Quality Engineering
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/lqen20

A FAST INITIAL RESPONSE SCHEME FOR THE EXPONENTIALLY WEIGHTED MOVING AVERAGE CONTROL CHART

Teri Reed Rhoads [a] , Douglas C. Montgomery [a] & Christina M. Mastrangelo [b]

[a] Department of Industrial and Management Science Engineering , Arizona State University , Box 875906, Tempe, Arizona, 85287-5906

[b] Department of Systems Engineering , University of Virginia , Thornton Hall, Charlottesville, Virginia, 22903

Published online: 19 Oct 2007.

Quality Engineering, 9(2), 317-327 (1996-97)

A FAST INITIAL RESPONSE SCHEME FOR THE EXPONENTIALLY WEIGHTED MOVING AVERAGE CONTROL CHART

Teri Reed Rhoads and Douglas C. Montgomery
Department of Industrial and Management Science Engineering
Arizona State University
Box 875906
Tempe, Arizona 85287-5906

Christina M. Mastrangelo
Department of Systems Engineering
University of Virginia
Thornton Hall
Charlottesville, Virginia 22903

Key Words

FIR EWMA; EWMA; Fast initial response EWMA; Average run length.

Introduction

Exponentially weighted moving average (EWMA) and cumulative sum (CUSUM) control charts have found extensive application in industry. Both the EWMA and the CUSUM control chart present significant advantages over traditional Shewhart control charts when small process shifts are of interest. Specifically, a properly designed CUSUM or EWMA control chart will have a smaller value of the average run length (ARL) for detecting a small shift (that is, a shift of up to about 1.5 standard deviations) while maintaining at least the same ARL as the Shewhart chart when the process is in control.

For this reason, CUSUM and EWMA control charts have been widely used in the chemical and process industries, where relatively small process disturbances often have serious economic consequences. In these industries, a common situation following an out-of-control signal is to take a corrective action that resets the process back to the target (or in-control) value of the mean. Lucas and Crosier (1) proposed a fast initial response (FIR) or "head start" modification for the CUSUM to improve its shift detection performance following a process adjustment that did not return the mean to the target value.

This article discusses a fast initial response modification for the EWMA control chart. We investigate two versions of the EWMA with the FIR feature: one proposed by Lucas and Saccucci (2), and a new procedure. Although both procedures can enhance the performance of the EWMA, the new procedure that we propose is superior and is the one we recommend.

EWMA and CUSUM Control Charts

We now give a brief summary of the EWMA and CUSUM control charts. For further details of the construction and operation of these charts, see Ref. 3.

The EWMA is defined as

$$Z_t = \lambda X_t + (1 - \lambda)Z_{t-1}, \tag{1}$$

where $0 < \lambda < 1$ and X_t is either an individual process measurement made at time t or a sample mean if subgroups of size n are observed. The starting value for the EWMA at time $t = 0$ (or the start-up value), Z_0, is the target value of the mean, which we will take in this study to be zero. The variance of the EWMA statistic Z_t is

$$\sigma_z^2 = \frac{\sigma^2}{n}\left(\frac{\lambda}{2 - \lambda}\right)[1 - (1 - \lambda)^{2t}] \tag{2}$$

which, as t increases, becomes

$$\sigma_z^2 = \frac{\sigma^2}{n}\left(\frac{\lambda}{2 - \lambda}\right). \tag{3}$$

The EWMA control chart is constructed by plotting Z_t versus time, and setting the control limits at

$$Z_0 \pm L\sigma_z. \tag{4}$$

Several approaches have been used to compute ARL values for the EWMA. Robinson and Ho (4) use an iterative autoregressive numerical technique utilizing successive Edgeworth expansions for approximating ARL values. Crowder (5) computes exact expressions for ARLs by using an integral-equation approach to obtain expressions for the moments of run length and then solves the linear equations. Lucas and Saccucci (2) compute ARL values by approximating the EWMA statistic with a finite-state Markov chain. Crowder (5), Lucas and Saccucci (2), and Montgomery (3) give recommendations for the design of an EWMA control chart; that is, selecting λ and L.

There are several ways to define CUSUM schemes. We prefer the tabular or algorithmic CUSUM in which one computes, at time t, an upper-side CUSUM statistic $S_H(t)$ and a lower-side CUSUM statistic $S_L(t)$ according to

$$S_H(t) = \max[0, x_t - (\mu_0 + K) + S_H(t - 1)] \tag{5}$$

and

$$S_L(t) = \max[0, (\mu_0 - K) - x_t + S_L(t - 1)] \tag{6}$$

where μ_0 is the target value of the mean and K is a constant often called the reference value or the slack value. If either $S_H(t)$ or $S_L(t)$ exceed the decision interval H, the process is out of control (or off target). At time $t = 0$ (or after a process adjustment), the starting value of the CUSUM is $S_H(0) = S_L(0) = 0$. Montgomery (3) discusses the choice of the parameters K and H to give a good ARL performance.

Lucas and Crosier (1) proposed using a "head start" or fast initial response (FIR) feature for the CUSUM to improve its performance in detecting an off-target process at CUSUM start-up. Their suggestion is very simple: Set $S_H(0) = S_L(0) = H/2$ (called a 50% head start) every time an adjustment or corrective action is made that has as its objective bringing the process mean back to the target value μ_0. They show considerable reduction in the average run length to detect an off-target value of the mean results from using the FIR feature. The reason for this is clear: If the process is still off target following an adjustment, the CUSUM only has to "rise" one-half the distance from the head start in comparison to the full distance H if no head start is used. The FIR feature does not materially affect the in-control ARL, because if the process is correctly adjusted back to the target value μ_0, both CUSUM statistics $S_H(t)$ and $S_L(t)$ will be driven to zero, at which time the head start value is effectively eliminated.

Fast Initial Response and the EWMA

Lucas and Saccucci (2) suggested that a FIR feature would be useful for the EWMA control chart, particularly while using small values of λ. Because this constitutes the majority of EWMA applications in the authors' experience, an FIR would seem to be generally applicable to the EWMA. Lucas and Saccucci implement the FIR by using two one-sided EWMAs: one with a starting value above the target, and the other with a starting value below the target. They illustrated a head start value of 50%; that is, the head start is one-half of the distance between the target value (the centerline Z_0 on the EWMA control chart) and the control limit. Both EWMA schemes are run until either an out-of-control signal is generated or the two one-sided EWMAs are within one standard deviation of each other—at which time the two EWMAs would be averaged and the control chart continued in the usual manner from that point.

All of Lucas and Saccucci's analysis was done using the steady-state value of σ_z in computing the control limits. This, of course, results in a set of fixed-width limits [see Eqs. (2) and (3)]. Correct use of the EWMA requires the use of Eq. (2) in constructing control limits. This results in control limits that are parabolic in shape for the first few observations. These limits increase in width with time t,

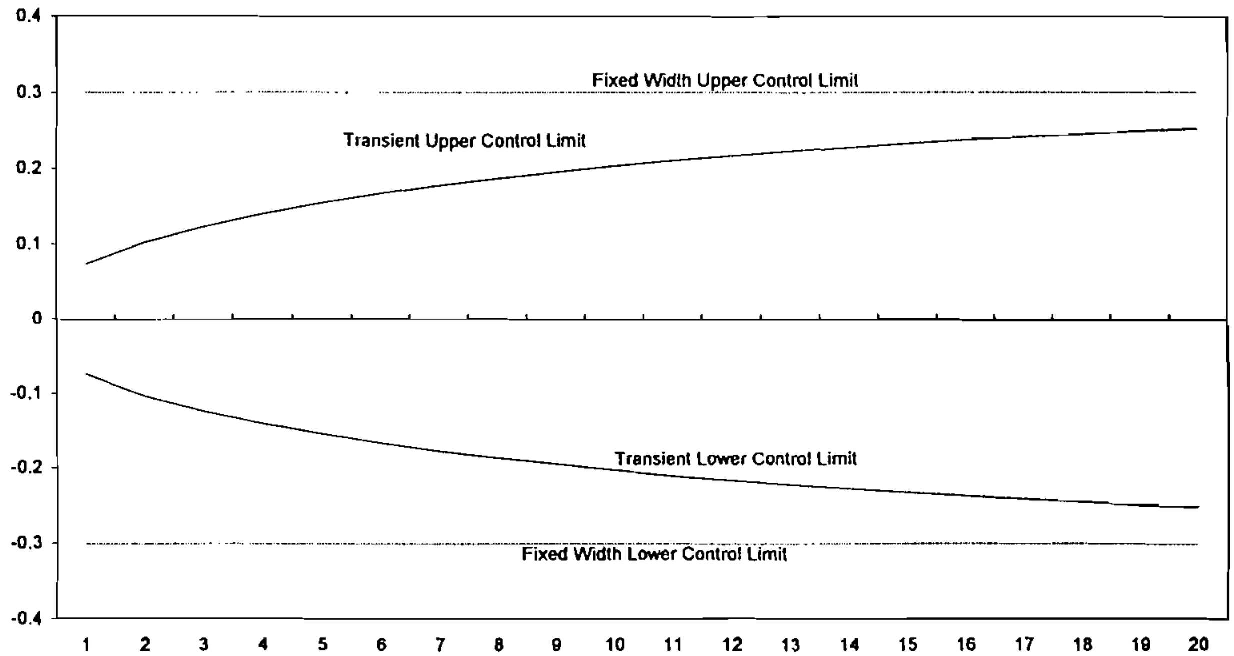

Figure 1. Transient versus fixed-width control limits, 20 periods, $\lambda = 0.03$.

converging to the steady-state or fixed-width limits that would result from Eq. (3). Generally, the speed of this convergence is a function of λ, with smaller values of λ requiring more time periods to converge. Figure 1 shows a comparison of the fixed-width limits versus the proper control limits with the parabolic section described above for $\lambda = 0.03$. Most standard computer programs compute the limits correctly (see, for example, Ref. 6).

We propose using the FIR feature with the EWMA in the manner suggested by Lucas and Saccucci but using the correct parabolic limits. We refer to this as an FIR EWMA with *transient* limits to indicate the time dependency in the behavior of the control limits. We will show that this version of the FIR procedure outperforms both the Lucas and Saccucci FIR scheme and the EWMA control chart without the FIR feature.

Average Run Length Comparisons

Before giving the ARL results, we first show some examples of how these different FIR EWMA schemes work. Lucas and Saccucci (2) present a set of 19 simulated observations with $\mu = 0$ and $\sigma = 1$. At the eleventh observation, a shift of 1σ occurs. The EWMA using transient limits and the FIR EWMA with transient limits and fixed limits operated identically for this case. Figure 2 shows the upper limits in this case. From this graph, it can be seen that all three methods converge eventually to the same values as do the transient and fixed limits. However, starting an analysis at observation 11, when the mean is already off target, produced different results. These comparisons are plotted in Figure 3. The FIR EWMA with transient limits and fixed limits both signal three periods before the regular EWMA with transient limits. Finally, a third example was constructed in which the value at the third time period was reduced by 0.1; then, only the FIR EWMA with transient limits signals early, as is shown in Figure 4. Therefore, clearly, the use of an FIR feature can improve EWMA control chart performance if the process is not correctly adjusted to the target value of the mean.

Computer simulation was used to study the ARL performance of the EWMA chart with and without transient (parabolic) limits but no FIR, and the two variations of the FIR scheme discussed in the previous section. Throughout, samples were drawn from a normal distribution with mean $\mu_0 = 0$ and unit standard deviation. The subgroup size was taken to be $n = 1$. Thus, our results can be interpreted as ARLs for a standardized EWMA control chart. ARLs were computed based on 1000 replicates of each test scenario.

Using this simulation procedure, a comparison was made among the four methods: the EWMA with fixed limits, the EWMA with transient limits, the FIR EWMA with

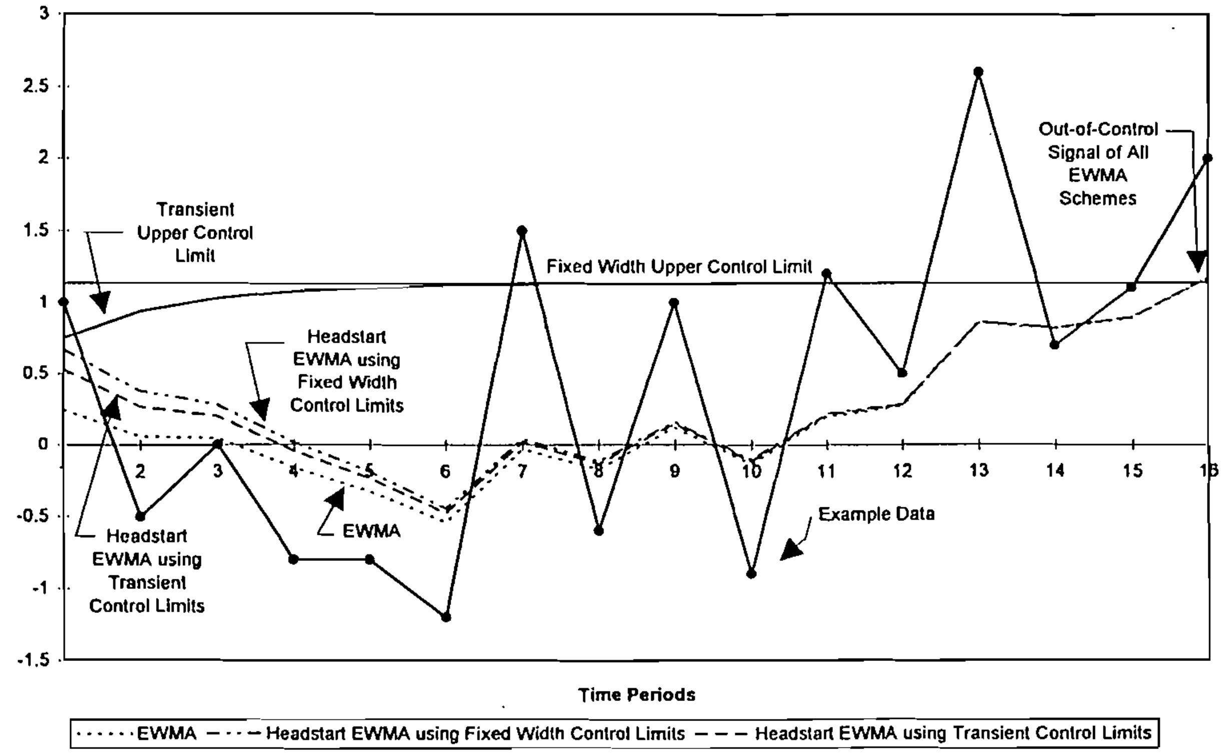

Figure 2. Example No. 1 from Ref. 2, adding a head start EWMA with transient control limits.

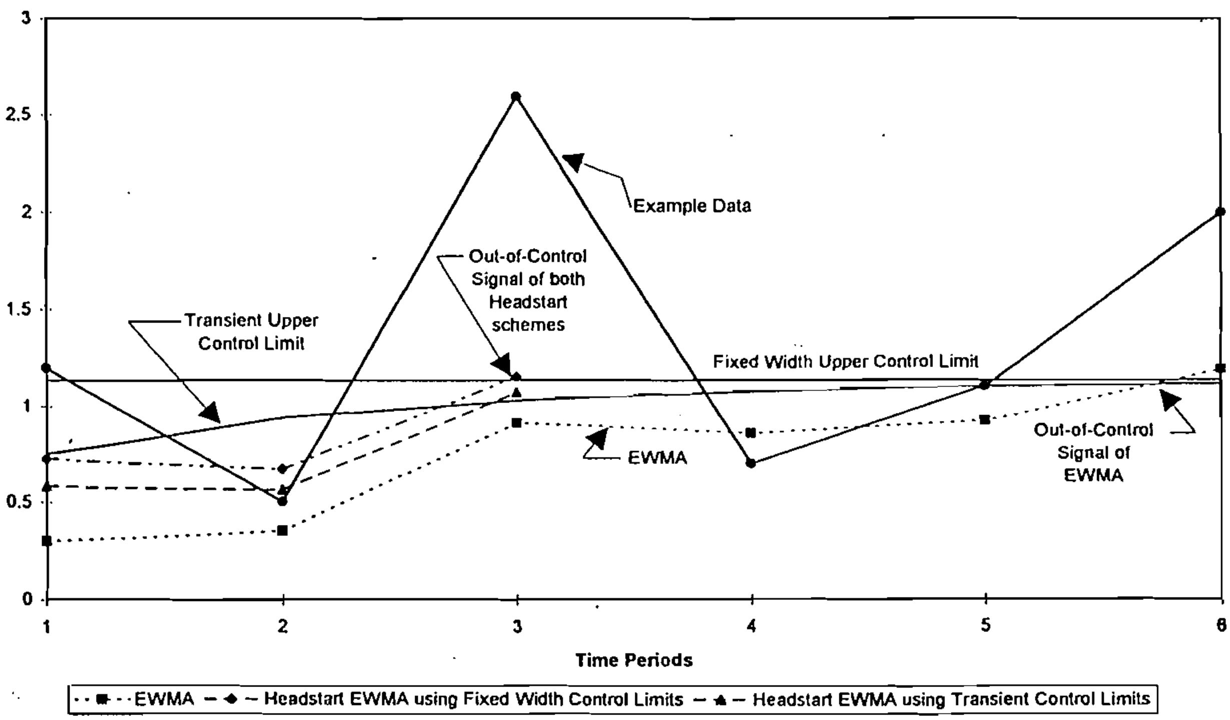

Figure 3. Example No. 2 from Ref. 2, adding a head start EWMA with transient control limits.

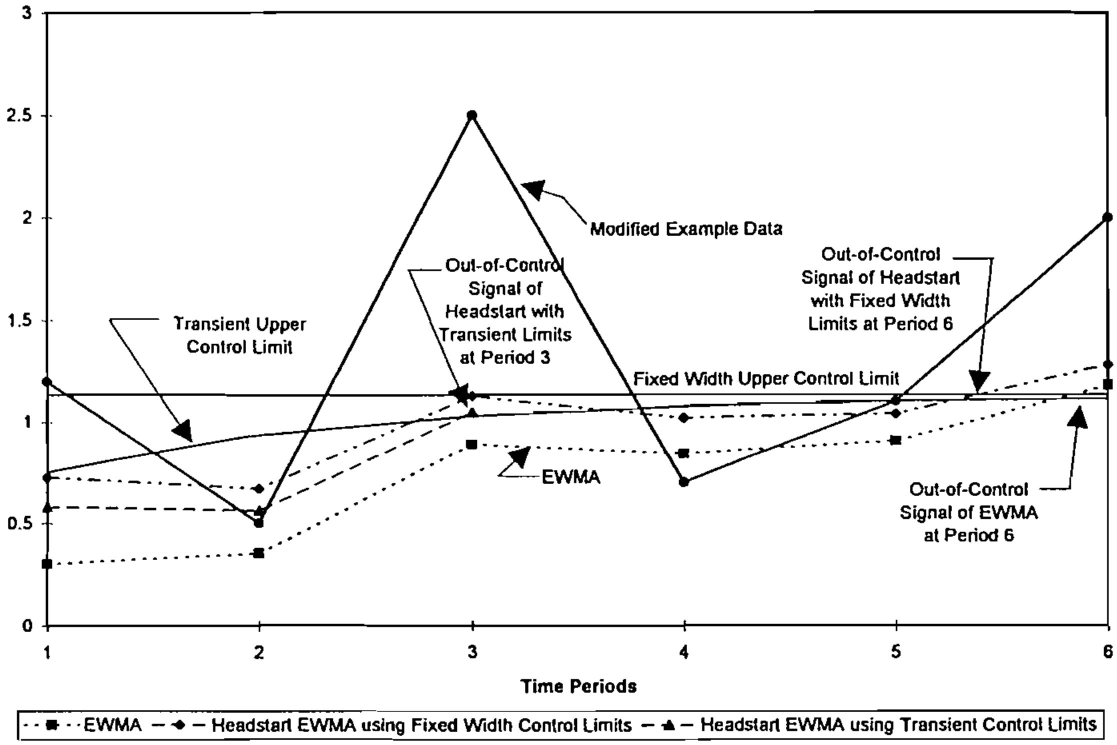

Figure 4. Example No. 2, from Ref. 2, modified.

fixed limits (as proposed in Ref. 2), and the FIR EWMA with transient limits. Table 1 is a summary of the ARLs for five sets of control limit widths (L), λ's, and shifts in the process mean of 0 (the mean is correctly adjusted back to the target), 0.5, 1, 1.5, 2, 3, and 4 as a multiple of $\sigma_{\bar{x}}$. Because the process is assumed to start with the mean at the shifted value indicated in the table, the ARL values are "initial state" ARLs, as described by Lucas and Saccucci. The values of L and λ used were selected from those given in Table 3 of Ref. 2. This choice of parameters gives an ARL of about 500 periods when the process is in control and no FIR feature is used.

Several observations may be made by examining Table 1. First, using any FIR scheme in conjunction with the EWMA results in an improved capability to detect an off-target process, in the sense that the ARL is reduced. For example, consider the case of a $1\sigma_{\bar{x}}$ shift of deviation in the mean from target. If $L = 2.615$ and $\lambda = 0.05$, then the ARLs for the fixed-limit EWMA and the transient-limit EWMA are 11.3 and 7.2 samples, respectively. Incorporation of the head start reduces these ARLs to 7.0 and 4.2, respectively. Second, the FIR EWMA with transient limits always has a shorter ARL than does the FIR EWMA with fixed limits. In many cases, this reduction in the ARL is very substantial, indicating that the FIR with transient limits is the preferred way to implement the "head start" feature for an EWMA scheme. Finally, the use of the FIR feature does not have a dramatic effect on the in-control ARL (shift = 0 in Table 1), except in the case where λ is relatively small, say $\lambda \leq 0.1$. In these cases, the head start used with the transient limit does result in a substantial reduction of the in-control ARL.

For the cases where $\lambda \leq 0.1$, we reran the simulations, increasing the values of L for the FIR EWMA with transient limits until the in-control ARL obtained matched approximately the in-control ARL for the FIR EWMA with fixed limits. We also ensured that the in-control ARL values were at least 370 samples so that the "false alarm rate" for the procedure would be at least equivalent to that for a Shewhart control chart. The ARLs for those cases are shown in Table 2. Note that by comparing the results with those in Table 1 for shift sizes $\geq 1\sigma_{\bar{x}}$, there are no substantial changes in the reported ARLs or standard deviations of run length. Therefore, by widening the control lengths in the cases with $\lambda \leq 0.1$, a shorter ARL than would be obtained with the fast initial response scheme with fixed-width

Table 1. Average Run Length and Standard Deviation of Run Length for EWMA Control Schemes

	$L = 3.054$, $\lambda = 0.40$ AVERAGE RUN LENGTH STANDARD DEVIATION				$L = 2.998$, $\lambda = 0.25$ AVERAGE RUN LENGTH STANDARD DEVIATION				$L = 2.814$, $\lambda = 0.1$ AVERAGE RUN LENGTH STANDARD DEVIATION			
SHIFT[a]	EWMA FIXED	EWMA TRANS	FIR FIXED	FIR TRANS	EWMA FIXED	EWMA TRANS	FIR FIXED	FIR TRANS	EWMA FIXED	EWMA TRANS	FIR FIXED	FIR TRANS
0	497.2	496.0	481.7	467.9	498.6	494.9	483.7	452.0	496.2	487.0	462.6	418.8
	493.3	494.1	487.6	484.7	505.7	506.3	507.4	490.7	518.3	520.5	520.3	526.0
0.5	71.7	68.1	65.9	63.3	48.3	46.9	42.1	39.3	32.0	28.8	24.2	20.7
	67.8	64.3	64.3	64.3	43.3	43.9	42.4	42.8	22.7	22.5	21.9	22.1
1	14.4	14.0	12.5	11.8	11.1	10.4	8.5	7.6	10.3	8.1	6.9	5.2
	11.9	12.0	11.8	11.9	7.4	7.5	7.1	7.3	4.6	4.9	4.3	4.9
1.5	5.9	5.5	4.7	4.1	5.5	4.8	3.9	3.2	6.1	4.1	3.7	2.4
	3.6	3.7	3.6	3.5	2.7	2.8	2.5	2.7	2.1	2.3	1.6	1.8
2	3.5	3.1	2.6	2.3	3.6	3.0	2.5	1.9	4.3	2.6	2.7	1.6
	1.7	1.7	1.6	1.6	1.3	1.5	1.3	1.3	1.2	1.4	1.0	1.0
3	2.0	1.7	1.5	1.3	2.2	1.6	1.5	1.1	2.9	1.5	1.8	1.1
	0.7	0.7	0.6	0.6	0.6	0.7	0.6	0.4	0.7	0.6	0.6	0.3
4	1.4	1.2	1.1	1.0	1.7	1.2	1.1	1.0	2.2	1.1	1.3	1.0
	0.5	0.4	0.3	0.2	0.5	0.4	0.3	0.2	0.4	0.3	0.5	0.1

	$L = 2.615$, $\lambda = 0.05$ AVERAGE RUN LENGTH STANDARD DEVIATION				$L = 2.437$, $\lambda = 0.03$ AVERAGE RUN LENGTH STANDARD DEVIATION			
SHIFT[a]	EWMA FIXED	EWMA TRANS	FIR FIXED	FIR TRANS	EWMA FIXED	EWMA TRANS	FIR FIXED	FIR TRANS
0	481.6	458.7	420.6	318.1	477.3	427.9	383.4	286.2
	462.0	464.9	460.3	427.5	430.5	432.0	411.5	394.0
0.5	28.8	23.0	19.7	15.7	29.7	20.0	18.6	12.8
	16.0	16.9	14.8	16.6	14.3	15.3	13.0	14.7
1	11.3	7.2	7.0	4.2	12.5	6.5	7.4	3.6
	4.1	4.7	3.6	4.4	4.1	4.6	3.4	3.8
1.5	7.1	3.6	4.1	2.1	8.1	3.3	4.6	1.9
	2.0	2.1	1.5	1.7	2.1	2.1	1.7	1.6
2	5.2	2.4	3.1	1.4	6.0	2.2	3.4	1.4
	1.2	1.3	1.0	0.9	1.3	1.2	1.0	0.8
3	3.5	1.4	2.1	1.1	4.0	1.3	2.4	1.0
	0.7	0.6	0.5	0.3	0.7	0.6	0.6	0.2
4	2.7	1.1	1.7	1.0	3.1	1.1	1.9	1.0
	0.5	0.2	0.5	0.05	0.5	0.2	0.4	0.03

[a]The shift or deviation of the mean from target as a multiple of $\sigma_{\bar{x}}$.

Table 2. Average Run Length Performance of the FIR EWMA with Adjusted Control Limit Width

	AVERAGE RUN LENGTH STANDARD DEVIATION					
	$\lambda = 0.1$		$\lambda = 0.05$		$\lambda = 0.03$	
SHIFT[a]	ORIGINAL $L = 2.814$	ADJUSTED $L = 3.0$	ORIGINAL $L = 2.615$	ADJUSTED $L = 2.72$	ORIGINAL $L = 2.437$	ADJUSTED $L = 2.535$
0	462.6	465.6	420.6	416.8	383.4	384.2
	520.3	551.4	460.3	534.6	411.5	562.9
0.5	24.2	22.2	19.7	17.0	18.6	14.9
	21.9	23.3	14.8	17.7	13.0	16.2
1	6.9	5.4	7.0	4.4	7.4	3.9
	4.3	5.1	3.6	4.3	3.4	3.8
1.5	3.7	2.4	4.1	2.2	4.6	2.0
	1.6	1.9	1.5	1.8	1.7	1.7
2	2.7	1.6	3.1	1.5	3.4	1.4
	1.0	1.0	1.0	0.9	1.0	0.9
3	1.8	1.1	2.1	1.1	2.4	1.1
	0.6	0.3	0.5	0.2	0.6	0.2
4	1.3	1.0	1.7	1.0	1.9	1.0
	0.5	0.09	0.5	0.04	0.4	0.03

[a]The shift or deviation of the mean from target as a multiple of $\sigma_{\bar{x}}$.

control limits is still the result. There is a larger change in the average run lengths and the standard deviation of run length in the scenarios with shifts $< 1\sigma_{\bar{x}}$; these changes are increases in both reported statistics. However, none of the changes observed would alter our conclusions about the effectiveness of the proposed fast initial response scheme for the EWMA.

Example

In this section, we show a practical application of the proposed fast initial response scheme with appropriate transient limits for the EWMA control chart. Many of the applications for this technique will be process industries, where the output variable is monitored by a control chart and adjustments are made to a manipulated variable when the output is "out of control" on a standard control chart. We will focus on that situation. The process chosen for our example is continuous polymerization of a product used as an additive in motor oil. The quality characteristic of interest in this product is number-average molecular weight. This characteristic should be controlled to a target value of 2500 moles if the product is to be acceptable to the end user. Molecular weight is controlled by making adjustments to catalyst feed rate. Historically, the molecular weight has been monitored using a Shewhart control chart and adjustments to the feed rate are only made when a point goes outside the control limits on this chart. We will demonstrate how this system would operate with the EWMA control chart using the head start and the transient limits feature we propose.

Figure 5 is an EWMA control chart for 50 observations on molecular weight. The EWMA parameter λ is equal to 0.1, the control limits are at $\pm 3\sigma$, and as is evident from the figure, the process is in control at the target value of 2500 moles. The standard deviation of molecular weight used to construct this control chart is 42.49214, based on the first 50 observations. Note that the chart has transient limits and that the head start scheme we have proposed is used. Because the process is operating on target, the fast initial response feature quickly converges to the standard EWMA statistic. This will happen in any situation where the process begins and stays on target for a period of time.

At period 50, an out-of-control signal is generated on Figure 5. This means that molecular weight has shifted off target by a statistically significant amount and that an adjustment to catalyst feed rate is required. The amount of the adjustment would be calculated by the process engineer to theoretically bring molecular weight back to the target value of 2500 moles. When this adjustment is made, how-

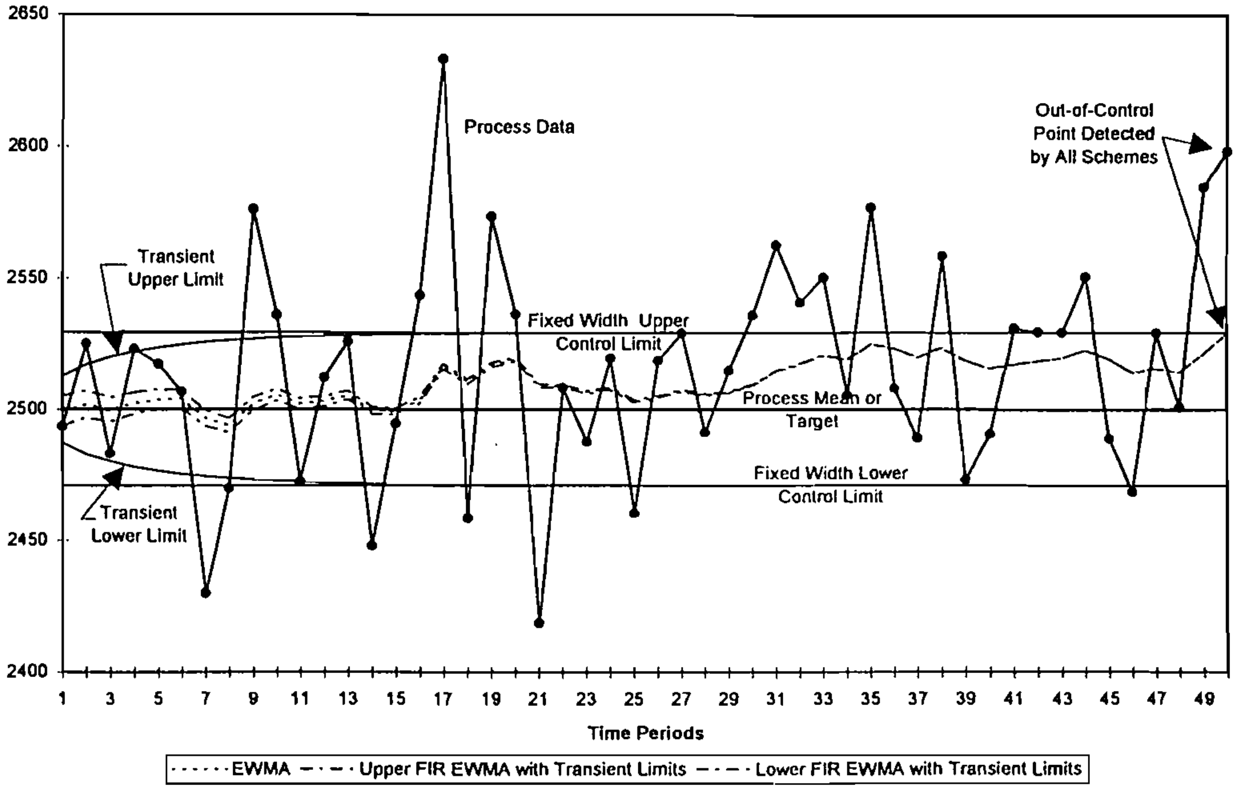

Figure 5. Polymer process data.

ever, following period 50, the actual adjustment applied apparently does not bring the process back on target. To see this, examine Figure 6. In this figure, we show the EWMA scheme with the head start applied to the observation on molecular weight in period 51. Note that the proposed scheme signals immediately, indicating that the process adjustment made following period 50 was not successful; that is, it did not result in the molecular-weight response being returned to the target value of 2500 moles or a value sufficiently close to 2500 moles for the process to be considered in control. Therefore, further process adjustments to catalyst feed rate are required.

It is interesting to observe what would have happened if an EWMA with fixed-width control limits and a head start feature had been employed instead of our scheme. Figure 7 contains this display. Note that when the process is restarted in period 51, after the process adjustment, the two EWMA schemes required by the head start feature move toward the upper control limit yet do not reach the limit until period 66. Thus, the time required to detect the off-target condition with this procedure is 16 periods, substantially greater than that required by the procedure that we recommend. Figure 8 presents the case where an ordinary EWMA with transient limits but no head start feature is applied following period 50. Note that this scheme does not signal the off-target condition until period 66. This performance is identical to what was obtained with the head start feature using the fixed-width control limits.

Summary

This article discussed and illustrated the use of head start or FIR schemes with the EWMA control chart. This feature is a useful one when the EWMA is used in a situation where the process is reset to target following a signal on the control chart. We have shown that the FIR EWMA with the appropriate transient control limits outperforms both the "standard" EWMA (without a head start) and the FIR EWMA with fixed-width limits suggested by Lucas and Saccucci (2). The performance improvement is especially significant with small values of λ, such as those that are often used in practice. It is our recommendation that the transient control limits be utilized in the application of the FIR technique in order to achieve the best performance

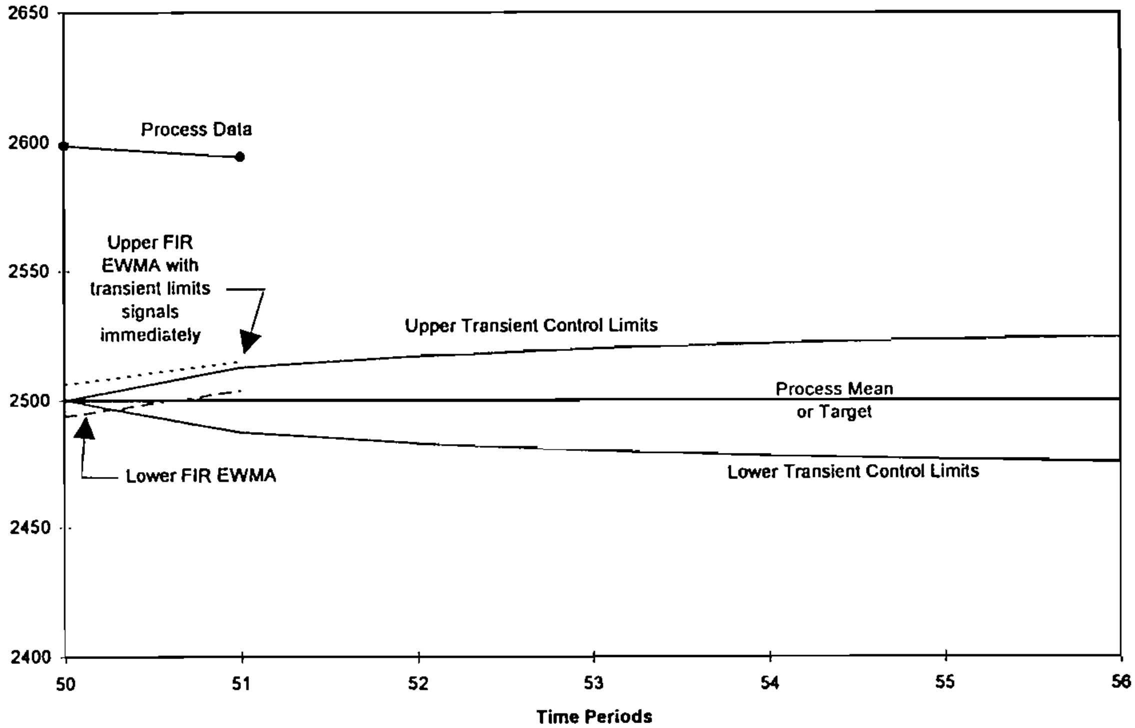

Figure 6. Polymer process data after an incorrect adjustment is made using transient control limits.

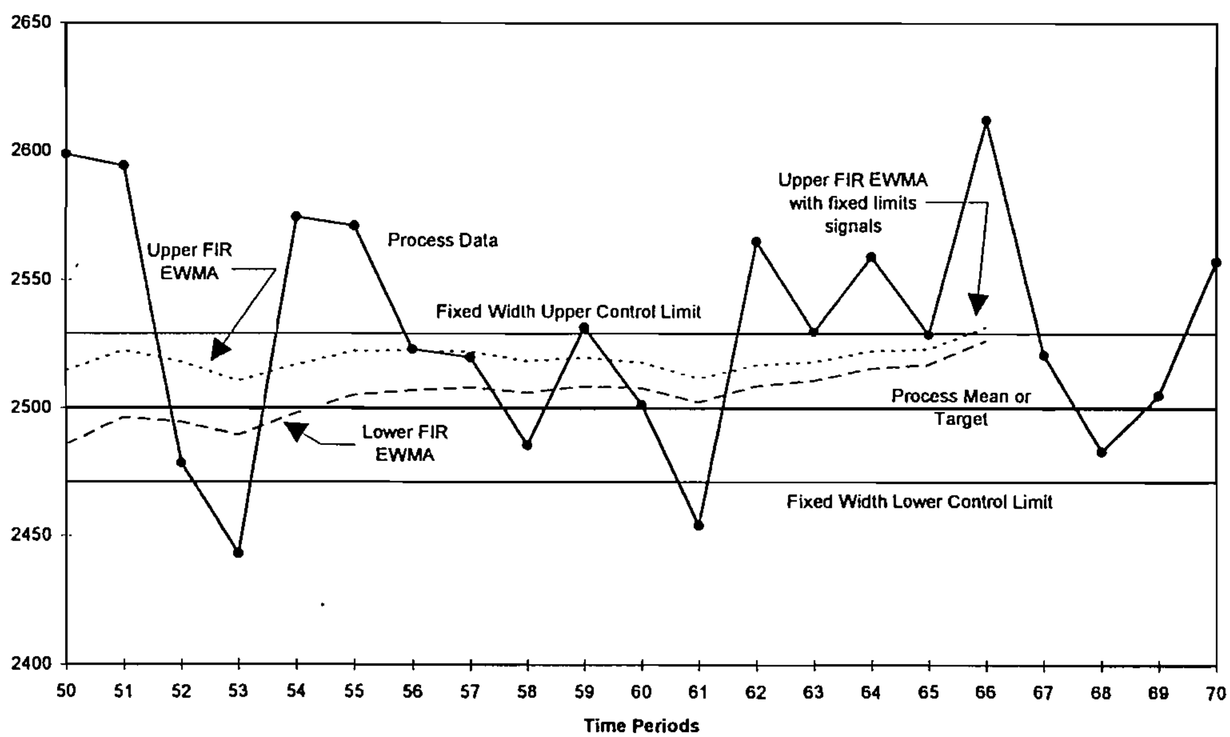

Figure 7. Polymer process data after an incorrect adjustment is made using fixed-width control limits.

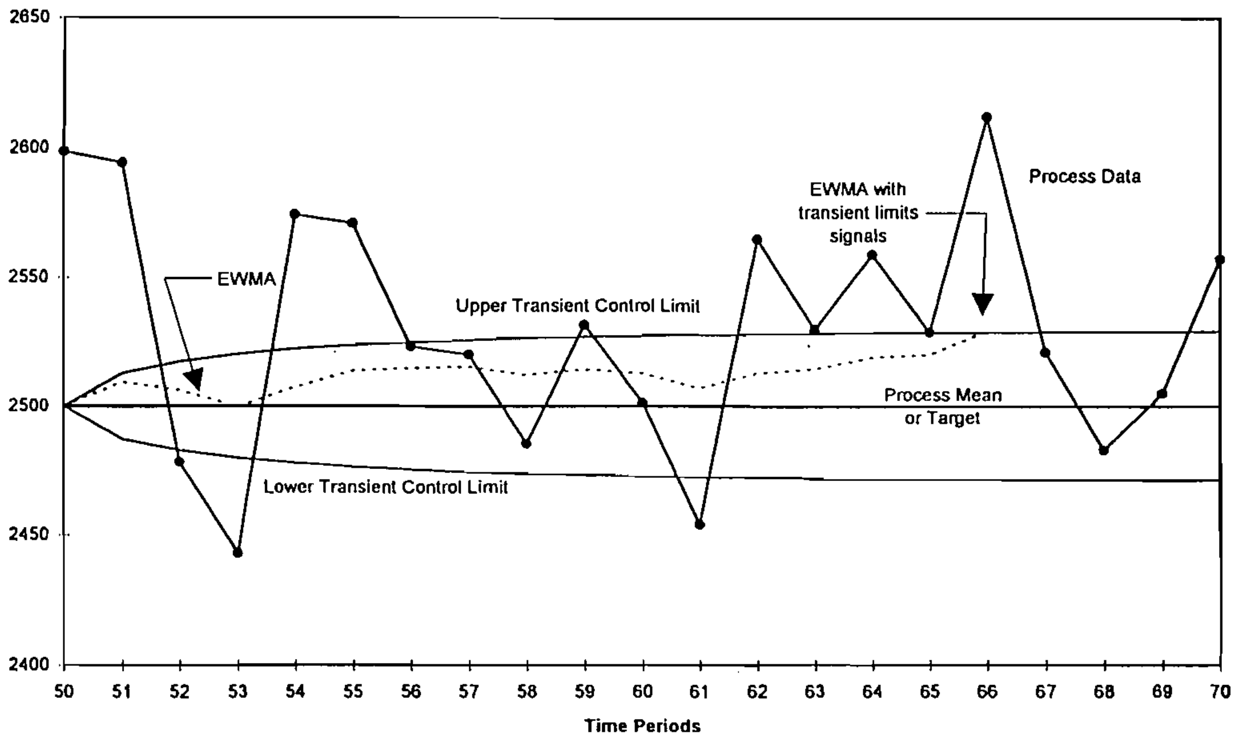

Figure 8. Polymer process data after an incorrect adjustment is made using no head start feature.

of the EWMA control scheme. Its greatest potential application is likely to be in situations where the process output is controlled to a target value by making periodic adjustments to some manipulated process variable.

Acknowledgment

The authors wish to thank the reviewer for his helpful comments that greatly improved an earlier version of this article.

References

1. Lucas, J. M. and Crosier, R. B., Fast Initial Response for CUSUM Quality-Control Schemes, *Technometrics*, 24, 199-206 (1982).
2. Lucas, J. M. and Saccucci, M. S., Exponentially Weighted Moving Average Control Schemes: Properties and Enhancements, *Technometrics*, 32, 1-12 (1990).
3. Montgomery, D. C., *Introduction to Statistical Quality Control*, 2nd ed., John Wiley & Sons, New York, 1991.
4. Robinson, P. B. and Ho, T. Y., Average Run Lengths of Geometric Moving Average Charts by Numerical Methods, *Technometrics*, 20, 85-93 (1978).
5. Crowder, S. V., Computation of ARL for Combined Individual Measurement and Moving Range Charts, *J. Qual. Technol.*, 19, 98-102 (1987).
6. *Statgraphics User's Manual*, Version 6, Manugistics, Inc., Cambridge, MA, 1992.

About the Authors: Teri Reed Rhoads is a doctoral student of quality and reliability engineering in the Department of Industrial and Management Systems Engineering at Arizona State University. She holds a bachelor degree in petroleum engineering from the University of Oklahoma and a masters in business administration from the University of Texas of the Permian Basin. She is a member of the Institute of Industrial Engineers and the Society of Petroleum Engineers.

Douglas C. Montgomery is professor of industrial and management systems engineering at Arizona State University. He is the author of 10 books, including widely used texts on statistical quality control and experimental design, and over 100 papers and articles. He is a Fellow of the American Society for Quality Control, a Fellow of the American Statistical Association, and a Fellow of the Institute of Industrial Engineers, and has held several national

offices in these organizations, including Director of the Statsitics Division of ASQC. He is the editor of the *Journal of Quality Technology*.

Christina M. Mastrangelo is an assistant professor of Systems Engineering at the University of Virginia. She holds degrees in industrial engineering from Arizona State University and her interests include quality engineering and experimental design. She is a member of the Institute of Industrial Engineers and the American Society for Quality Control.