Appl. Statist. (1998)
47, *Part* 2, *pp.* 203–216

Grouped data exponentially weighted moving average control charts

Stefan H. Steiner†
University of Waterloo, Canada

[Received October 1996. Final revision April 1997]

Summary. In the manufacture of metal fasteners in a progressive die operation, and other industrial situations, important quality dimensions cannot be measured on a continuous scale, and manufactured parts are classified into groups by using a step gauge. This paper proposes a version of exponentially weighted moving average (EWMA) control charts that are applicable to monitoring the grouped data for process shifts. The run length properties of this new grouped data EWMA chart are compared with similar results previously obtained for EWMA charts for variables data and with those for cumulative sum (CUSUM) schemes based on grouped data. Grouped data EWMA charts are shown to be nearly as efficient as variables-based EWMA charts and are thus an attractive alternative when the collection of variables data is not feasible. In addition, grouped data EWMA charts are less affected by the discreteness that is inherent in grouped data than are grouped data CUSUM charts. In the metal fasteners application, grouped data EWMA charts were simple to implement and allowed the rapid detection of undesirable process shifts.

Keywords: Cumulative sum; Exponentially weighted moving average; Grouped data; Process shifts

1. Introduction

In quality control, exponentially weighted moving average (EWMA) control charts are used to monitor process quality. EWMA charts, and other sequential approaches like cumulative sum (CUSUM) charts, are alternatives to Shewhart control charts that are especially effective in detecting small persistent process shifts. Although introduced by Roberts (1959), EWMA charts have only recently had their properties evaluated analytically (Crowder, 1987; Lucas and Saccucci, 1990). The EWMA also has optimal properties in some forecasting and control applications (Box *et al.*, 1974).

For monitoring a process, an EWMA control chart consists of plotting

$$z_t = \lambda x_t + (1 - \lambda) z_{t-1}, \qquad 0 < \lambda \leqslant 1, \tag{1}$$

versus time t, where x_t is an estimate of the process characteristic that we wish to monitor, λ is a constant weighting factor and the starting value z_0 equals an *a priori* estimate of the parameter of the monitored process. In equation (1), x_t may represent the sample mean, sample standard deviation or any other empirically estimated process parameter. When the recursion in equation (1) is written out, the EWMA test statistic z_t equals an exponentially weighted average of all previous observations, i.e.

†*Address for correspondence*: Department of Statistics and Actuarial Sciences, University of Waterloo, Waterloo, Ontario, N2L 3G1, Canada.
E-mail: shsteine@setosa.uwaterloo.ca

 0035–9254/98/47203

$$z_t = \lambda x_t + \lambda(1-\lambda)x_{t-1} + \lambda(1-\lambda)^2 x_{t-2} + \ldots + (1-\lambda)^t z_0.$$

In contrast, tabular CUSUM charts assign equal weight to all past observations since the CUSUM statistic last equalled 0 (Montgomery (1991), section 7.2). In quality monitoring applications of EWMA control charts, typical values for the weight λ are between 0.05 and 0.25.

From equation (1), the mean and variance of z_t, denoted μ_{z_t} and $\sigma^2_{z_t}$ respectively, are easily derived (Montgomery, 1991). Assuming that the x_t are independent random variables with mean μ_x and variance σ^2_x gives

$$\mu_{z_t} = \mu_x$$

and

$$\sigma^2_{z_t} = \sigma^2_x \left(\frac{\lambda}{2-\lambda}\right) \{1-(1-\lambda)^{2t}\} \cong \sigma^2_x \left(\frac{\lambda}{2-\lambda}\right) \qquad \text{as } t \to \infty. \tag{2}$$

Control limits for an EWMA control chart are typically derived on the basis of $\pm L$ σ-limits, where L is usually equal to 3, as in the design of Shewhart control chart limits. The fact that the z_t are not independent is ignored. Thus, the control limits of an EWMA chart used to monitor the process mean are

$$\mu_{z_t} \pm L\sigma_{z_t} = \mu_x \pm L\sigma_x \sqrt{\left[\left(\frac{\lambda}{2-\lambda}\right)\{1-(1-\lambda)^{2t}\}\right]}, \tag{3}$$

where, in applications, μ_x and σ_x are typically estimated from preliminary data as the sample mean and sample standard deviation respectively. The process is considered to be out of control whenever the EWMA test statistic z_t falls outside the range of the control limits given by equation (3). In the limit with $\lambda = 1$, the EWMA chart is identical with a Shewhart $\bar{X}$ control chart.

In some industrial situations, however, collecting variables data on critical quality dimensions is either impossible or prohibitively expensive, and the data are grouped. The widespread occurrence of binomial pass–fail attribute data in industry attests to the economic advantages of collecting pass–fail data over exact measurements. In general, variables data provide more information, but gauging, or classifying observations into groups based on a critical dimension, is often preferred as it takes less skill, is faster, is less costly and is a tradition in certain industries (Schilling (1981), p. 333, and Ladany (1976)). Grouped data are a natural compromise between the low data collection and implementation costs of binomial data and the high information content of variables data. Grouped data occur in industry because of multiple pass–fail gauges, step gauges or other similar measurement devices (Steiner *et al.*, 1994). A step gauge with $k-1$ gauge limits yields k-group data. Pass–fail binomial attribute data represent the special case of two-group data. For more information on grouped data, see Haitovsky (1982).

EWMA control charts are designed for variables data and it is not clear how to adapt the charts to handle grouped data, or what effect grouped data may have on the run length properties of EWMA charts.

The development of control charting methodology for use with grouped data other than binomial data started with Stevens (1948), who proposed two simple *ad hoc* Shewhart control charts for simultaneously monitoring the mean and standard deviation of a normal distribution using three-group data. Beja and Ladany (1974) proposed using three-group data to test for one-sided shifts in the mean of a normal distribution with known process dispersion. In the methodology of sequential quality control, Schneider and O'Cinneide (1987) proposed a CUSUM scheme for monitoring the mean of a normal distribution with two-group data. Geyer *et al.* (1996) extended this CUSUM to the use of three-group data, with gauges symmetric about the midpoint between the target mean and the out-of-control mean that the chart should detect quickly. Gan (1990) proposed a modified EWMA chart for use with binomial data. The modified form of the EWMA uses equation (1) but rounds off the EWMA test statistic and calculates the run length properties by using a Markov chain. Unfortunately, this solution approach is appropriate only when many failures are expected in a sample. As a result, the solution procedure typically requires large samples, especially when the probability of failure is small. Steiner *et al.* (1994, 1996a) were the first to consider the general k-group case. They developed methodology for one-sided and two-sided acceptance sampling plans, acceptance control charts and Shewhart-type control charts. In addition, Steiner *et al.* (1996b) considered k-group sequential probability ratio tests and CUSUM procedures. These k-groups control charts use the likelihood ratio to derive an efficient test statistic. Steiner *et al.* (1994, 1996a, b) also gave design procedures for the various types of k-group control charts, calculated run length properties and addressed the question of optimal gauge design. These references show that k-group control charts are efficient alternatives to standard variables-based techniques.

This paper addresses the derivation of the general k-group EWMA control chart. Grouped data EWMA procedures bridge the gap between the efficiency of binomial attribute procedures and that of variables-based EWMA charts. An important matter, addressed later, is how this loss of information in the data affects the performance of the grouped data EWMA chart in comparison with variables-based EWMA charts. For grouped data, EWMA charts may be a better choice than a CUSUM chart since, because of the exponential weighting of past observations, the EWMA smooths out the inherent discreteness. This is an advantage that allows more flexibility in the design of grouped data EWMA charts than with grouped data CUSUM charts.

A good example of grouped data in industry occurs in the manufacture of metal fasteners in a progressive die environment, where the opening gap dimension of a metal clamp, called a robotics clamp (Fig. 1), was considered critical. This problem was previously considered by Steiner *et al.* (1994, 1996a).

Obtaining exact measurements of the gap dimension on the shop-floor was prohibitively difficult and expensive. The metal used in the clamp is fairly pliable, and as a result using calipers distorts the opening gap dimension. Another alternative, an optical measuring device, is expensive and not practical for on-line quality monitoring. As a result, the only

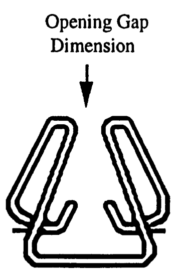

Fig. 1. Robotics clamp

economical alternative on the shop-floor is to use step gauges, where clamps are classified into different groups based on the smallest diameter pin that the clamp's opening gap does not fall through. The step gauge employed consisted of three pins of diameters 53, 54 and 55 thousandths of an inch. Using the given step gauge, units are classified into four intervals with corresponding interval midpoint weights of 52.5, 53.5, 54.5 and 55.5

From previous measurements, the process mean is known to be currently stable, producing clamps with an average open gap dimension of 54.2 thousandths of an inch (μ_x) and standard deviation of 1.3 (σ_x). We wish to monitor the stability of the mean width of the opening gap. Steiner *et al.* (1996a) proposed a grouped data Shewhart control chart that has an in-control average run length (ARL) approximately equal to 370, and an out-of-control ARL, at a mean shift of half a standard deviation unit, of approximately 15.5 and 12.7 for positive shifts and negative shifts respectively. This was accomplished with a sample of size 12 units. Since this is a fairly small process shift we would expect to do better with an EWMA chart.

This paper is organized in the following manner. Section 2 discusses two possible grouped data scoring procedures and recommends unbiased estimate scores for EWMA charts. In Sections 3 and 4, EWMA control charts for the k-group case are developed and the run length properties of grouped data and variables data-based EWMA charts and grouped data CUSUM charts are compared. Section 5 discusses in more detail the metal fasteners example that motivated this work and Section 6 briefly discusses optimal gauge placement. Appendix A shows how the run length distribution of grouped data EWMA charts can be approximated by using a Markov chain.

2. Sequential scoring procedures for grouped data

When using grouped data in control charts, the need arises to assign the grouped observations a numerical value based on their grouping. For pass–fail gauges, observations are usually treated singly as Bernoulli random variables. However, when observations are grouped into multiple intervals, a number of different scoring or weighting procedures are feasible. This paper considers two scoring schemes, namely midpoint scores and unbiased estimate scores. In industry, group interval midpoints are used most often. However, as will be shown, midpoint weights have some undesirable properties, and if some additional process information is available unbiased estimate scores are a better choice.

Throughout this paper, it is assumed that, although the data are grouped, there is an underlying continuous measurement that is unobservable. Let X represent the underlying measurement, and let $t_1 < t_2 < \ldots < t_{k-1}$ denote the $k-1$ end points or gauge limits used to derive the k-group data. We assume, for the moment, $k-1$ predetermined gauge limits. In many applications, the grouping criterion is fixed since it is based on some standard classification device or procedure. Section 6 addresses the relaxation of this assumption. Assume that the random variable X has probability distribution $f(x; \theta)$ and cumulative distribution function $F(x; \theta)$, where θ is the process parameter of interest. Let w_j be the group weight or score assigned to all observations falling into the jth group. Defining $t_0 = -\infty$ and $t_k = \infty$, the probability that an observation falls into the jth interval is given by

$$\pi_j(\theta) = F(t_j; \theta) - F(t_{j-1}; \theta), \qquad j = 1, 2, \ldots, k. \tag{4}$$

Ideally, the group weights chosen have a physical interpretation. This makes the interpretation of the resulting control charts easier for the industrial personnel. This implies that the weight for each group should lie somewhere between the group gauge limits. Strictly applying this criterion precludes the use of likelihood ratio weights as suggested by Steiner *et al.* (1996a) in the Shewhart control chart context.

Furthermore, often the group weights are used not only in a control chart but also to estimate the current process mean and variance so that we can calculate process capability measures. Typically the process mean and variance are estimated as the sample mean and variance of the group weights. As a result, the properties of these estimates are of interest. Ideally the sample mean and variance are unbiased estimates of the true process parameters. However, this is not possible with group data for all true parameter values. For any weighting scheme w_j, the expected value of the process mean estimate and process standard deviation estimate at the parameter value θ are respectively

$$\begin{aligned} E(w) &= \hat{\mu} = \sum_{j=1}^{k} w_j \, \pi_j(\theta), \\ \operatorname{var}(w) &= \hat{\sigma}^2 = \sum_{j=1}^{k} w_j^2 \, \pi_j(\theta) - \hat{\mu}^2 \end{aligned} \tag{5}$$

where $\pi_j(\theta)$ is given by equation (4).

These parameter estimates can be substantially different from the true process values. Naturally, any bias in parameter estimation adversely affects process capability calculations and our understanding of the process.

The midpoint approach is a very simple scoring procedure and is often used in industry. Each observation falling into a particular group is assigned a score equal to the group interval midpoint. With gauge limits $\mathbf{t}$, midpoint group weights are given as

$$w_j^{(\mathrm{m})} = \begin{cases} (3t_1 - t_2)/2 & \text{for } j = 1, \\ (t_{j-1} + t_j)/2 & \text{for } 2 \leqslant j \leqslant k-1, \\ (3t_{k-1} - t_{k-2})/2 & \text{for } j = k. \end{cases} \tag{6}$$

The midpoint scheme is attractive because it is very simple, and the scores retain a clear physical meaning. In addition, the midpoint scores can be determined without knowledge of the underlying process distribution. However, calculating the sample mean and variance of the midpoint scores can yield biased estimates of the true process mean and variance. Using equations (5) with group weights defined in equation (6) we may derive the expected bias in the estimates of the process mean and variance. Fig. 2 shows the results for a range of true process mean values and $\mathbf{t} = [-2, -1, 0, 1, 2]$. Fig. 2 illustrates that, using the midpoint weights when the process is in control, the sample mean is an unbiased estimate of the process mean (when the gauge limits are placed symmetrically), but the process variance is typically overestimated.

The midpoint score approach has further difficulty as intervals that extend to $-\infty$ or ∞ do not have true midpoints. In definition (6), end groups are assigned scores based on the most extreme gauge limits and the distance to the second most extreme scores on either side. Clearly, although this approach seems reasonable if the groups are of equal width, other definitions of the weights are possible.

The unbiased estimate weights are derived so that, when the process is in control, the expected sample mean and variance are unbiased estimates of the process mean and variance. However, these two conditions do not specify unique weights. As a result, the recommended unbiased estimate weights also have the smallest sum of squared bias terms at μ_1 and μ_{-1},

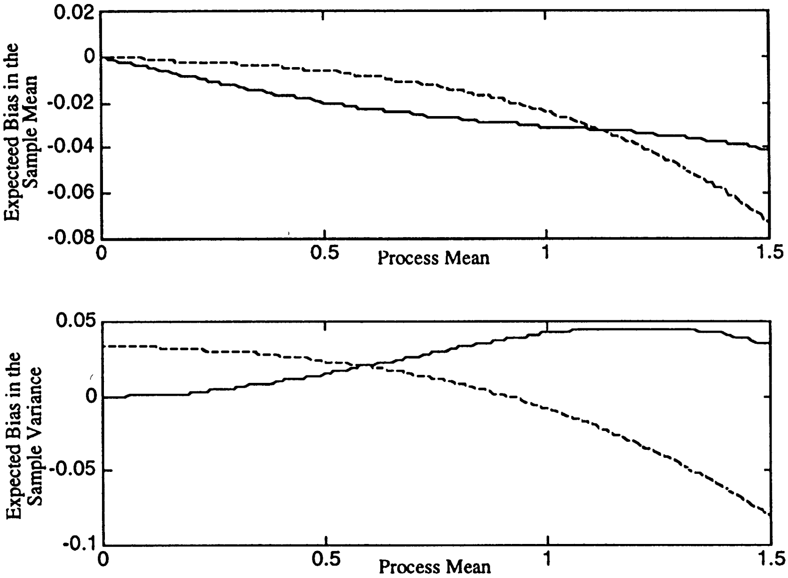

Fig. 2. Expected bias in the parameter estimates: ———, unbiased estimate weights with $\mu_1 = 0.5$ and $\mu_{-1} = -0.5$; - - - -, midpoint weights ($\mathbf{t} = [-2, -1, 0, 1, 2]$)

where μ_1 and μ_{-1} are process mean values on each side of the null that we wish to detect quickly. Thus, the unbiased estimate weights are derived as the $w^{(u)}$ weights that minimize

$$\{E(w^{(u)}|\mu_1) - \mu_1\}^2 + \{E(w^{(u)}|\mu_{-1}) - \mu_{-1}\}^2 \tag{7}$$

subject to

$$E(w^{(u)}|\mu_0) = \mu_0,$$

$$\text{var}(w^{(u)}|\mu_0) = \sigma_0^2.$$

For example, when $\mathbf{t} = [-2, -1, 0, 1, 2]$ and choosing $\mu_1 = 0.5$ the unbiased estimate weights are −2.8, −1.4, −0.4, 0.4, 1.4 and 2.8. These weights are different from the midpoint weights −2.5, −1.5, −0.5, 0.5, 1.5 and 2.5.

Fig. 2 shows the expected bias of the midpoint and unbiased estimate weights for various values of the true process mean μ when the underlying process is normally distributed with variance equal to 1. In control, i.e. at $\mu = 0$, both methods yield unbiased estimates of the process mean; however, the estimate of the process variance is biased for the midpoint approach.

Group weights defined as the conditional expected value of an observation that falls into a particular group, as given by $w_j^{(c)} = E(X|X \in j\text{th group}, \mu = \mu_0)$, were also considered. However, conditional expected value weights have a negative bias when estimating the process variance while $\mu = 0$ and generally introduce more bias into both the process mean and the variance estimates as the actual process mean changes than either the midpoint or unbiased estimate weights. As a result, conditional expected value weights are not considered further in this paper.

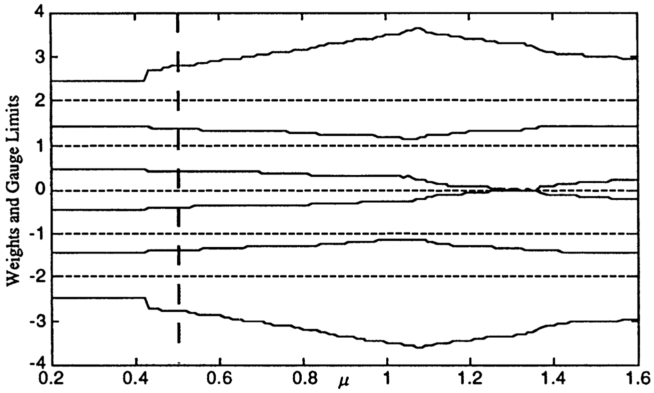

Fig. 3. Unbiased estimate weights for **t** = [−2, −1, 0, 1, 2]

Calculating the unbiased estimate weights requires knowledge of the underlying process distribution, or at least information about the group probabilities at both μ_0 and μ_1. The unbiased estimate weights are desirable when the group weights are used to estimate process measures such as process capability directly. Fig. 3 gives an example of how the optimal weights derived according to expression (7) change with different values of μ_1 $(= -\mu_{-1})$ when $\mathbf{t} = [-2, -1, 0, 1, 2]$. The broken vertical line shows the resulting weights when $\mu_1 = 0.5$.

As group weights are often used both in control charts and in process capability calculations, the unbiased estimate scoring approach is recommended if a good understanding of the underlying distribution is available. Otherwise, the group midpoint approach provides reasonably good results. However, the solution methodology that will be used to derive group data EWMA control charts works with any scoring procedure as long as each observation that falls into a particular group is assigned the same weight.

3. Exponentially weighted moving average control charts with grouped data

The proposed EWMA control charts for grouped data are based on expression (1), where x_t equals the average group weight assigned to a sample. When the process is being monitored for mean shifts, the group weights can be given by any weighting procedure that retains the group ordering, such as one of the possibilities discussed in Section 2. Using grouped data there are only finitely many possible average group weights, the number of weights being based on the number of groups utilized and the sample size. Thus there are finitely many possibilities for x_t. The test statistic z_t in equation (1), however, also depends on the previous value z_{t-1}. As a result, the repeated use of formula (1) smooths out the discreteness that is inherent in the observed values. This section addresses the questions of grouped data EWMA control chart design and performance.

The performance of EWMA control charts is usually discussed in terms of the run length. Crowder (1987) used an integral equation approach to derive the run length properties of EWMA control charts based on variables data. Crowder gave tables of run length results for various combinations of the parameters λ and L in equations (1) and (3). Unfortunately, Crowder's approach cannot handle the discreteness that is inherent in the grouped data EWMA case. An alternative solution procedure, presented in Appendix A, involves modelling the situation with a Markov chain. For control charts, the Markov chain solution

Table 1. ARLs for two-sided grouped data EWMA charts, $X \sim N(0, 1)$

$\sqrt{n}\mu_x/\sigma_x$	*ARLs for the following conditions:*							
	Continuous		t = [−2, −1, 0, 1, 2]		t = [−1, 0, 1]		t = [−1, 1]	
	$\lambda = 0.25$, $L = 2.998$	$\lambda = 0.10$, $L = 2.814$	$\lambda = 0.25$, $L = 2.991$	$\lambda = 0.10$, $L = 2.802$	$\lambda = 0.25$, $L = 2.821$	$\lambda = 0.10$, $L = 2.763$	$\lambda = 0.25$, $L = 2.981$	$\lambda = 0.10$, $L = 2.837$
0.0	500	500	498	500	511	498	515	487
±0.5	48	31	52	34	53	35	63	41
±1.0	11.1	10.3	12.1	11.0	13.1	12.1	14.9	13.0
±1.5	5.5	6.1	6.0	6.6	7.0	7.7	7.4	7.8
±2.0	3.6	4.4	4.1	4.8	5.1	6.1	5.3	6.1
±3.0	2.3	2.9	3.1	3.5	4.1	5.1	4.1	5.1
±4.0	1.7	2.2	3.0	3.1	4.0	5.0	4.0	5.0

approach was first developed by Page (1954) to evaluate the run length properties of a CUSUM chart. Grouped data with their inherent discreteness appear well suited to the Markov approach, since the Markov framework requires a discretization of the state space. The proposed Markov chain solution methodology can also provide approximate solutions for EWMA control charts based on variables data, and the method proposed was used to verify the results reported in Crowder (1987).

Table 1 gives ARL values for the EWMA charts based on variables (continuous) data, and EWMA control charts for different grouping criteria, given by t. The EWMA charts are designed to detect shifts in the mean of a normal process with an in-control mean of 0 and variance equal to 1. The process shifts are given in standard error units ($\sigma_x/\sqrt{n} = 1/\sqrt{n}$). The group data EWMA charts are designed to match the in-control ARL of the variables-based EWMA as closely as possible, but because of the discreteness of the group weights an exact match is not always possible. The run lengths shown in Table 1 are all derived assuming that the EWMA starts in the zero state when the process shift occurs. Steady state results provide a more realistic approximation. However, as shown by Lucas and Saccucci (1990), the zero state and steady state run lengths are very similar. Fig. 4 plots the results from Table 1 on a log-scale. The results in Table 1 are generated for the unit sequential case, i.e. $n = 1$. For

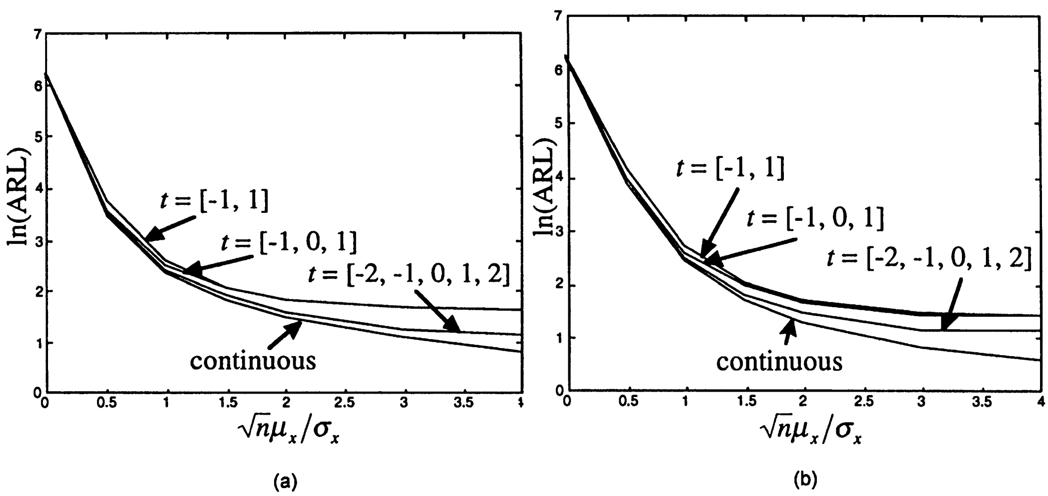

Fig. 4. ARL plots comparing grouped data EWMAs with variables data EWMAs: (a) $\lambda = 0.10$; (b) $\lambda = 0.25$

larger sample sizes the grouped data results are slightly better, since there is less discreteness, but the difference is not substantial for small sample sizes.

Table 1 and Fig. 4 show, as expected, that the grouped data EWMA charts are not as effective as a variables-based EWMA for detecting process mean shifts. This decrease in efficiency is due to the information lost when grouping the data. However, the loss of efficiency in detecting fairly small process mean shifts (say of the order of one standard deviation unit) is quite small. For very large process shifts, in contrast, the grouped data charts perform poorly since there is a maximum weight value that any observation can take. In applications, EWMA charts are typically used to detect fairly small process shifts. This suggests that grouped data EWMA charts are a viable alternative when collecting variables data is prohibitively expensive or impossible.

All EWMA control charts have two design parameters, namely λ and L, as defined in equations (1) and (3). Often EWMA charts are designed by specifying a desired in-control run length and the magnitude of the shift in the process that is to be detected quickly. Lucas and Saccucci (1990) provided a look-up table of optimal parameter values for the variables data case. Alternatively, Crowder (1987) provided extensive tables that are also useful for determining good initial values for λ and L. The same general procedure is suggested for grouped data charts. However, because of the inherent discreteness, the desired ARLs may not be precisely attainable. As not all state values are attainable, changes to L do not necessarily change the ARL of the EWMA. Generally, small λ-values are good for detecting small process shifts but are poor for larger shifts, and vice versa for large λ. Using the solution methodology presented in Appendix A, n and L are adjusted until the desired in-control and out-of-control ARLs are closely met. Large values of L lead to large ARLs, whereas increasing the sample size n decreases the out-of-control ARL and the problem discreteness. A step-by-step design procedure is given below.

(a) Find the suggested optimal λ- and L-values for EWMA charts based on continuous data from Lucas and Saccucci (1990). Set the sample size n equal to 1.
(b) Keeping λ fixed, adjust L until the desired in-control ARL is attained. The methodology presented in Appendix A can be used to find the in-control ARL for each combination of λ and L.
(c) Determine the out-of-control ARL at the current values of n, λ and L.
(d) If the desired out-of-control ARL is exceeded, increment n, and repeat this procedure starting at step (b).

It is possible that using this iterative design procedure the sample size becomes too large to be practical. If that occurs, the desired run length properties are too stringent and cannot be achieved economically. To alleviate this problem either the desired in-control ARL must be decreased or the acceptable out-of-control ARL must be increased.

4. Comparison with cumulative sum procedures for grouped data

Both EWMA charts and CUSUM charts are designed to detect small persistent process shifts. Past researchers (Lucas and Saccucci, 1990) found that there is very little difference between EWMA and CUSUM procedures in terms of the ARL for detecting persistent process mean shifts. In this section, the performance of grouped data and variables-based EWMA charts and CUSUM charts for detecting process mean shifts are compared. The in-control process is assumed to be normally distributed and, without loss of generality, the in-control process mean and variance are set to 0 and 1 respectively. As EWMA charts are inherently two sided, they are compared with a two-sided tabular CUSUM. A tabular CUSUM procedure to detect increases in the process mean consists of plotting $Y_t = \max(0, Y_{t-1} + x_t - k)$, where $Y_0 = 0$ and k is a design parameter that specifies an indifference region (Montgomery (1991), p. 291). The CUSUM chart concludes that the process mean has shifted upwards whenever $Y_t \geqslant h$, where h is another design parameter. A two-sided tabular CUSUM is created by simultaneously monitoring two one-sided CUSUM charts, where the aim of one is to detect upward mean shifts, whereas the aim of the other is to detect downward shifts (Montgomery (1991), p. 291). For both the EWMA and the CUSUM charts based on grouped data we used the midpoint weights as discussed in Section 2, though similar qualitative results have been derived for other weighting schemes.

The ARL results are given in Table 2. The variables-based CUSUM chart with $h = 5$ and $k = 0.5$ has an in-control ARL of 430, and an out-of-control ARL at a one σ-unit shift in the mean of 10.2. These ARL values are used as the standard. The variables-based EWMA chart is designed to match these standard ARL values. The grouped data charts are designed so that their in-control run lengths match the target 430. For the grouped data CUSUM charts, ARL values are determined using the methodology presented in Steiner *et al.* (1996b). The run length results are matched by altering the value of h. However, because of the inherent discreteness of grouped data, the desired in-control ARL of 430 is not precisely obtainable. To make the ARLs easier to compare, the grouped data CUSUM ARLs, presented in Table 2, are theoretical values estimated using linear interpolation between the two closest cases. For the EWMA grouped data charts the value of L was altered to yield the desired in-control run length. For the EWMA grouped data chart more flexibility is available and the desired in-control run length was obtained without using interpolation. This design advantage of grouped data EWMA charts is discussed in more detail later.

Table 2 shows that for grouped data as well as variables data there is very little difference between the EWMA and CUSUM charts in terms of run length performance. The CUSUM chart seems to be slightly better at detecting process shifts that are smaller than the shift that the chart was designed to detect, whereas EWMA charts appear slightly better for larger process shifts. However, this pattern is reversed for smaller values of λ.

Although there is little difference between grouped data EWMA charts and grouped data CUSUM charts in terms of the ARL, there are other reasons why an EWMA chart may be preferable. First, the grouped data EWMA charts considered here are two sided by design,

Table 2. ARL comparison between two-sided grouped data EWMA charts and two-sided grouped data CUSUM charts

$\sqrt{n}\mu_x/\sigma_x$	*ARLs for the following conditions:*							
	Continuous		t = [−2, −1, 0, 1, 2]		t = [−1, 0, 1]		t = [−1, 1]	
	CUSUM, $k = 0.5$, $h = 5.0$	*EWMA*, $\lambda = 0.2045$, $L = 2.915$	*CUSUM*, $k = 0.5$, $h \cong 4.48$	*EWMA*, $\lambda = 0.2045$, $L = 2.897$	*CUSUM*, $k = 0.25$, $h \cong 4.074$	*EWMA*, $\lambda = 0.2045$, $L = 2.8$	*CUSUM*, $k = 0.5$, $h \cong 3.691$	*EWMA*, $\lambda = 0.2045$, $L = 2.78$
0.0	430	430	430	430	430	430	430	430
±0.5	37	39	42	42	42	44	48	52
±1.0	10.2	10.2	11.1	11.0	11.7	12.0	13.2	13.5
±1.5	5.7	5.4	6.0	5.9	6.8	6.7	7.5	7.1
±2.0	4.0	3.7	4.2	4.1	5.3	5.0	5.7	5.2
±3.0	2.5	2.3	3.1	3.1	4.5	4.1	4.7	4.1
±4.0	2.0	1.8	3.0	3.0	4.4	4.0	4.5	4.0

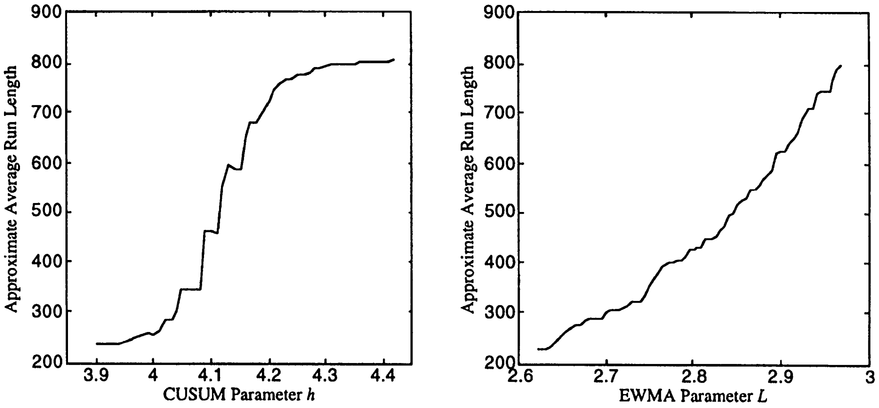

Fig. 5. Comparison of the discreteness in the in-control ARL of grouped data CUSUM and EWMA charts with $\mathbf{t} = [-1, 0, 1]$

whereas a two-sided CUSUM chart requires either the use of the awkward V-mask or two one-sided tabular CUSUM charts (Montgomery, 1991). As a result, if two-sided monitoring of the process is required, variables-based or grouped data EWMA charts are easier to implement. Second, for grouped data, over time the EWMA test statistic smooths out the inherent discreteness in the average group weight, whereas a grouped data CUSUM test statistic remains a simple linear combination of the initial group weights. As a result, grouped data EWMA charts are more flexible in their design than grouped data CUSUM charts, especially when the sample size is small. Fig. 5 illustrates this point very effectively. Fig. 5 shows the discreteness in the resulting in-control ARL for grouped data CUSUM and EWMA charts when the design parameters h and L are changed for fixed parameters $k = 0.5$ and $\lambda = 0.2$. As the parameter values h and L increase, the ARL of the chart should also increase. Fig. 5 shows that there are many more possible ARL values for an EWMA grouped data chart. This is a clear advantage when designing grouped data EWMA charts since typically sequential control charts are designed to have certain ARL characteristics. In Fig. 5, for the grouped data CUSUM, the slight decreases observed in the ARL of the grouped data CUSUM as h increases represent some errors in the approximation of the ARL.

5. Metal fasteners example

This section illustrates the application of a grouped data EWMA chart to the metal fasteners example. To aid comparisons with the previously proposed Shewhart chart (Steiner *et al.*, 1996a), the EWMA sample size was fixed at 12, and since the expected process shift is relatively small a λ-value of 0.1 was used. Setting the L-value so that the in-control ARL of the EWMA chart matches the Shewhart chart we derive an EWMA chart with $\lambda = 0.1$, $L = 2.54$ and $n = 12$. This grouped data EWMA chart has an in-control ARL of 370 and out-of-control ARLs of around 7.8 and 5.6 for positive and negative mean shifts of half a standard deviation unit respectively. These values are significantly better than the corresponding Shewhart chart with the same sample size. Fig. 6 shows the resulting EWMA chart using the data of Steiner *et al.* (1996a). The process was in control for the first 10 samples and was shifted down approximately one standard deviation unit starting at observation 11. The EWMA chart shown in Fig. 6 signals at observation 12.

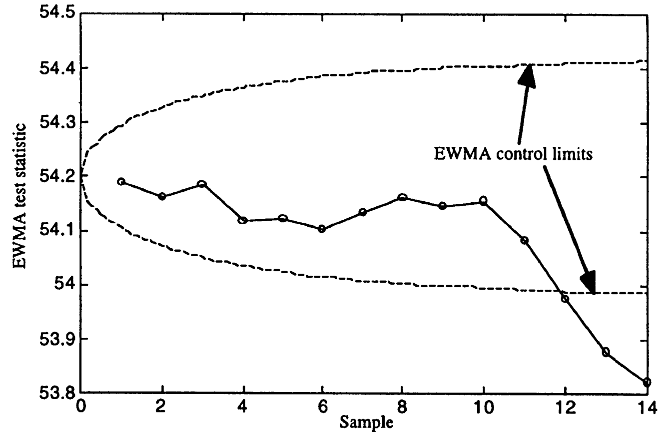

Fig. 6. EWMA control chart

In this application of the methodology, the resulting EWMA control charts detected the process shift in the same time that the corresponding group data Shewhart control chart detected the change. However, the observed process shift was fairly large here, of the order of one standard deviation unit. For smaller process shifts the derived grouped data EWMA chart would perform substantially better than the grouped data Shewhart chart.

6. Optimal grouping criterion

The run length results and comparisons presented in this paper assume fixed group intervals. This is often a reasonable assumption because of the use of standardized gauges. However, in some circumstances the placement of the group limits may be under our control. In that case, the question of optimal group intervals for the grouped data EWMA arises. Finding the best gauge limits requires a definition of optimal. One possibility is to find the gauge limits that yield the shortest out-of-control ARL at a given mean shift while having an in-control ARL of at least ARL_0. This is an attractive definition of optimal gauge limit but requires a solution for different in-control ARLs and different out-of-control shifts. Another approach to finding good gauge limits is to determine the grouping criterion that gives the best estimate of the process mean while the process is in control. This is an attractive option since usually the process will remain in control most of the time, and there is a connection between good estimation and effective hypothesis testing. The best gauge limits for estimation are found by maximizing the expected Fisher information that is available in the grouped data. The expected Fisher information provides a measure of the efficiency of the grouped data compared with variables data. Steiner *et al.* (1996a) derived the best estimation gauge limits for grouped data to detect shifts in the normal mean or standard deviation, and Steiner (1994) derived the optimal limits to detect a shift in Weibull parameters.

Acknowledgement

This research was supported, in part, by the Natural Sciences and Engineering Research Council of Canada.

Appendix A

This appendix derives the expected value and variance of the run length of grouped data EWMA control charts. The solution procedure utilizes a Markov chain where the state space between the control limits is divided into $g-1$ distinct discrete states and the out-of-control condition corresponds to the gth state. The states are defined as

$$\sigma = (s_1, s_2, \ldots, s_{g-2}, s_{g-1}) = (\text{LCL} + y, \text{LCL} + 2y, \ldots, \text{UCL} - 2y, \text{UCL} - y),$$

where $y = (\text{UCL} - \text{LCL})/g$ and UCL and LCL are the control limits as given by equation (3).

The transition probability matrix is given by

$$P = \begin{pmatrix} p_{11}, & p_{12}, \ldots, & p_{1g} \\ p_{21}, & \ldots, & p_{2g} \\ \vdots & & \vdots \\ p_{g1}, & \ldots, & p_{gg} \end{pmatrix} = \begin{pmatrix} R, & (I-R)\mathbf{1} \\ 0, \ldots, 0, & 1 \end{pmatrix}, \tag{8}$$

where I is the $(g-1) \times (g-1)$ identity matrix, $\mathbf{1}$ is a $(g-1) \times 1$ column vector of 1s and p_{ij} is the transition probability from state s_i to state s_j. The last row and column of the matrix P correspond to the absorbing state that represents an out-of-control signal. The R-matrix equals the transition probability matrix with the row and column that correspond to the absorbing (out-of-control) state deleted.

The group probabilities π_a, $a = 1, \ldots, k$, as defined by equation (4), and the group weights w_a, $a = 1, \ldots, k$, as given by equations (5) or the solution to equations (6), set the transition probabilities in the matrix R. Using the defined states **s** as a discretization, the transition probabilities p_{ij} are

$$p_{ij} = \begin{cases} \pi_a & \text{if } s_j - y/2 < \lambda w_a + (1-\lambda)s_i < s_j + y/2, \\ 0 & \text{otherwise,} \end{cases} \quad \text{for } j = 1, 2, \ldots, g-1 \tag{9}$$

$$p_{ig} = \begin{cases} \sum \pi_a & \text{for all } a \text{ such that } \lambda w_a + (1-\lambda)s_i \geqslant s_{g-1} + y/2 \text{ or } \lambda w_a + (1-\lambda)s_i \leqslant s_1 - y/2, \\ 0 & \text{if no such } a \text{ exists.} \end{cases}$$

The expected run length and the variance of the run length are found by using the matrix R. Letting γ denote the run length of the EWMA, we have

$$\Pr(\gamma \leqslant t) = (I - R^t)\mathbf{1},$$

and thus

$$\Pr(\gamma = t) = (R^{t-1} - R^t)\mathbf{1} \qquad \text{for } t \geqslant 1. \tag{10}$$

Therefore,

$$E(\gamma) = \sum_{t=1}^{\infty} t \Pr(\gamma = t) = \sum_{t=1}^{\infty} R^t \mathbf{1} = (I - R)^{-1}\mathbf{1}. \tag{11}$$

Similarly, the variance of the run length is

$$\text{var}(\gamma) = 2R(I - R)^{-2}\mathbf{1}, \tag{12}$$

where equations (11) and (12) are $(g-1) \times 1$ vectors that give the mean and variance of the run length from any starting value or state s_i. The mean and variance of the run length that correspond to starting at z_0 are easily found by finding the i such that $s_i - y/2 \leqslant z_0 \leqslant s_i + y/2$. If the control limits are symmetric about z_0 the corresponding state is $s_{g/2}$.

This Markov chain solution approach approximates the solution, with the accuracy of the approximation depending on the number of states (g) used. Larger values of g tend to lead to better approximations. However, unfortunately, because of discreteness, the ARL value does not smoothly approach the true value as g increases. As a result, the regression extrapolation technique suggested by

Brook and Evans (1972), to find the ARL of a variables-based CUSUM scheme approximated by a Markov chain, is not applicable here. However, fairly close approximations of the true run length properties can be obtained by taking the average result obtained using a few fairly large values of g. For example, the results presented in this paper estimate the true value $E(\gamma)_{g=\infty}$ and $\text{var}(\gamma)_{g=\infty}$ by averaging the results derived with $g = 100, 110, 120, 130, 140, 150$. Verification of this approach by using simulation suggests that the derived estimates for the mean and variance of the run length differ from the true value by less than 2–3% for most group limit designs of interest, with the approximation becoming worse as the size of the process shift increases.

References

Beja, A. and Ladany, S. P. (1974) Efficient sampling by artificial attributes. *Technometrics*, **16**, 601–611.

Box, G. E. P., Jenkins, G. M. and MacGregor, J. F. (1974) Some recent advances in forecasting and control: Part II. *Appl. Statist.*, **23**, 158–179.

Brook, D. and Evans, D. A. (1972) An approach to the probability distribution of cusum run length. *Biometrika*, **59**, 539–549.

Crowder, S. V. (1987) Run-length distributions of EWMA charts. *Technometrics*, **29**, 401–407.

Gan, F. F. (1990) Monitoring observations generated from a binomial distribution using modified exponentially weighted moving average control chart. *J. Statist. Comput. Simuln*, **37**, 45–60.

Geyer, P. L., Steiner, S. H. and Wesolowsky, G. O. (1996) Optimal SPRT and CUSUM procedures using compressed limit gauges. *IIE Trans.*, **28**, 489–496.

Haitovsky, Y. (1982) Grouped data. In *Encyclopedia of Statistical Sciences* (eds S. Kotz, N. L. Johnson and C. B. Read), vol. 3, pp. 527–536. New York: Wiley.

Ladany, S. P. (1976) Determination of optimal compressed limit gaging sampling plans. *J. Qual. Technol.*, **8**, 225–231.

Lucas, J. M. and Saccucci, M. S. (1990) Exponentially weighted moving average control schemes: properties and enhancements (with discussion). *Technometrics*, **32**, 1–29.

Montgomery, D. C. (1991) *Introduction to Statistical Quality Control*, 2nd edn. New York: Wiley.

Page, E. S. (1954) Continuous inspection schemes. *Biometrika*, **41**, 100–114.

Roberts, S. W. (1959) Control chart tests based on geometric moving averages. *Technometrics*, **1**, 239–250.

Schilling, E. G. (1981) *Acceptance Sampling in Quality Control*. New York: Dekker.

Schneider, H. and O'Cinneide, C. (1987) Design of CUSUM control charts using narrow limit gauges. *J. Qual. Technol.*, **19**, 63–68.

Steiner, S. H. (1994) Quality control and improvement based on grouped data. *PhD Thesis*. McMaster University, Hamilton.

Steiner, S. H., Geyer, P. L. and Wesolowsky, G. O. (1994) Control charts based on grouped data. *Int. J. Prodn Res.*, **32**, 75–91.

———(1996a) Shewhart control charts to detect mean and standard deviation shifts based on grouped data. *Qual. Reliab. Engng Int.*, **12**, 345–353.

———(1996b) Grouped data sequential probability ratio tests and cumulative sum control charts. *Technometrics*, **38**, 230–237.

Stevens, W. L. (1948) Control by gauging (with discussion). *J. R. Statist. Soc.* B, **10**, 54–108.