number of observations elapsed since the change point, before the control chart signals. Given an estimate of the shift size, one can calculate Type II error probability ( $\beta_{S_2|k}$ ), and the run length probability distribution is simply the geometric distribution with the parameter  $\beta_{S_2|k}$ :

$$P(\text{detection at } S_2 | K = k, K \in (S_1, S_2]) = \beta_{S_2|k}^{S_2-k} (1 - \beta_{S_2|k}). \tag{4}$$

Equation (4) can be explained as follows. When the mean is shifted at  $k$ , the shift is not detected (with Type II error probability  $\beta_{S_2|k}$ ) until observation  $S_2$  (with probability  $1 - \beta_{S_2|k}$ ). To calculate  $\beta_{S_2|k}$ , one should know first how much the mean is shifted. Since the shift occurred at  $k$ , the shifted mean can be estimated by (1) or (2).

$$\hat{\beta}_{S_2|k} = \Phi\left(3 - \frac{\hat{\mu}_{S_2|k} - T}{\sigma}\right) - \Phi\left(-3 - \frac{\hat{\mu}_{S_2|k} - T}{\sigma}\right). \tag{5}$$

$P(K = k | K \in (S_1, S_2])$  is the probability that the shift occurs at  $k$  given that the shift is known to be in the lag-window. To calculate this probability, we need to consider the probability for a shift to occur at any time. Here, we first assume that the occurrence probability at any given time is constant  $p$ . Under this assumption, we can obtain:

$$P(k = k | K \in (S_1, S_2]) = \frac{p(1-p)^{S_2-S_1-1}}{\begin{bmatrix} S_2 - S_1 \\ 1 \end{bmatrix} p(1-p)^{S_2-S_1-1}} = \frac{1}{S_2 - S_1}. \tag{6}$$

Substituting (4) and (6) into (3), we have

$$P_{k|S_2} = \frac{\hat{\beta}_{S_2|k}^{S_2-k} (1 - \hat{\beta}_{S_2|k})}{\sum_{l=S_1+1}^{S_2} \hat{\beta}_{S_2|l}^{S_2-l} (1 - \hat{\beta}_{S_2|l})}. \tag{7}$$

We are now ready to estimate the process mean at the detection point:

$$\hat{\mu}_{S_2} = \sum_{K=S_1+1}^{S_2} \hat{\mu}_{S_2|K} P_{K|S_2}, \tag{8}$$

where  $\hat{\mu}_{S_2|K}$  is the estimated mean of  $X_{S_2}$  given a known change point at  $K$  and  $P_{K|S_2}$  is the probability for the shift to occur at  $K$  given a control chart detection at  $S_2$ .

When the process is not shut down after the control chart’s signaling and observation samples taken from the process are continued, then the process mean can be estimated for later observation points, say  $j$  and  $j > S_2$ , in the same manner:

$$\hat{\mu}_j = \sum_{K=S_1+1}^{S_2} \hat{\mu}_{j|K} P_{K|S_2}, \tag{9}$$

where

$$P_{K|S_2} = \frac{\hat{\beta}_{j|K}^{S_2-K} (1 - \hat{\beta}_{j|K})}{\sum_{l=S_1+1}^{S_2} \hat{\beta}_{j|l}^{S_2-l} (1 - \hat{\beta}_{j|l})} \tag{10}$$

and

$$\hat{\beta}_{j|K} = \Phi\left(3 - \frac{\hat{\mu}_{j|K} - T}{\sigma}\right) - \Phi\left(-3 - \frac{\hat{\mu}_{j|K} - T}{\sigma}\right).$$

As mentioned earlier, the mean estimate given in (9) is in essence a weighted average of the observations. When  $\hat{\mu}_{j|K}$  is estimated by  $\bar{X}_{j|K}$ , we can summarize the weights for each observation  $X_i$  as follows:

$$\hat{\mu}_j = \sum_{i=S_1+1}^{S_2-1} \sum_{K=S_1+1}^i \frac{P_{K|S_2}}{j - K + 1} X_i + \sum_{i=S_2}^j \sum_{K=S_1+1}^{S_2} \frac{P_{K|S_2}}{j - K + 1} X_i. \tag{11}$$

As an example, in the oxidation process the target SiO2 thickness is 250 Å. After preventive maintenance (PM), the process is in operation with process mean set to the target. Also, the thickness SD is known to be 1.5 Å. A Shewhart control chart is then constructed to monitor the oxidation condition. We have 70 observations. The thickness mean is shifted at the 50th observation with a magnitude 3.0 Å to become 253 Å. Figure 2 shows the corresponding Shewhart control chart. In the figure we assume that the true mean is known before the 50th observation.

As can be seen, at observation 57 the control chart gives an alarm. We now apply the mean estimate to this detection observation and the observations that follow. The lag-window includes observations 1 to 57. That is, the shift could have actually occurred at any of these observations. To compute the mean estimate for the detection observation 57, we first calculate sample means for  $\bar{x}_{57|57}, \bar{x}_{57|56}, \dots, \bar{x}_{57|1}$ . Corresponding to these sample means, we can estimate Type II error probabilities  $\beta_{57|57}, \beta_{57|56}, \dots, \beta_{57|1}$  using (5):

$$\beta_{57|K} = \Phi\left(3 - \frac{\hat{\mu}_{57|K} - 250}{1.5}\right) - \Phi\left(-3 - \frac{\hat{\mu}_{57|K} - 250}{1.5}\right) \text{ for } K = 1, \dots, 57.$$

The shift occurrence probability given signaling at 57 can then be estimated using (7):

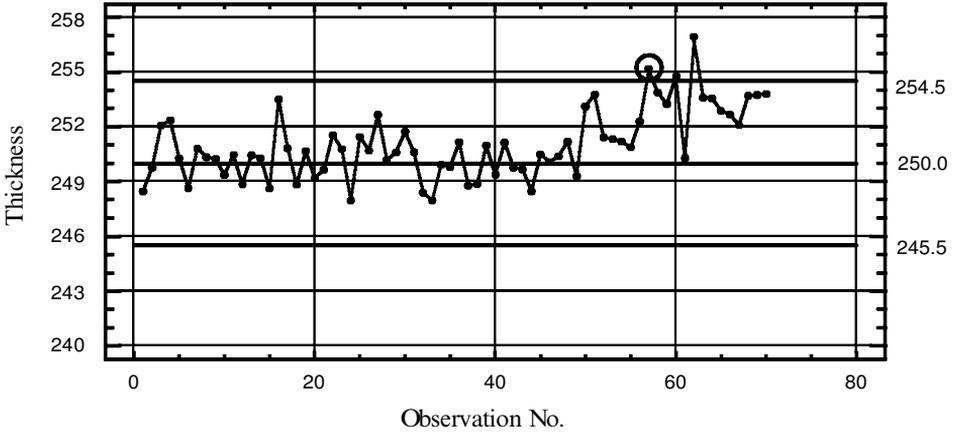


Figure 2. The Shewhart control chart for oxidation thickness.

$$P_{K|57} = \frac{\beta_{57|K}^{57-k}(1 - \beta_{57|K})}{\sum_{l=1}^{57} \beta_{57|l}^{57-l}(1 - \beta_{57|l})} \quad \text{for } K = 1, \dots, 57.$$

The mean estimate for observation 57 is calculated as

$$\hat{\mu}_{57} = \sum_{K=1}^{57} \bar{x}_{57|K} P_{K|57}.$$

The same procedure is repeated for estimating means for observations 58 and beyond. Figure 3 shows the estimated mean together with the observations.

One can observe from figure 3 that the mean estimate at the signal observation overshoots a little given only one observation available after detection. The mean

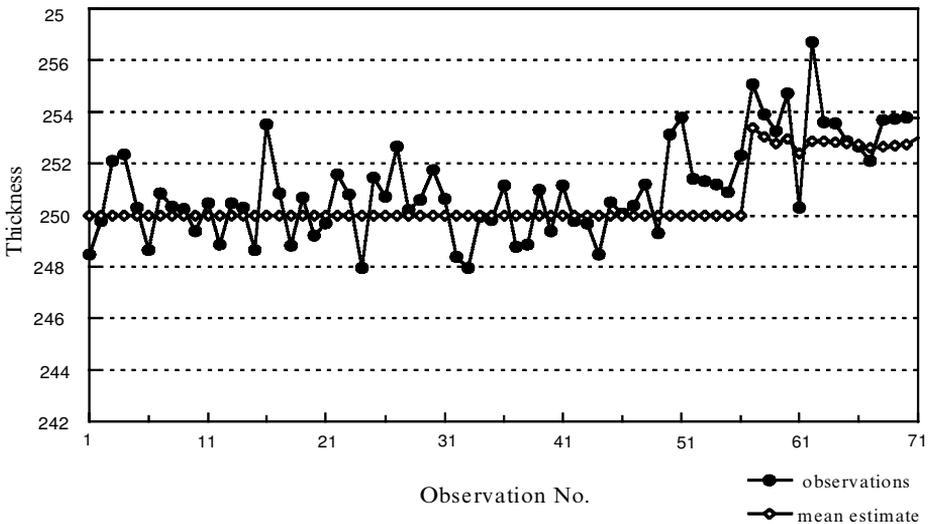


Figure 3. Mean estimate for Shewhart chart monitored process.

estimate is also seen to oscillate before it gradually stabilizes. Using the Bayes estimate given by (2), when enough experience has been gained, can significantly reduce the overshooting and the oscillation at the beginning of the estimation. Suppose that the shifted mean has learned to vary around 252.5 Å with S.D. 2.0 Å. Then

$$\hat{\mu}_{j|K} = \frac{4.0\bar{x}_{j|K} + \frac{2.25}{j-k+1}252.5}{\frac{2.25}{j-k+1} + 4.0}, \quad \text{for } K = 1, \dots, 57.$$

Figure 4 shows a comparison between mean estimates with and without the Bayes stabilization.

The Shewhart-based mean estimate is a weighted average of available observations. It is interesting to see how the weights vary over the age of the observation. Figure 5 shows the weights assigned by the estimator to the past observations.

From Figure 5, the proposed mean estimate looks very much like a dynamic weighted average of available observations. For example, for the detection observation's mean estimate, the weight taken on observation 57 itself is about 0.55 and then decreases exponentially for older observations. When the mean estimate continues for observations after signaling, the weight distribution is dynamically adjusted. While the weights exponentially decrease for those observations sampled before detection, the observations after signaling are equally weighted. Take observation 62 as an example, the mean estimate for this data point takes observations 57 to 61 with an equal weight about 0.1. However, weights for observations 56 and earlier decrease drastically. This kind of weight distribution seems quite plausible for the situation where the out-of-control occurrence takes only the form of mean shift and its occurrence is solely tested by the Shewhart chart.

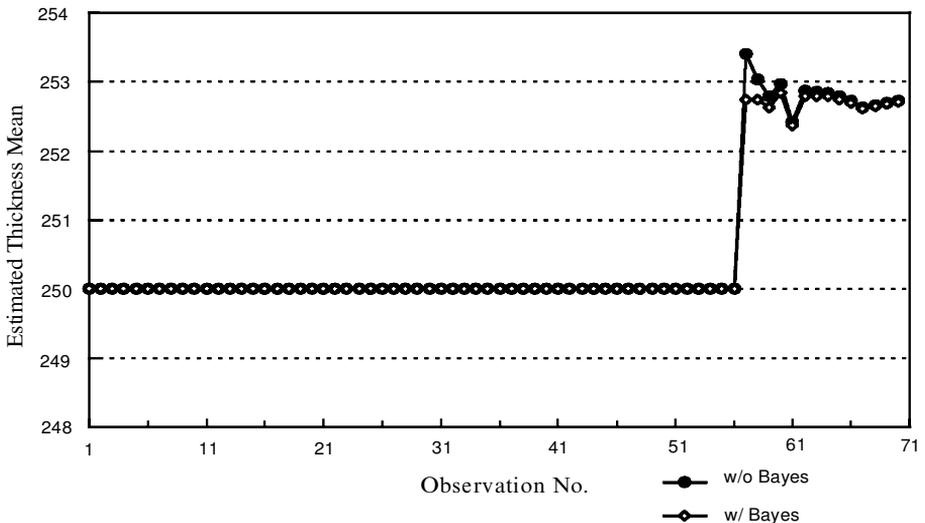


Figure 4. Mean estimates with and without Bayes stabilization.

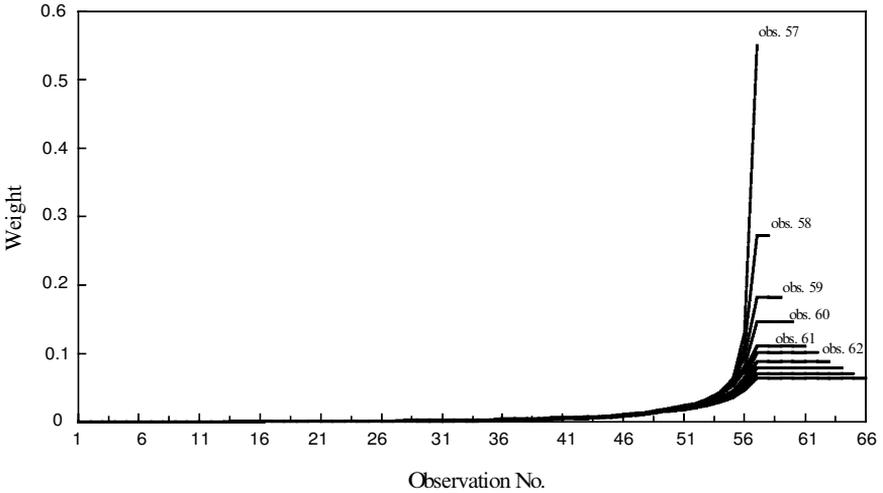


Figure 5. Weights for observations taken by Shewhart-based mean estimate.

**4. Mean estimate for other SPC charts and Shewhart chart with runs rules**

It is well known that the original Shewhart chart is not effective in detecting smaller process shifts. The popularity of the Shewhart chart lies not only on its simplicity but also on the development of customized runs rules. In addition to improving the control chart’s ability in detecting shifts, runs rules (*Statistical Quality Control Handbook* 1956) can be installed to identify patterns that are rather difficult to detect by the single out-of- $3\sigma$  rule.

The effectiveness of a Shewhart chart is usually evaluated by Type I and Type II error probabilities. When runs rules are applied to Shewhart charts or other types of SPC charts, such as EWMA or CUSUM charts (Montgomery 1996), are used, the testing is not only based on a single observation but also on past observations. Here, run lengths (RLs) are used instead to evaluate the control chart’s ability. While the in-control RL ( $RL_0$ ) assesses how frequently the control chart gives false alarms, the out-of-control RL ( $RL_1$ ) measures how quickly the control chart signals a process shift. To evaluate the RLs, one can apply the Markov chain approach to find its probability distribution (Brook and Evans 1972, Champ and Woodall 1987, and Lucas and Saccucci 1990).

As explained in the previous section, the out-of-control run length ( $RL_1$ ) probability given an approximated shift size is needed in order to calculate  $P_{k|S_2}$ . That is, we need to calculate

$$\begin{aligned}
 P_{S_2|k} &= P(\text{detection at } S_2 | K = k, K \in (S_1, S_2]) \\
 &= P(RL_1 = S_2 - k + 1 | \text{shift} = \hat{\mu}_{j|k} - T)
 \end{aligned}
 \tag{12}$$

for  $j \geq S_2$ .

Depending on the runs rules or types of control charts used, the calculation of the run length probability can become extensive. For example, it takes 216 states in the Markov chain to calculate the run length distribution when the four runs rules listed above are implemented simultaneously. As for EWMA and CUSUM charts, the larger the number of states, the more accurate the RL distribution. For on-line

production use, computer programs will be required for the calculation of RL distributions and mean estimate. For the calculation procedure using the Markov chain approach, please refer to the references given above.

Because of the complexity of RL distribution calculations, we propose an approximation based on the simple geometric distribution. That is

$$\beta_{j|k} \approx 1 - \frac{1}{ARL_1}$$

and

$$P_{S_2|k} = P(RL_1 = S_2 - k + 1 | shift = \hat{\mu}_{j|k} - T) = \left(1 - \frac{1}{ARL_1}\right)^{S_2-k} \frac{1}{ARL_1}, \quad j \geq S_2, \tag{13}$$

where  $ARL_1 = E(RL_1 = S_2 - k + 1 | shift = \hat{\mu}_{j|k} - T)$ . Since the calculations of average RLs (ARLs) are much simpler, usually by using an integral equation (Crowder 1989, Fellner 1990), than computing the entire distribution probabilities, this approximation will allow us to easily implement the proposed mean estimate.

The CUSUM chart is known for its superior ability in detecting small process shifts. The tabular CUSUM statistics are computed as follows:

$$C_i^+ = \max[0, x_i - (\mu_0 + k\sigma) + C_{i-1}^+]$$

$$C_i^- = \max[0, x_i - (\mu_0 + k\sigma) + C_{i-1}^-],$$

where the starting values  $C_0^+ = C_0^- = 0$ . If either  $C_i^+$  or  $C_i^-$  exceeds  $h\sigma$ , the process is considered out of control. To retain  $ARL_0$  around 370 and a relatively good performance of  $ARL_1, k$  and  $h$  are usually set equal to 1/2 and 4.5, respectively. The ARLs of this CUSUM scheme for different shift sizes can be easily approximated by the following formula (Siegmond 1985):

$$ARL \approx \begin{cases} \frac{\exp(-2\Delta b) + 2\Delta b - 1}{2\Delta^2} & \text{for } \Delta \neq 0 \\ b^2 & \text{for } \Delta = 0, \end{cases}$$

where  $\Delta = \delta - k$  for the upper one-sided CUSUM,  $b = h + 1.166$ , and  $\delta = (\text{expected shift} - \mu_0)/\sigma$ . To test our mean estimate approach for CUSUM charts, we use Siegmund's ARL formula and the geometric distribution approximation for the calculations of run length probabilities. We shall also compare our approach to a mean estimator (Montgomery 1996) usually used for the tabular CUSUM chart:

$$\hat{\mu} = \begin{cases} \mu_0 + k\sigma + \frac{C_i^+}{N^+} & \text{if } C_i^+ > H \\ \mu_0 + k\sigma + \frac{C_i^-}{N^-} & \text{if } C_i^- > H, \end{cases}$$

where  $N^+$  and  $N^-$  represent the number of consecutive periods that  $C_i^+$  or  $C_i^-$  have been nonzero.

The performance of the mean estimators is measured by mean squared deviation (MSD) of 50 mean estimates from the true mean after the first alarm of the CUSUM

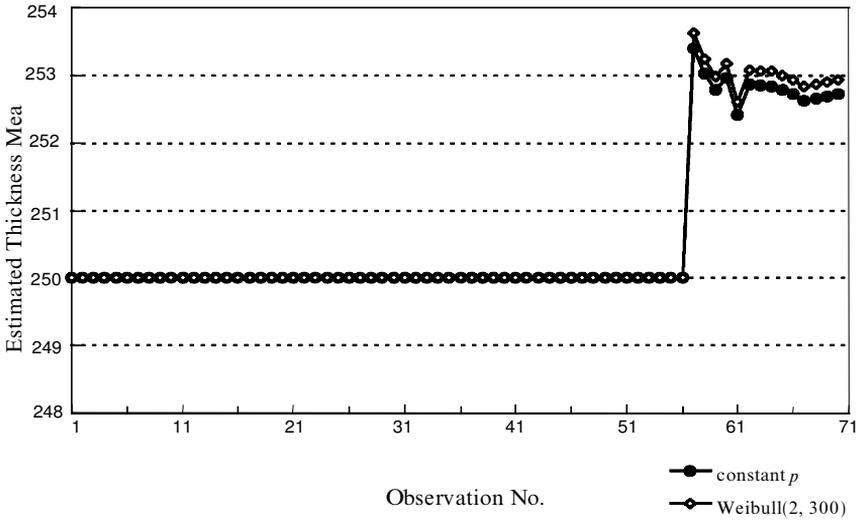


Figure 6. Mean estimates for constant shift occurrence probability and Weibull distributed inter-occurrence time.

chart. This mean squared deviation is then normalized against the process standard deviation  $\sigma$ :

$$\frac{\sqrt{MSD_{50}}}{\sigma} = \sqrt{\frac{\sum_{i=S_2}^{S_2+49} (\mu_i - \hat{\mu}_i)^2}{50\sigma^2}}$$

Only fifty observations after the out-of-control signal are taken to calculate the mean squared deviation because corrective actions are likely to be taken within these runs and the mean estimators become relatively steady after 50 observations. Also, since the CUSUM chart’s design is usually aimed for smaller process shifts. We compare the performance of mean estimators for shift sizes:  $0.25\sigma$ ,  $0.5\sigma$ ,  $0.75\sigma$ ,  $1\sigma$ ,  $1.25\sigma$  and  $1.5\sigma$ . Results, shown in figure 6, are based on 2000 simulation runs with standard errors less than 0.5%.

It can be seen from figure 6 that our mean estimator outperforms the CUSUM mean estimator for process shifts less than  $1.25\sigma$ . The estimator, however, suffers from unstable approximations of ARL and RL probability distribution and becomes less accurate for mean shifts greater than  $1.25\sigma$ . Since the CUSUM scheme is especially designed to detect smaller process shifts, the usual CUSUM mean estimator apparently lacks the estimation power for such shifts and our proposed estimator should be used instead. It can be also observed from figure 6 that our mean estimator is more robust than the CUSUM mean estimator especially for mean shifts less than  $0.5\sigma$ .

**5. Mean estimate for general shift occurrence mechanism**

We assume in the basic estimation procedure that the probability for a shift occurrence at any observation is constant  $p$ . We now consider a more realistic case where the probability of the shift occurrence is not constant and the shift inter-occurrence time follows a general continuous distribution with pdf  $f(t)$ . We

assume that the shift occurrence is a renewal process; that is, the time duration between consecutive shifts is positive, independent, identically distributed with *pdf*  $f(t)$ . Because of the availability of a large number of continuous distributions, the continuous time renewal process will allow us to model a wide range of shift occurrence mechanisms.

Under the above assumptions, the starting point of the inter-occurrence renewal cycle actually coincides with the starting point of the lag-window defined earlier; that is, the cycle starts at  $S_1 + 1$ . Then

$$P_{k|(S_1, S_2]} = P(K = k | K \in (S_1, S_2]) \\ = \frac{P(\text{a shift occurs in } (k - 1, k], \text{ say } t, \text{ and no shifts in } (t, S_2])}{P(N(hS_2 - hS_1) = 1)}$$

where  $h$  is the time interval between two consecutive observations and  $N(hS_2 - hS_1) = 1$  implies that the number of shifts in the interval  $(S_1, S_2]$  is one. Further results can be carried out as follows (Cox 1962):

$$P(N(hS_2 - hS_1) = 1) = P(N(hS_2 - hS_1) \geq 1) - P(N(hS_2 - hS_1) \geq 2) \\ = F(hS_2 - hS_1) - F_2(hS_2 - hS_1) \\ = \int_0^{hS_2 - hS_1} [1 - F(hS_2 - hS_1 - t)]f(t) dt$$

where

$$F(T) = \int_0^T f(t) dt = P(N(T) \geq 1) \text{ and } F_2(T) = \int_0^T F(T - t)f(t) dt = P(N(T) \geq 2)$$

Similarly, the probability that a shift occurs in  $(k - 1, k]$ , say at  $t$ , and no other shifts occur in  $(t, S_2]$  can be calculated as:

$$\int_{h(k-S_1-1)}^{h(k-S_1)} P(N(hS_2 - hS_1 - t) = 0)f(t) = \int_{h(k-S_1-1)}^{h(k-S_1)} [1 - F(hS_2 - hS_1 - t)]f(t) dt.$$

Thus

$$P_{k|(S_1, S_2]} = \frac{\int_{h(k-S_1-1)}^{h(k-S_1)} [1 - F(hS_2 - hS_1 - t)]f(t) dt}{\int_0^{hS_2 - hS_1} [1 - F(hS_2 - hS_1 - t)]f(t) dt} \tag{14}$$

When the inter-occurrence time is exponentially distributed with parameter  $\lambda$ , then

$$P_{k|(S_1, S_2]} = \frac{\int_{h(k-S_1-1)}^{h(k-S_1)} e^{-\lambda(hS_2 - hS_1 - t)} \lambda e^{-\lambda t} dt}{\lambda(S_2 - S_1) h e^{-\lambda(S_2 - S_1)h}} = \frac{1}{S_2 - S_1}.$$

This is the same result as that of (6). In fact, the constant probability of shift occurrence model results in a geometric distribution of the shift inter-occurrence time which is a discrete ‘version’ of the exponential distribution. Both of them share the ‘memoryless property’.

Now, we consider a more general case by substituting (12) and (14) into (3) to obtain

$$P_{k|S_2} = \frac{P_{S_2|k} P_{k|(S_1, S_2)}}{\sum_{l=S_1+1}^{S_2} P_{S_2|l} P_{l|(S_1, S_2)}}. \tag{15}$$

Finally, (9) can be rewritten for general cases where runs rules are used and the shift occurrence follows a renewal process:

$$\hat{\mu}_j = \sum_{K=S_1+1}^{S_2} \hat{\mu}_{j|K} \frac{P_{S_2|K} P_{K|(S_1, S_2)}}{\sum_{l=S_1+1}^{S_2} P_{S_2|l} P_{l|(S_1, S_2)}}, \tag{16}$$

where  $\hat{\mu}_{j|K}$  is from (1) or (2) and  $P_{S_2|K}$  and  $P_{K|(S_1, S_2)}$  are from (12) (or (13)) and (14), respectively.

Back to the semiconductor manufacturing problem: when an adjustment mechanism is present, the objective would be to adjust the process such that the mean of the thickness is as close as possible to the target. In the oxidation process, the furnace temperature is usually used as the adjustment variable. Given the estimate of the shifted mean in (16), the process engineer can now adjust the temperature to compensate for the thickness deviation  $\hat{\mu}_j - T$ .

Continuing the earlier example, suppose that the process engineer has collected information on the frequency of Shewhart chart signals, other than false alarms. The time interval between two successive signals can be properly treated and modeled to serve the modeling of shift inter-occurrence time. Weibull distribution is often used for modeling ‘time to failure’ of a piece of device or equipment and is also appropriate for modeling ‘time to shift’ of a process. The *pdf* of the Weibull distribution is:

$$f(t) = \begin{cases} \alpha\beta^{-\alpha} t^{\alpha-1} e^{-(t/\beta)^\alpha} & \text{if } t > 0 \\ 0 & \text{otherwise} \end{cases}$$

where  $\alpha$  is the shape parameter and  $\beta$  is the scale parameter. For the oxidation process, suppose that  $\alpha$  and  $\beta$  are estimated to be 2 and 300, respectively; i.e., the mean time to shift is about 266 hours. The specific chemical process required to deposit a layer of 250 Å of SiO<sub>2</sub> takes about 2.5 hours to complete; that is, thickness samples are taken every 2.5 hours ( $h = 2.5$ ). Using (14), we obtain:

$$P_{k|(0,57]} = \frac{\int_{2.5(k-1)}^{2.5k} (2)(300)^{-2} t \exp \left[ -\left(\frac{142.5 - t}{300}\right)^2 - \left(\frac{t}{300}\right)^2 \right]}{\int_0^{142.5} (2)(300)^{-2} t \exp \left[ -\left(\frac{142.5 - t}{300}\right)^2 - \left(\frac{t}{300}\right)^2 \right]}.$$

$P_{k|(0,57]}$ , together with the previously calculated  $P_{57|k}$  using  $\beta_{57|k}$  is substituted into (15), we obtain new mean estimates as shown in figure 7. Again, we assume that the true mean is known before the 50th observation.

In figure 7 the newly estimated means are contrasted to the mean estimates assuming constant shift occurrence probabilities. As can be seen, the estimated mean also overshoots at the beginning of the process but stabilizes closer to 253 Å. This is because the weights assigned to the more recent data have shifted upward as shown in figure 8.

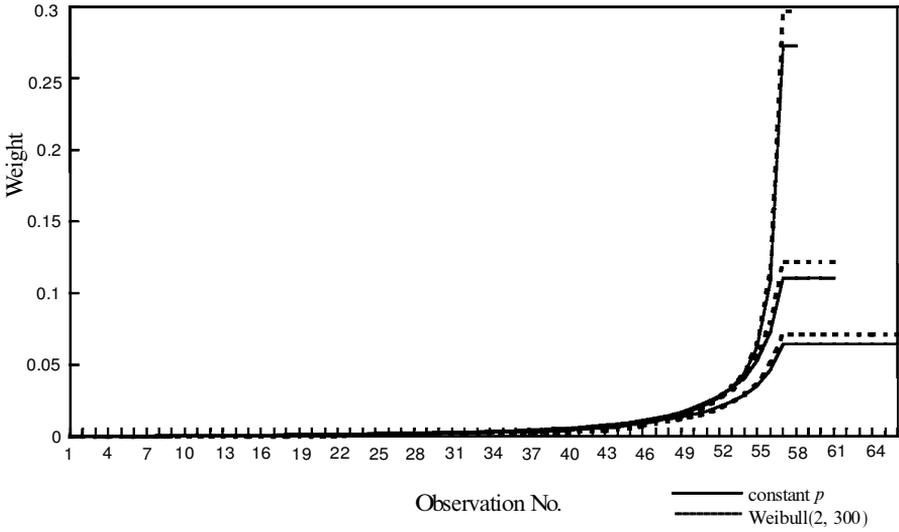


Figure 7. Weights taken by mean estimates for constant shift occurrence probability and Weibull distributed inter-occurrence time.

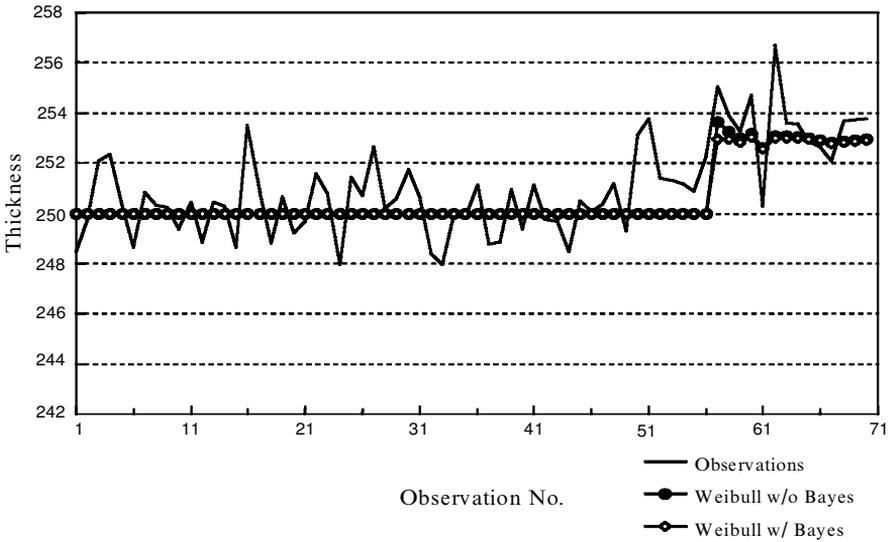


Figure 8. Mean estimates for Weibull distributed inter-occurrence time with Bayes stabilization.

Unlike the constant occurrence probability, the occurrence probability for the Weibull model increases as time elapses. That is, more recent data are assigned more weights as described in (11) and (15) and demonstrated in figure 8.

Finally, the Bayes estimate can be used again to reduce the overshooting and fluctuation at the beginning of the estimation.

## 6. Conclusions

In this paper we have developed methodologies to estimate the process mean for processes monitored by SPC charts to test possible process shifts. The proposed mean estimate was actually a dynamically weighted average of the historical data. It estimated the SPC chart's test sensitivity and dynamically adjusted the weights of the historical data. The mean estimate was further refined by utilizing prior knowledge on the shift occurrence rate and shift size. The use of the mean estimate was then demonstrated through an example of the semiconductor oxidation process. The performance of the proposed mean estimator was shown to outperform a CUSUM mean estimator for shift sizes less than  $1.25\sigma$ .

## References

- ÅSTRÖM, K. J., 1970, *Stochastic Control Theory* (New York: Academic Press).
- BARNARD, G. A., 1959, Control charts and stochastic processes. *Journal of the Royal Statistics Society, Series B*, **21**, 239–257.
- BONING, D., HURWITZ, A., MOYNE, J., MOYNE, W., SMITH, T., TAYLOR, J. and TELFEYAN, R., 1995, Run by run control of chemical mechanical polishing. *IEEE Transactions on Components, Packaging and Manufacturing Technology — Part C*, **19**, 307–314.
- BOX, G. E. and JENKINS, G. M., 1968, Some recent advances in forecasting and control. *Applied Statistics*, **17**, 91–109; 1970, *Time Series Analysis: Forecasting and Control* (San Francisco: Holden-Day).
- BOX, G. E. and KRAMER, T., 1992, Statistical process monitoring and feedback adjustment – a discussion. *Technometrics*, **34**, 251–285.
- BROOK, D. and EVANS, D. A., 1972, An approach to the probability distribution of CUSUM run length. *Biometrika*, **59**, 539–549.
- CHAMP and WOODALL, 1987, Exact results for Shewhart control charts with supplementary runs rules. *Technometrics*, **29**, 393–399.
- CHERNOFF, H. and ZACKS, S., 1964, Estimating the current mean of a normal distribution which is subject to change in time. *Annals of Mathematics and Statistics*, **35**, 999–1018.
- COX, D. R., 1962, *Renewal Theory* (New York: Chapman and Hall).
- CROWDER, S. V., 1989, Design of exponentially weighted moving average schemes, **21**, 155–162.
- DAVIS, M. H. A. and VINTER, R. B., 1985, *Stochastic Modeling and Control* (New York: Chapman & Hall).
- FELLNER, W. H., 1990, Algorithm AS 258. Average run length for cumulative sum schemes. *Applied Statistics*, **39**, 402–412.
- HARRISON, P. J. and STEVENS, C. F., 1971, A Bayesian approach to short-term forecasting. *Operational Research Quarterly*, **22**, 341–362; 1976, Bayesian forecasting. *Journal of Royal Statistics Society, Series B*, **38**, 205–247.
- HOERL, R. W. and PALM, A., 1992, Discussion: integrating SPC and APC. *Technometrics*, **34**, 268–272.
- INGOLFSSON, A. and SACHS, E., 1993, Stability and sensitivity of an EWMA controller. *Journal of Quality Technology*, **25**, 271–287.
- KELTON, W. D., WALTON, M., HANCOCK, M. and BISCHAK, D. P., 1990, Adjustment rules based on quality control charts. *International Journal of Production Research*, **28**, 385–400.
- LEHMANN, E. L., 1991, *Theory of Point Estimation* (Pacific Grove, California: Wadsworth), pp. 236–249.
- LUCAS, J. M. and SACCUCI, M. S., 1990, Exponentially weighted moving average control schemes: properties and enhancements. *Technometrics*, **32**, 1–27.
- MACGREGOR, J. F., 1987, Interfaces between process control and online statistical process control. *Computing and Systems Technology Division Communications*, **10**, 9–20; 1988, On line statistical process control. *Chemical Engineering Progress*, Oct., 21–31.
- MONTGOMERY, D. C., 1996, *Introduction to Statistical Quality Control* (New York: Wiley).
- PAGE, E. S., 1954, Continuous inspection schemes. *Biometrika*, **41**, 100–116; 1955, A test for a change in a parameter occurring at an unknown point. *Ibid.*, **42**, 523–526.

- SACHS, E., HU, A. and INGOLFSSON, A., 1995, Run by run process control: combining spc and feedback control. *IEEE Transactions on Semiconductor Manufacturing*, **8**, 26–43.
- SIEGMUND, D., 1985, *Sequential Analysis: Tests and Confidence Intervals* (New York: Springer-Verlag).
- SHEWHART, W. A., 1931, *Economic Control of Quality Improvement* (Princeton: Van Nostrand Reinhold).
- NISHINA, K., 1992, A comparison of contro charts from the viewpoint of change-point estimation. *Quality and Reliability Engineering International*, **8**, 537–541.
- TUCKER, W. T., FALTIN, F. W. and VANDER WIEL, S. A., 1993, Algorithmic statistical process control: an elaboration. *Technometrics*, **35**, 363–375.
- VANDER WIEL, S. A., TUCKER, W. T., FALTIN, F. W. and DOGANAKSOY, N., 1992, Algorithmic statistical process control: concepts and an application. *Technometrics*, **34**, 286–297.
- Western Electric Company, 1958, *Statistical Quality Control Handbook* (New York: Western Electric Company).
- WOODALL, W. H. and REYNOLDS, M. R. R., Jr., 1983, A discrete Markov chain representation of the sequential probability ratio test. *Communications in Statistics — Sequential Analysis*, **2**, 27–44.
- YASHCHIN, E., 1995, Estimating the current mean of a process subject to abrupt changes. *Technometrics*, **37**, 311–323.