UNIVERSITY OF GÖTEBORG

Department of Statistics

RESEARCH REPORT 1993:1
ISSN 0349-8034
Revised nov 93

COMPARISON BETWEEN TWO METHODS OF SURVEILLANCE: EXPONENTIALLY WEIGHTED MOVING AVERAGE VS CUSUM

by

Marianne Frisén

and

Göran Åkermo

Statistiska institutionen

Göteborgs Universitet

Viktoriagatan 13

S-411 25 Göteborg

Sweden

COMPARISON BETWEEN TWO METHODS OF SURVEILLANCE: EXPONENTIALLY WEIGHTED MOVING AVERAGE vs CUSUM

M Frisén and G Åkermo
Department of Statistics, University of Göteborg
Viktoriagatan 13, S-411 25 Göteborg, Sweden.

When control charts are used in practice it is necessary to know the characteristics of the charts in order to know which action is appropriate at an alarm. The probability of a false alarm, the probability of successful detection and the predictive value are three measures (besides the usual ARL) used for comparing the performance of two methods often used in surveillance systems. One is the "Exponentially weighted moving average" method, EWMA, (with several variants) and the other one is the CUSUM method (V-mask). Illustrations are presented to explain the observed differences. It is demonstrated that a high probability of alarm in the beginning (although it gives good ARL properties) might cause difficulties since a low predicted value makes action redundant at early alarms.

KEY WORDS: Quality control; Control charts; EWMA; FIR; V-mask; Predicted value; Performance;

Methods for continual surveillance to detect some event of interest, usually presented in the form of control-charts, are used in many different areas, e.g. industrial quality control, detection of shifts in economic time series, medical intensive care and environmental control.

A wide variety of methods have been suggested, see e.g. Zacks (1983) and Wetherill and Brown (1990). Some methods (like the Shewhart test) only take the last observation into account. Others (simple sums or averages) give the same weight to all observations. For most applications it is relevant to use something in between. That is, all observations are taken into account but more weight is put on recent observations than on old ones. The CUSUM and the EWMA are such methods. They are much discussed and both are nowadays often recommended. Both these methods include the extremes mentioned above as special cases and the relative weight on recent observations and old ones can be continuously varied by varying their two parameters. A description of the methods is given in Section 2.

Several extensive comparisons of these methods have been done, see e.g. Ng and Case (1989), Lucas and Saccucci (1990) and Domangue and Patch (1991). These comparisons are made for cases where the out-of-control state is present when the surveillance starts. The study by Domangue and Patch includes the case where the out-of-control state is a linearly increasing change, but also this state is assumed to start at the same time as the surveillance starts. The comparisons have not demonstrated any great differences. This is not surprising since by the two parameters the methods can be designed to fulfil two conditions. The methods can thus be designed to have the same average run length, ARL, (see Section 3.2) for both the in-control and the out-of-control state. Nearly all comparisons have been based on the ARL. Here a study is made of the remaining differences when the methods have the same ARL.

The usual measures of a test's performance, namely the significance level and the power, would have to be generalized in any of many possible ways to take into account the dependence on the length of the period of surveillance and the time point t' where the change occurs. Here, a systematic demonstration is made to show how these variables (which vary between practical situations) influence different measures (for different methods). Other variables such as the rate of change (if the change takes place successively) will also influence the performance of a method of surveillance. However, the following discussion will be restricted to the influence of the first two mentioned variables which always influence the performance. Thus only sudden changes to another constant level will be studied in the examples and simulations.

This paper uses three measurements of performance suggested by Frisén (1992) for the comparison of the two methods in cases where, by the choice of design parameters, the first moment of the run-length distributions are set equal. The main interest is the influence of time and the different risks of false judgements involved when repeated decisions will be made about hypotheses which might successively change.

In Section 1 the situations considered are specified and some notations are introduced. In Section 2 the two methods are presented. In Section 3 measures to be used in the evaluations are introduced. In Section 4 the results are given and in Section 5 the results are discussed.

1. SPECIFICATIONS

Let $X = \{X_t: t = 1,2,...\}$ be the observation of interest. It may be an average or some other derived statistic. In the case of surveillance of the fetal heart rate, X is a recursive residual of a measure of variation. The random process which determines the state of the system is denoted $\mu = \{\mu_t: t = 1,2,...\}$.

In the examples below the case of shift in the mean of Gaussian random variables from an acceptable value μ^0 (zero) to an unacceptable value μ^1 (one) is considered. It is assumed that if a change in the process occurs, the level suddenly moves to another constant level, μ^1, and remains on this new level. That is $\mu_t = \mu^0$ for $t= 1,...,\tau-1$ and $\mu_t = \mu^1$ for $t= \tau, \tau+1,....$

Here μ^0 and μ^1 are regarded as known values and the time point τ where the critical event occurs is regarded as a random variable with known density. The incidence of a change, inc(t'), is the probability that the stochastic time τ for the change takes the value t', conditioned on $\tau > t'-1$. The incidence is assumed constant in the following examples.

Our problem is to discriminate between the states of the system at each decision time s, s = 1,2,... by the observation $X(s) = \{X_t: t \leq s\}$ under the assumption that $X_1 - \mu_1$, $X_2 - \mu_2$,... are independent Gaussian random variables with mean zero and with the same known standard deviation(in the examples $\sigma=1$).

2. METHODS

Since repeated decisions are made and hypotheses might change over time theory of ordinary hypothesis testing does not apply. Two specific methods of surveillance often used in quality control will be described below. For more exhaustive descriptions of methods used in quality control see e.g. Wetherill and Brown (1990). The two methods will be evaluated by the measures suggested in Section 3. Thus their principal differences will be enlightened. However, the two methods are by no means the only ones to be considered. Similar comparisons of other methods was made by Frisén (1992). The present methods are chosen because they are much discussed methods of similar type. The EWMA- and the CUSUM-method both take past observations into account by summation. They also have two parameters each. They can thus have the same ARL both with and without a specific shift. To make the methods comparable the parameters of the methods are set by the requirement that the ARL^0 and ARL^1 (as described in Section 3) are the same. The actual values used in this study is for the in-control-state $ARL^0=330$ and for the out-of-control state of a shift to $\mu^1=1$ at the start of the surveillance, $ARL^1=9.6$. Very extensive simulations were used to find parameter sets which resulted in the same values of ARL and for the figures. Thus only one set of parameters is used. However, this is enough to prove that important differences might exist in spite of equal ARL values. The results will also support the general discussion about which qualities we should require.

Two-sided methods are used in the examples and simulations. The methods (with the same parameters as in the simulations) are illustrated in Figures 2 - 5 with data (Figure 1) used by Lucas and Crosier (1982) and Lucas and Saccucci (1990). In order to get simulation results which are suitable for comparisons between methods the same random numbers are used for all methods in each control sequence. The value for the first time point (and in some figures also the second one) is achieved by exact calculation.

Although discrete time is considered continuous curves are drawn by linear connections between values to simplify the pictures.

2.1 *CUSUM*

Page (1954) suggested that the cumulative sums of observed values should be used in a specific way to detect a shift in the mean of a normal distribution. His suggestion was that you calculate C_t=sum(x_i-μ^0), i=1,...,t , and that there will be an alarm for the first t with |C_t-C_{t-i}| is greater than h+ki for some i, and C_0=0. Sometimes (see e.g. Siegmund 1985) the CUSUM test is presented in a more general way by likelihood ratios (which in the normal case reduce to C_t-C_{t-i}). The test might be performed by moving a V-shaped mask over a diagram until any earlier observation is outside the limits of the mask (see Figure 2). Thus the method is often referred to as "the V-mask method". Another name used in some fields of the literature is "Hinkley's method".

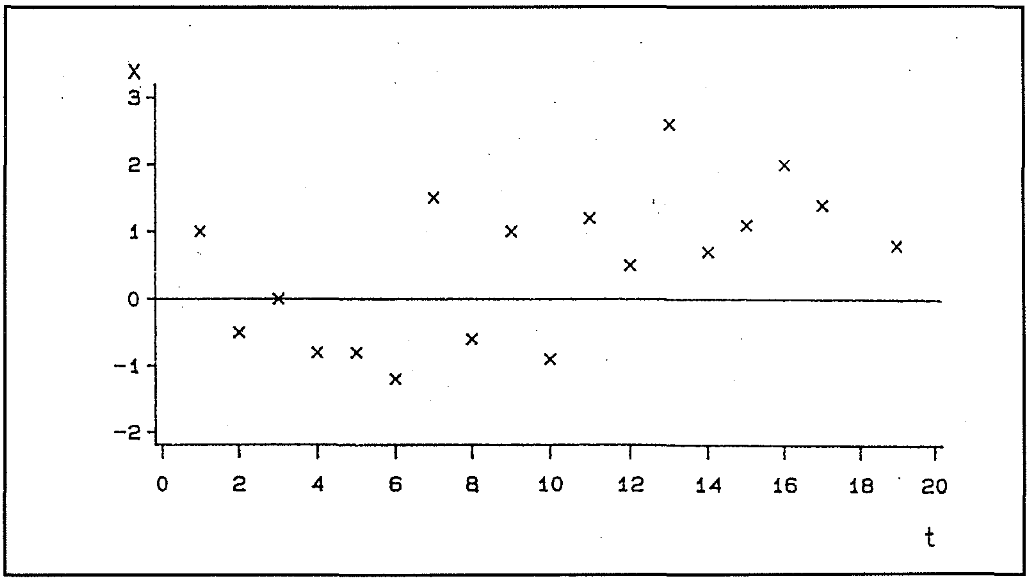

Figure 1. *The observed values X for each time t were generated by Lucas and Crosier (1982) by a process with constant mean (zero) for the first 10 observations and with a shift in mean of one standard deviation (one) for the last observations.*

The properties of this method are determined by the value of the parameters k and h. The information from earlier observations is handled differently depending on the position in the time series. Recent observations have more weight than old ones. If h=0, the V-MASK-test degenerate to a SHEWHART-test with the alarm-limit equal to k. With a shift of size μ^1-μ^0 and a constant variance, k=(μ^1-μ^0)/2 is usually recommended (see e.g. Bissel (1969)). This value of k is supposed to give a test having the shortest ARL^1 (for this specific shift) for a given ARL^0. Here the main aim is to demonstrate that important differences exist in spite of equal ARL. The choice of parameters in the examples is thus not important but was made for

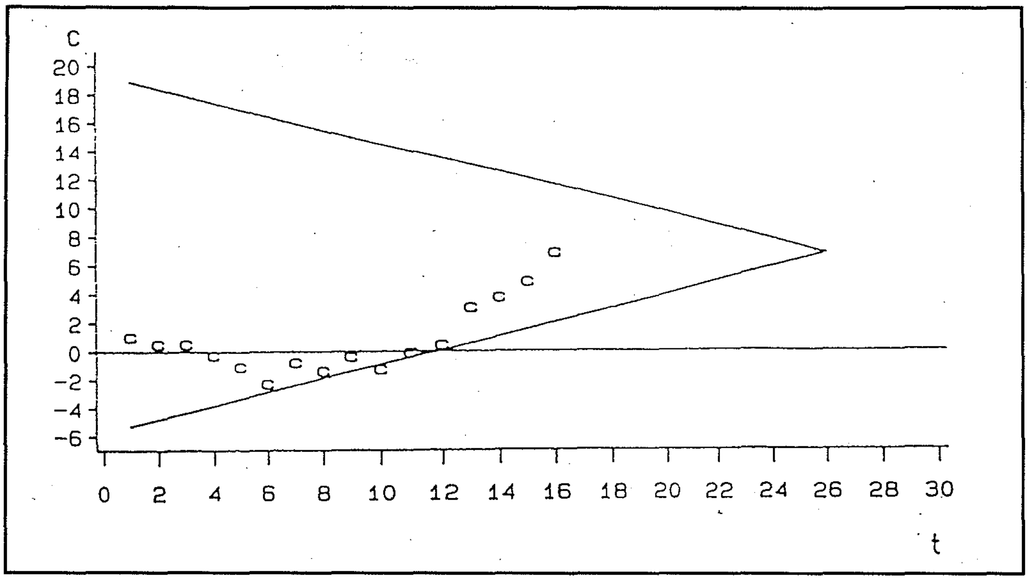

Figure 2. *CUSUM. An alarm occurs at the first time any C_t falls outside the V-shaped mask. In this case the first alarm is at time t=16*

practical reasons. The examples are however very similar to those in Lucas and Saccucci (1990)). The average run lengths have been fixed at $ARL^0=330$ and $ARL^1=9.7$. The parameter h, which determines the distance between the last observation and the apex of the "V" is set to 4.73 and the parameter k, which determines the slopes of the legs is set to 0.49. The alarm region for the first two steps is illustrated in Figure 6. The region for the first three steps is illustrated in Figure 7a.

Several variants of the method have been suggested. Lucas (1982) suggested a combination with the Shewhart method. Observe that a CUSUM always will give an alarm if any observation deviates more than h+k from the target value. Also the standard version of CUSUM can thus be regarded as a combination with a Shewhart test with the limit h+k. Yashchin (1989) has suggested that the weights of different observations should be separately chosen to meet some specific purposes. Here the original version of the method by Page (1954) is studied. The method has certain optimality properties as described in Moustakides (1986), Pollak (1987) and Frisén and de Maré (1991).

2.2 *EXPONENTIALLY WEIGHTED MOVING AVERAGE*

Exponentially weighted forecasts have been advocated by e.g. Muth (1960). A method for surveillance based on exponentially weighted moving averages, here called EWMA, was introduced in the quality control literature by Roberts (1959) but has for a long time been rarely used. Recently it has got more attention as a process monitoring and control tool. This may be due to papers by Robinson and Ho (1978), Crowder (1987), Lucas and Saccucci (1990), Ng and Chase (1989) and Domange and Patch (1991) in which techniques to study the properties of the method and also positive reports of the quality of the method are given.

The statistic is

$$Z_i = (1-\lambda)Z_{i-1} + \lambda X_i, \quad i=1,2,..$$

where $0<\lambda<1$ and in the standard version of the method $Z_0 = \mu^0$. The statistic is sometimes referred to as a geometric moving average since it can equivalently be written as

$$Z_i=\lambda\sum_{j=0}^{i-1}(1-\lambda)^j X_{i-j}+(1-\lambda)^i Z_0$$

EWMA gives the most recent observation the greatest weight, and gives all previous observations geometrically decreasing weights. If λ is equal to one only the last observation is considered and the resulting test is a Shewhart test. If λ is near zero all observations have approximately the same weight.

If the observations are independent and have a common standard deviation σ_X, the standard deviation of Z_i is

$$\sigma_{Z_i}=\sqrt{\frac{\lambda}{2-\lambda}(1-(1-\lambda)^{2i})}\,\sigma_X$$

For the first observation σ_Z takes the value $\lambda\sigma_X$, and as i increases σ_Z increases to its limiting value

$$\sigma_Z=\sqrt{\frac{\lambda}{(2-\lambda)}}\,\sigma_X$$

An out-of-control alarm is given if the statistic $|Z_i|$ exceeds an alarm limit, usually chosen as $L\sigma_Z$, where L is a constant. It might seem natural (and is sometimes advocated) to use the actual value of the standard deviation of Zi. However, usually the limiting value σ_Z rather than σ_{Zi} is used in the alarm-limits for EWMA control charts (see e.g. Roberts (1959), Robinson and Ho (1978), Crowder (1989) and Lucas and Saccucci (1990)). For a two-sided control chart this results in two straight warning-limits, one on each side of the nominal level of Z.

This variant is therefore called the "straight EWMA" in the following. See Figure 3.

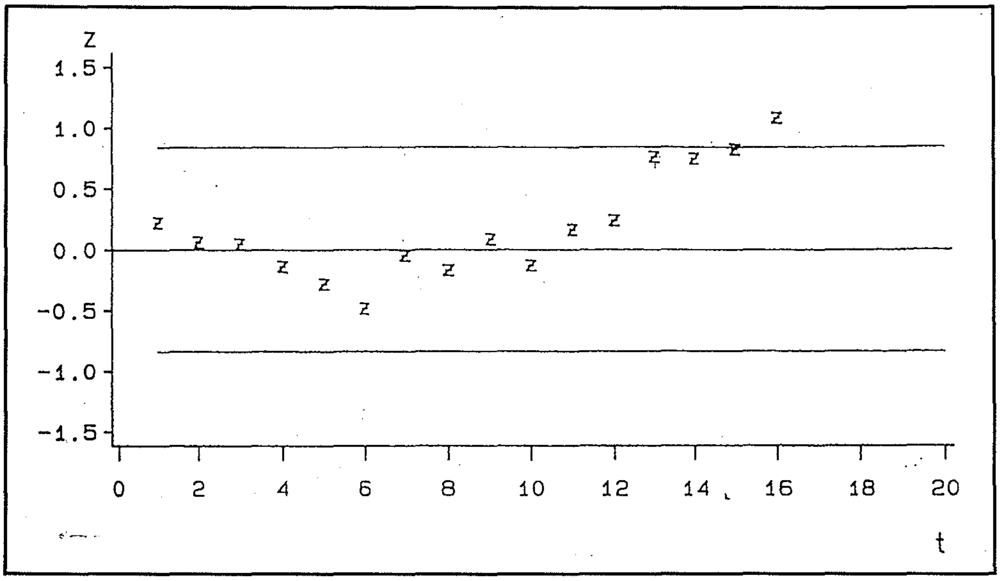

Figure 3. *Straight EWMA. Z denotes the exponentially weighted sum of the observations X. The straight alarm limits are at a distance $L\sigma_Z$ from the target value.*

By using σ_{Zi} the alarm limits start at a distance of $L\lambda\sigma_X$ from the target value and increases to $L\sigma_Z$. This variant is called "variance corrected EWMA" in the following. See Figure 4.

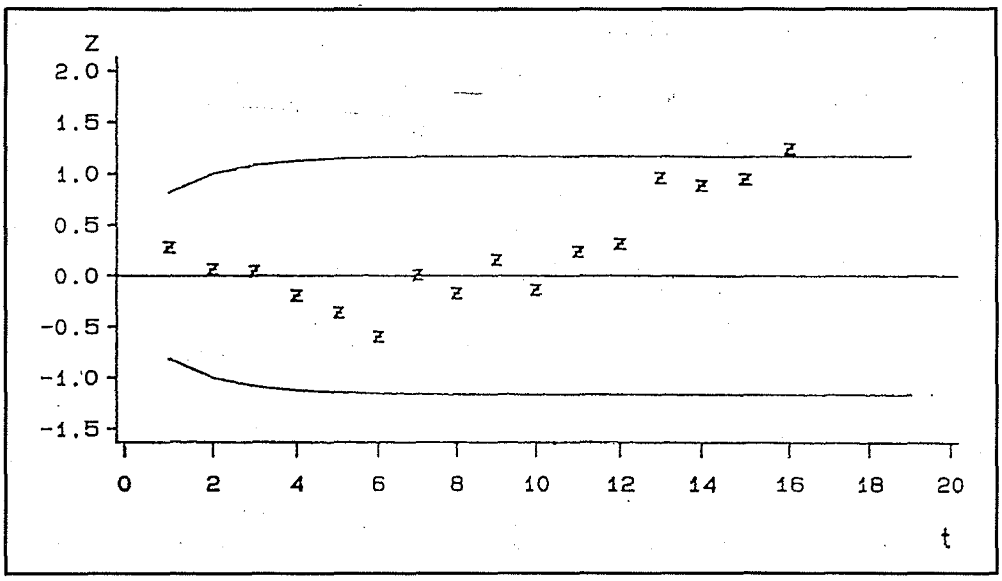

Figure 4. *Variance corrected EWMA. The alarm limits are based on the actual values of the variance of Z for each time point.*

Lucas and Saccucci (1990) recommend that instead of the standard starting value $Z_0 = \mu^0 = 0$, another value should be used to achieve a "Fast Initial Response", FIR. Two one-sided EWMA control schemes are simultaneously implemented. One is implemented with Z_0 = a and one with Z_0 = -a. There is an alarm if any of the one-sided schemes exceeds its constant limit. We will now study the relation between the different variants of EWMA more closely and concentrate on the one-sided upper limits for simplicity.

Let

$$c = L\sigma_Z = L\sqrt{\lambda/(2-\lambda)}\,\sigma_X$$

The straight EWMA gives alarm for $Z_i > c$.

The variance corrected EWMA gives alarm for

$$Z_i > c\sqrt{1-(1-\lambda)^{2i}}$$

The FIR have the same alarm value c as the straight EWMA but because of the starting value we have

$$Z_i' = Z_i + a(1 - \lambda)^i > c$$

that is

$$Z_i > c - a(1 - \lambda)^i$$

If

$$a = L\sigma_X\{(\lambda/(2 - \lambda))^{1/2} - \lambda\}/(1 - \lambda)$$

then the upper limit for the first observation will be the same as for the variance corrected EWMA which has the limit

$$L\sigma_{Z_i}$$

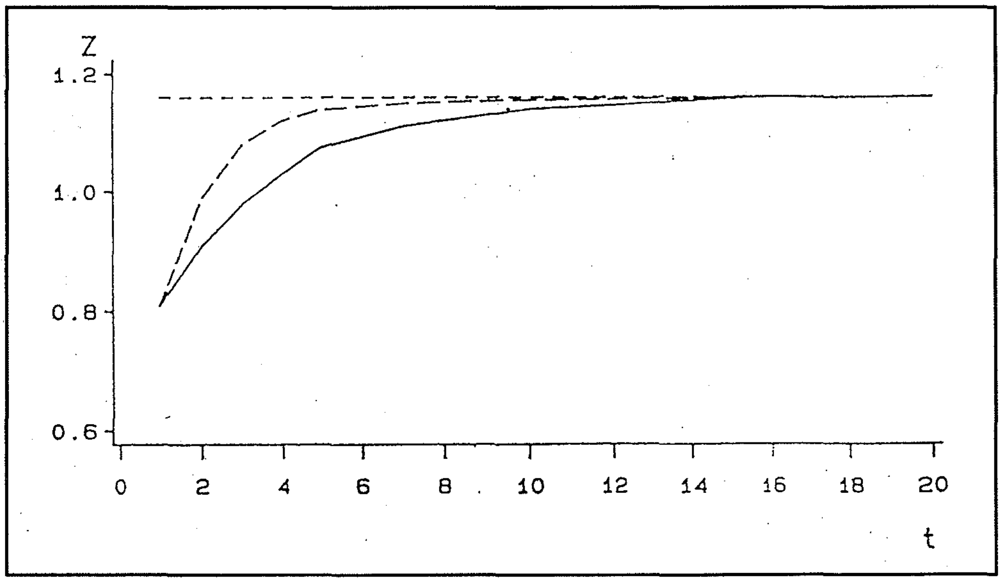

Figure 5. *Straight EWMA------, Variance corrected EWMA— — —, FIR————*

Both the FIR and the variance corrected EWMA have the same alarm limit as the straight EWMA for late observations. However the limits will converge faster to the constant limit for the last mentioned method than for the FIR method for all values of λ as can be proved by direct evaluation of the difference between the limits. See Figure 5 where the three variants (with the same λ and L as used for the variance corrected method in the other figures) are compared. In this figure the parameters (λ=0.283 and L=2.858) are not chosen to give the same ARL but to give the alarm limit the same asymptotic value and to give the FIR and the variance corrected variants the same limit at time t=1.

Also other variants of EWMA have been proposed, e.g. for multivariate problems (Lowry et.al. 1992). In the present study the characteristics of the straight and the variance corrected EWMA as described above are studied in detail. The parameter values are chosen to give the same average run lengths (see below) as the CUSUM both when there is no shift and when there is a shift to $\mu^1 = 1$, ARL^0=330 and ARL^1=9.7. The parameter values are for the straight EWMA L=2.385 and λ=.220 and for the variance corrected EWMA L=2.858 and λ=.283. Except for Figure 5 these parameter values are used in all figures and simulations. Alarm regions for the first two steps are illustrated in Figure 6. The alarm region for the first three steps is illustrated for the variance corrected EWMA in Figure 7.

The EWMA is not exactly optimal in the sense of Frisén and de Maré (1991) for any situation.

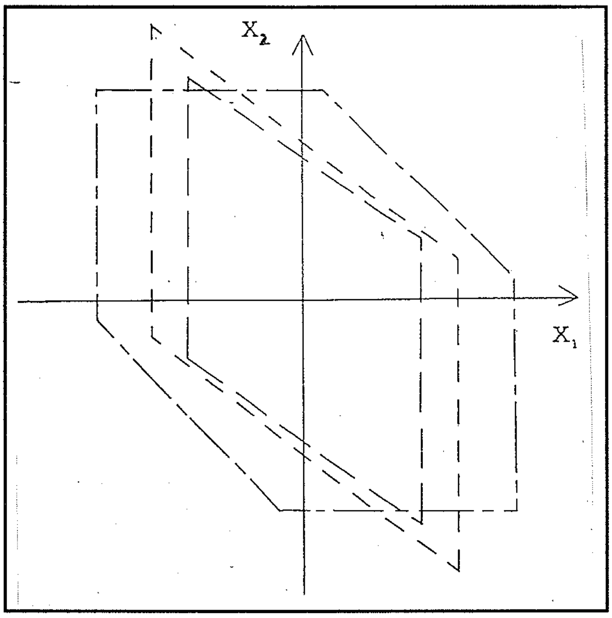

Figure 6. Detailed comparison between CUSUM and EWMA for the first two observations. The parameters in this and the following figures are the same as in Figure 2 - 4. Limits for alarm not later than at the second observation.
CUSUM —— - ——,
Straight EWMA- - - - -,
Variance corrected EWMA— — — —.

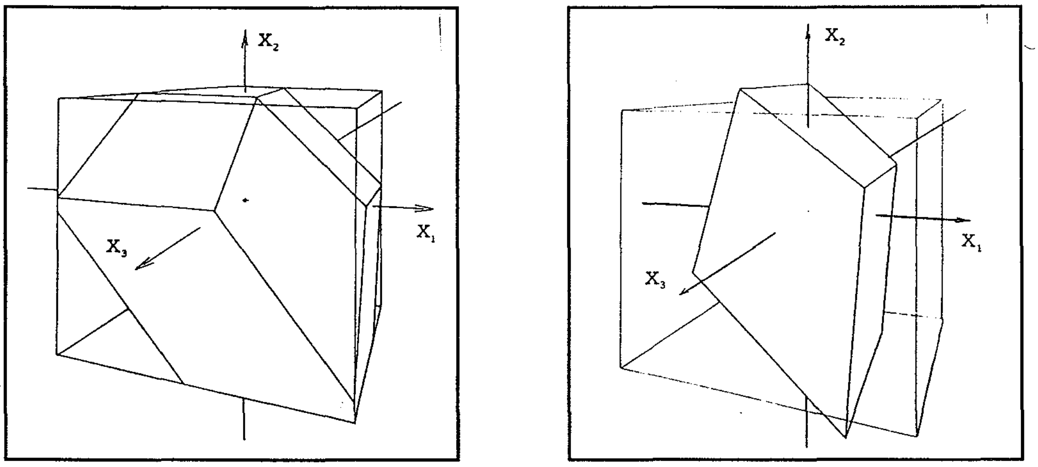

Figure 7.Limits for alarm not later than at the third observation. For reference the cube that is the limit for the Shewhart method with alarm limit 5.22 for each time point is included.
a.CUSUM b.Variance corrected EWMA

3. MEASURES OF THE PERFORMANCE

3.1 *RUN LENGTH DISTRIBUTION*

The run-length distributions for all interesting cases (also those where the change appears after the start of the surveillance) contains the information necessary for an evaluation of a method or a comparison between some methods. The actual comparison is usually based on some of the run-length distributions characteristics, mostly the average run length, but also the median or some other percentile could be considered. Several authors e.g. Zacks (1980), Crowder (1987) and Yashchin (1989) have pointed out that only one summarizing measure of the distribution is not enough. Run-length distributions are usually skew to the left, especially those connected to the alternative hypotheses (see Figures 8 - 11).

3.2 *ARL*

A measure which is often used in quality control is the average run length (ARL) until an alarm e.g. Wetherill and Brown (1990). It was suggested already by Page (1954). The average run length under the hypothesis of a stable process, ARL^0, is the average number of runs before an alarm when there is no change in the system under surveillance. The average run length under the alternative hypothesis, ARL^1, is the mean number of decisions that must be taken to detect a true level change that occurred at the same time as the inspection started.

Values of the ARL are much used information for the design of control charts for specific applications. Roberts (1966) has given very useful diagrams of the ARL. Later several authors e.g. Saccucci and Lucas (1990), Champ and Rigdon (1991), Champ et.al. (1991), Yashchin (1992) and Yashchin (1993) have studied the ARL of specific methods and models. However, ARL-curves do not contain all information about the methods. The distribution of the "run length" is generally markedly skew, so the ARL will give limited information. This has been pointed out by e.g. Woodall (1983).

Since both the EWMA and the CUSUM methods have two parameters they can be constructed to give the same ARL both for the null- and for an alternative situation (here $\mu^1=1$). By the choice of design parameters ARL^0 is set to 330 and ARL^1 to 9.7 for the methods compared below. Here the remaining differences are of main interest.

Because of a complicated time dependence, and the dependence of the incidence of the change to be detected, other measures (Frisén 1986, 1992) than the average run length should be considered in the evaluation of different methods. Beckman et al. (1990) advocate similar measures as those in Sections 3.4 and 3.5 for the case of flood warning systems.

3.3 *THE PROBABILITY OF FALSE ALARM*

The distribution when the process is under control is described by a measure α_t which corresponds to the probability of erroneous rejection of the null hypothesis, the level of significance, but is a function of the time t. α_t is the probability of an alarm no later than at t given that no change has occurred.

$$\alpha_t = P(RL \leq t \mid \mu_t = \mu^0)$$

3.4 *THE PROBABILITY OF SUCCESSFUL DETECTION*

The distance between the change and the alarm, sometimes called "residual RL" (RRL) is of interest in many cases. The optimality conditions by Girshick and Rubin (1952) and Shiryaev (1963) are based on this distance. One characterization of the distribution of the RRL is the probability that the RRL is less than a certain constant d (the time limit for successful rescuing action). This measure, PSD(d), the probability of successful detection, is the probability to get an alarm within d time units after the change has occurred, conditioned that there was no alarm before the change. The PSD is a function of the time distance d, the time of the change t' and μ^1.

$$PSD(d, t', \mu^1) = P(RL < t' + d \mid RL \geq t')$$

3.5 *PREDICTIVE VALUE*

PV, the predictive value of an alarm is the relative frequency of motivated alarms among all alarms at a certain point in time. This measure is a function of the incidence inc, μ^1 and the time t" of the alarm. It gives information on whether an alarm is a strong indication of a change or not. Let τ be the (stochastic) time of change, then

$$PV(t'', inc, \mu^1) = P(\tau \leq t'' \mid RL = t'').$$

Sometimes a late alarm is regarded with some doubt (cf. e.g. Johnson 1961). This might be for the same reason as a significant result at a very big sample size is considered less impressing than a significant result at a small sample size. However there is no analogy here unless you only consider cases where the change appears at the same time as the surveillance starts. The trust you should have in an alarm is measured by the predictive value.

4. RESULTS

The alarm regions up to the first two observations are given in Figure 6. Considering the first and second observation the CUSUM has an "acceptance region" which contains that of the straight EWMA-method, except the extreme situation with two observations on the boundary, one in each direction. This is the case called "worst possible" discussed by Yashchin (1987). The differences in size of the areas illustrate the different alarm probabilities at the first time points. Notable is also the shape of the regions, determined by the choice of the weight parameter λ and the reference value k.

In Figure 7 above, the tree-dimensional regions of alarm at any of the runs 1, 2 or 3 are given.

In Figures 8 and 9 below, the cumulative probabilities of false alarms illustrate the differences (in spite of equal ARL) between the methods. The probabilities are estimated by simulation of at least 100,000 replicates of each situation. The variance corrected EWMA has a greater probability (about 1%) of false alarm in a great part of the beginning than the straight EWMA which in turn has a slightly greater false alarm rate at the start than the CUSUM. The median is much smaller (about 230) than the ARL (330) which illustrates the skewness of the distribution. The probability to exceed the ARL is about 30%.

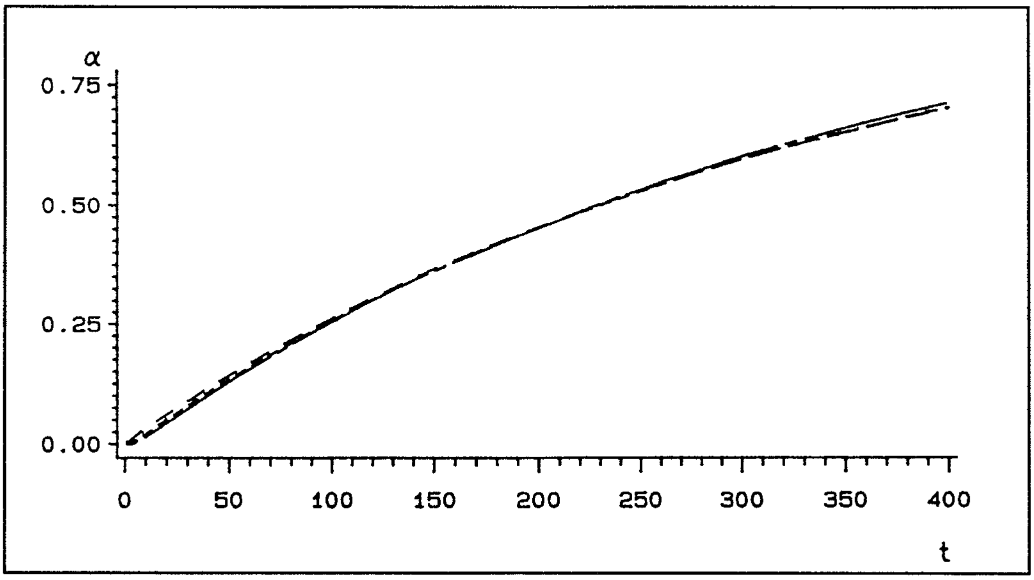

Figure 8. *The probability α of an alarm not later than at time t given that no change has occurred. The lines are linear connections between the values for each time point. Overview up to t=400.*
CUSUM —— - ——
Straight EWMA - - - - - -
Variance corrected EWMA — — — —

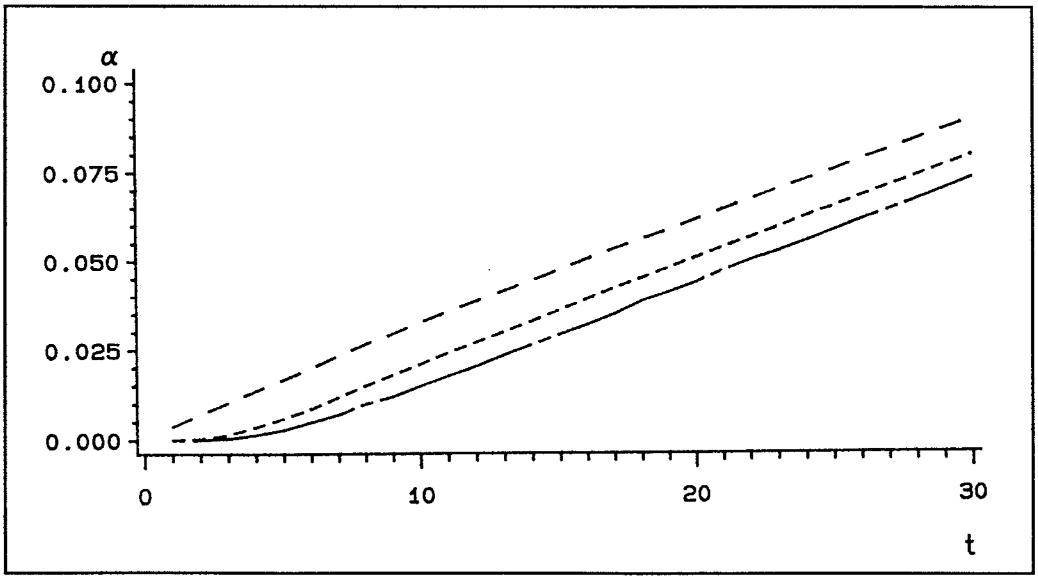

Figure 9. *As Figure 8, but detailed picture up to t=* 30

In Figures 10 and 11 the probability distributions of the residual run length are given for different times of shift. The results are based on at least 40,000 replicates. In Figure 10 it is given for the case where the shift occurs at the same time as the surveillance starts. ARL[1] is the expected value under this assumption. It is the same 9.7 for all methods. The median is however one unit less (equal to 7) for the variance corrected EWMA than for the CUSUM (equal to 8) which is an indication of the different shapes of the distributions, as is also seen in the figure. The EWMA has a higher density for small run lengths. In Figure 11 the distributions are given for the case where the shift occurs at the 9th run. The expected value in these distributions are not at all equal to ARL^1.

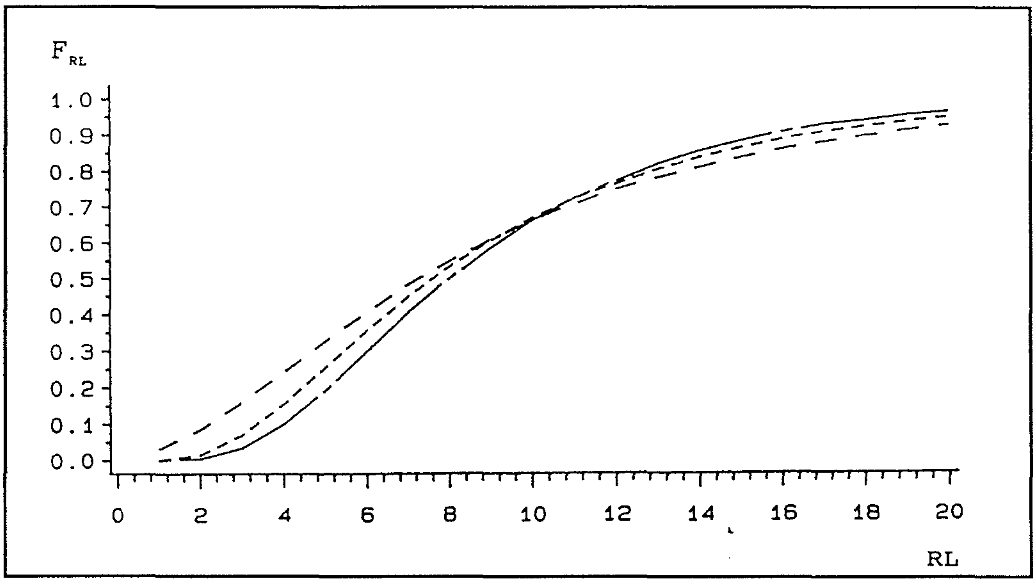

Figure 10. *The distribution of run lengths after a shift to $\mu^1=1$ at different times t'. The run length distribution F_{RL} when t'=1.*
CUSUM ——·——
Straight EWMA - - - - - -
Variance corrected EWMA — — — —

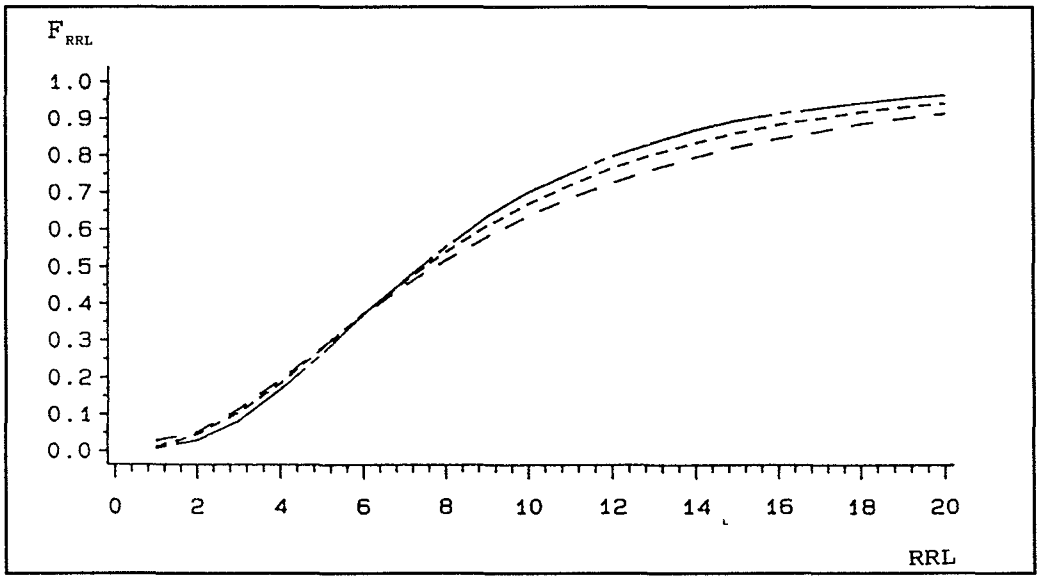

Figure 11. *As Figure 10, but the residual run length distribution F_{RRL} when t'=9.*

In Figure 12 it is demonstrated that the probability of successful detection within d=1 unit, that is immediately after the shift, is best for the variance corrected EWMA and better for the straight EWMA than for the CUSUM. The differences are most pronounced if the shift occurs soon after the surveillance has started (small t').

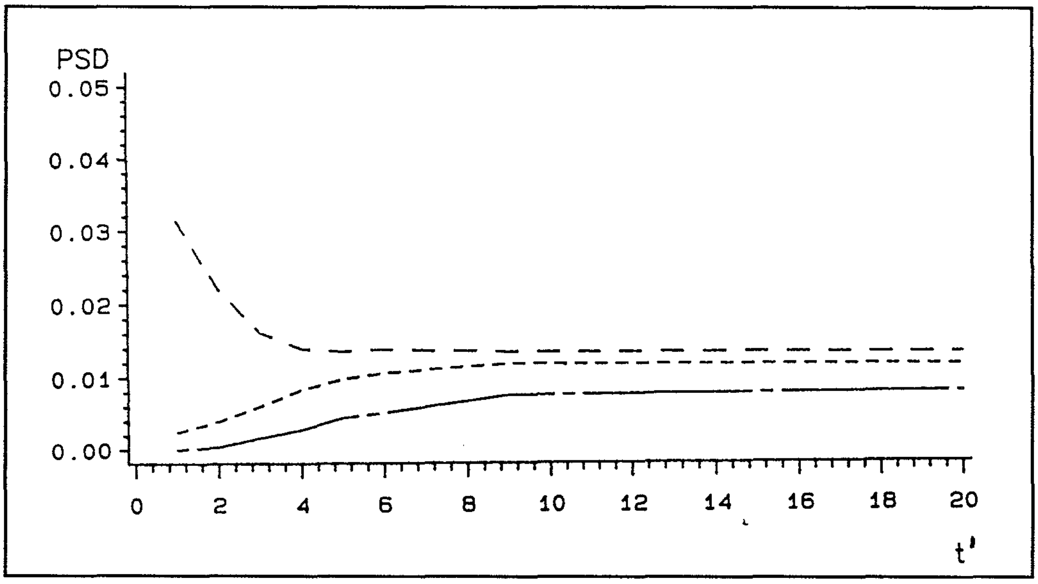

Figure 12. *Probability of successful detection, PSD. The probability of an alarm within d time units after the time t' of a shift to* $\mu^1=1$*, given that there was no alarm before t'. d=1.*

CUSUM ——- ——
Straight EWMA - - - - - -
Variance corrected EWMA — — — —

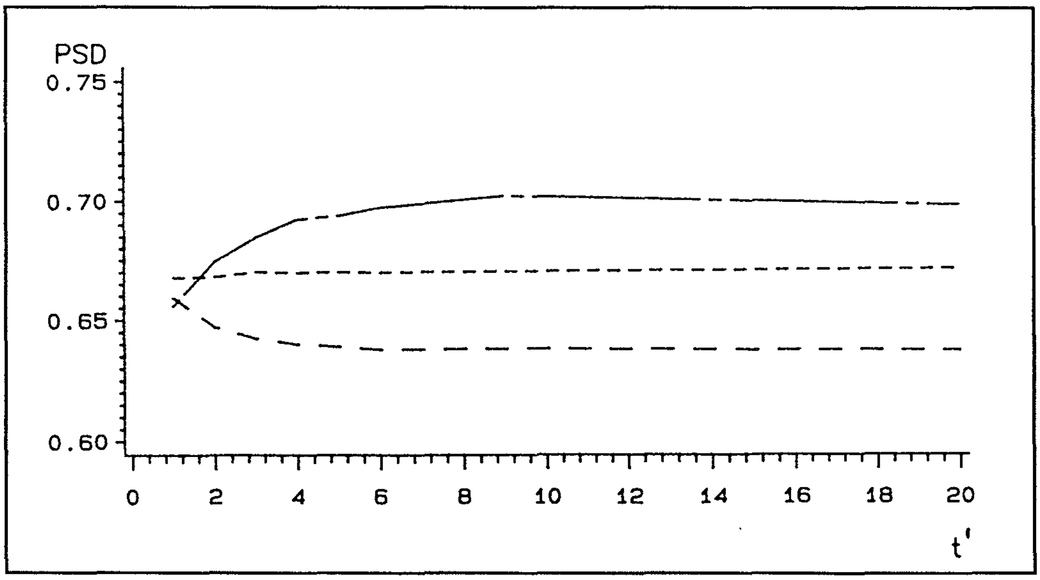

Figure 13. *As Figure 12, but d=10.*

Each point in Figures 12 an 13 is based on at least 40,000 replicates.In Figure 13 it is demonstrated that the differences are in the opposite direction for detection within d=10 units. Then the differences are least pronounced soon after the start of the surveillance.

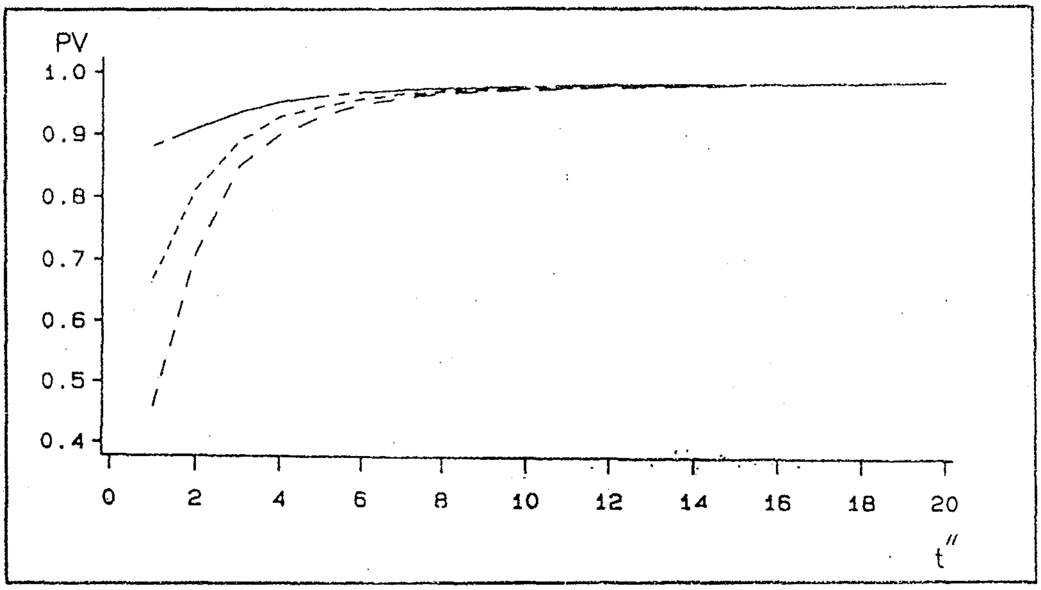

Figure 14. *Predicted value of an alarm, PV. The incidence of a shift to $\mu^1=1$ is 0.1.*
CUSUM —·—
Straight EWMA - - - - - -
Variance corrected EWMA — — — —

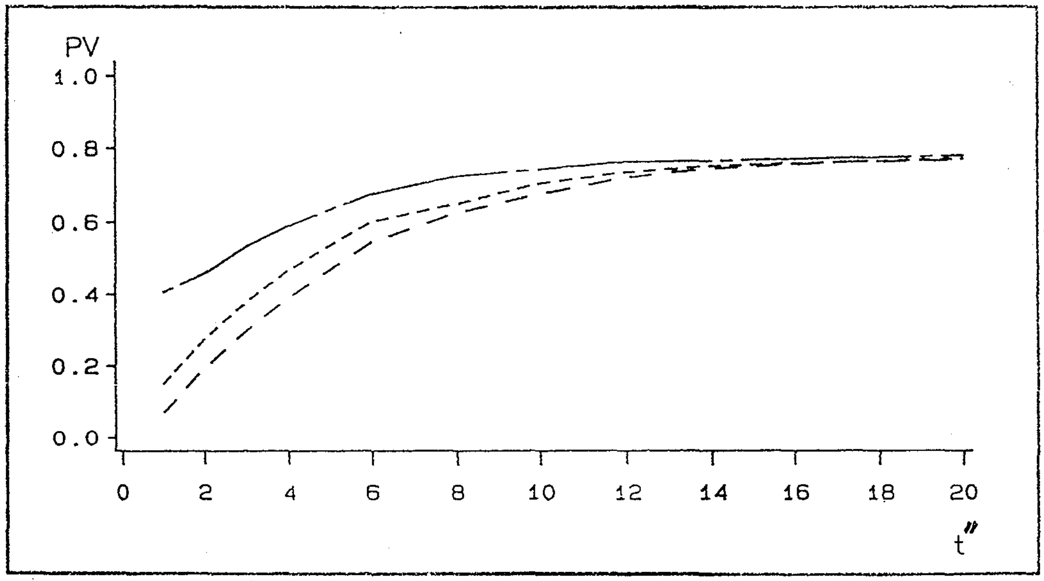

Figure 15. *As Figure 14 but the incidence of a shift to $\mu^1=1$ is 0.01*

In Figures 14 and 15 it is seen that the predicted value is low and varying for the EWMA methods (specially the variance corrected one) at early time-points. This implies that early alarms for the EWMA are very hard to interpret.

Each point in the figures is calculated as a function of the probabilities of false alarms and motivated alarms. The probabilities of false alarms are estimated by simulations of at least 100,000 replicates while the estimates of the probabilities of motivated alarms are based on at least 40,000 replicates. For t"=1 the probabilities are calculated exactly and for t"= 2, 3 and 4 the probabilities of motivated alarms are based on 1,000,000 replicates. The fact that the curve (for small values of t") in Figure 15 is not a smooth one is thus not due to uncertainty in the simulations. In fact, the predicted value is not always an increasing function of t" (Frisén (1992)).

5. DISCUSSION

As was also commented in the results, Figures 6 and 7 illustrate a difference in shape of the alarm region between the EWMA and the CUSUM which is general and which explains why the EWMA has bad "worst possible" properties (Yashchin 1987) while the CUSUM has minimax optimality (Moustakides 1986).

In Figures 6 and 7 interesting differences in symmetry are also illustrated. The alarm area is symmetrical for the CUSUM but not for the EWMA methods. That is, for the probability of an alarm not later than at t all observations up to X_t have the same weight for the CUSUM. For the EWMA methods the older ones have more weight. However for the probability of an alarm at time t the last observations have the greatest weight both for CUSUM and EWMA methods.

In Figures 10 and 11 it is also demonstrated that for changes which occurred at the same time as the surveillance started the probability of a detection within a short time (shorter than 10) is better for the examined EWMA methods than for the CUSUM, while the opposite is true for times longer than 10. If the shift occurs some time after the start the short time is less than 10. In most studies only the case of a shift at the same time as the start of the surveillance is studied. As was seen above CUSUM compares more favourable with EWMA in other cases.

As is illustrated by the relative size of the rejection areas in Figures 6 and 7, and more generally seen by the formulas for the methods, the examined EWMA methods have a higher probability than the CUSUM for alarms shortly after the surveillance has started - both false and motivated ones. This does not mean that it is higher shortly after the shift has appeared, if the shift occurs later, as is seen in Figures 10-13.

One balance between the false and motivated alarms is given by the predictive value. In the simulations the predicted value is never better for the EWMA than for the CUSUM. In Figures 14 and 15 it is seen that the low and variable predicted value for the EWMA methods (specially the variance corrected one) at early time-points makes the early alarms for the EWMA methods very hard to interpret. This may make the variance corrected EWMA worthless shortly after the start. In the beginning when the predicted value of an alarm is very low and varying no alarm could be trusted. In the example with $ARL_1=9.7$ the alarms by the EWMA before the 9th run have so low predicted value that they for most applications must be disregarded. Thus the benefit of a higher probability of an alarm in the beginning cannot be taken advantage of.

The general conclusion from the comparisons is that there might be important differences in characteristics in spite of equal ARL^0 and ARL^1. Even though only one set of parameters were examined for each method this is enough to demonstrate that differences exist.

In this paper only constant incidences are considered and the above discussion is relevant for this case. However, in some applications a higher incidence at the start of the surveillance might be relevant. The properties of the EWMA methods (especially the FIR variant) will then be more favourable. Only an approximately constant predictive value makes the method easily usable since only then it is possible to have the same kind of action independently of how far from the start the alarm is.

ACKNOWLEDGEMENT

This work has been supported by the Swedish Council for Research in the Humanities and Social Sciences.

REFERENCES

Beckman, S-I., Holst, J. and Lindgren, G. (1990) "Alarm characteristics for a flood warning system with deterministic components," *Journal of Time Series Analysis*, 11, 1-18.

Bissel, A. F. (1969) "CUSUM techniques for quality control," *Applied Statistics*, 18, 1-30.

Champ, C. W. and Rigdon, S. E. (1991) "A comparison of the Markov chain and the integral equation approaches for evaluating the run length distribution of quality control charts,"*Communications in Statistics. Simulation Comput.,* 20, 191-204.

Champ, C. W., Woodall, W. H. and Mohsen, H. A. (1991), "A generalized quality control procedure,"*Statistics & Probability Letters*, 11, 211-218.

Crowder, S. V. (1987), "A simple method for studying run-length distribution of exponentially weighted moving average charts," *Technometrics*, 29, 401-407.

Crowder, S V. (1989), "Design of Exponentially Weighted Moving Average Schemes,". *Journal of Quality Technology*, 21, 155-162.

Domangue, R. and Patch, S. C. (1991), "Some omnibus exponentially weighted moving average statistical process monitoring schemes," *Technometrics,* 33, 299-313.

Frisén, M. (1986), "On measures of goodness of statistical surveillance," *Proceedings, First World Congress of the Bernoulli Society, Tashkent*.

Frisén, M. (1992), "Evaluations of methods for statistical surveillance," *Statistics in Medicine*, 11, 1489-1502.

Frisén, M. and de Maré, J. (1991), "Optimal surveillance," *Biometrika*, 78, 271-280.

Girshick, M. A. and Rubin, H. (1952), "A Bayes approach to a quality control model," *The Annals of Mathematical Statistics*, 23, 114-125.

Johnson, N. L. (1961), "A simple theoretical approach to cumulative sum control charts," *Journal of the American Statistical Association*, 56, 835-840.

Lowry, C. A., Woodall, W. H., Champ, C. W. and Rigdon, S.E. (1992) "A multivariate exponentially weighted moving average control chart," *Technometrics*, 34, 46-53.

Lucas, J. M. (1982), "Combined Shewhart-CUSUM Quality Control Schemes," *J of Qual Technology*, 12, 1451-59.

Lucas, J. M. and Crosier, R. B. (1982), "Fast initial response for cusum quality control schemes: give your cusum a head start," *Technometrics*, 24, 199-205.

Lucas, J. M. and Saccucci, M. S. (1990), "Exponentially weighted moving average control schemes: properties and enhancements," *Technometrics*, 32, 1-12.

Moustakides, G. V. (1986), "Optimal stopping times for detecting changes in distributions," *Annals of Statistics*, 1379-87.

Muth, J. F. (1960), "Optimal properties of exponentially weighted forecasts," *Journal of the American Statistical Association,* 55, 299-306.

Ng, C. H. and Case, K. E. (1989), "Development and Evaluation of Control Charts Using Exponentially Weighted Moving Averages," *Journal of Quality Technology*, 21, 242-250.

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

Pollak, M. (1987), "Average run length of an optimal method of detecting a change in distribution," *Annals of Statistics*, 15, 749-779.

Roberts, S. W. (1959), "Control Chart Tests Based on Geometric Moving Averages," *Technometrics*, 1, 239-250.

Roberts, S.W. (1966), "A comparison of some control chart procedures," *Technometrics*, 8, 411-430.

Robinson, P. B. and Ho, T. Y. (1978), "Average Run Lengths of Geometric Moving Average Charts by Numerical Methods," *Technometrics*, 20, 85-93.

Saccucci, M. S. and Lucas, J. M. (1990), "Average Run Lengths for Exponentially Weighted Moving Average Control Schemes Using the Markov Chain Approach," *Journal of Quality Technology*, 22, 154-162.

Shiryaev, A. N. (1963), "On optimum methods in quickest detection problems," *Theory of Probability and its Applications*, 8, 22-46.

Siegmund, D. (1985), *Sequential analysis. Tests and confidence intervals*, Springer.

Wetherill, G.B. and Brown, D.W. (1990), *Statistical process control*, London: Chapman and Hall.

Woodall, W.H. (1983), "The distribution of the run length of one-sided cusum procedures for continuous random variables," *Technometrics*, 25, 295-301.

Yashchin, E. (1987), "Some aspects of the theory of statistical control schemes," *IBM Journal of Research and develop*, 31, 199-205.

Yashchin, E. (1989), "Weighted Cumulative Sum Technique," *Technometrics*, 31, 321-338.

Yashchin, E. (1992), "Analysis of CUSUM and other Markov-type control schemes by using empirical distributions," *Technometrics*, 34, 54-63.

Yashchin, E. (1993), "Performance of CUSUM control schemes for serially correlated observations", *Technometrics*, 35, 37-52.

Zacks, S. (1980), "Numerical determination of the distributions of stopping variables associated with sequential procedures for detecting epochs of shift in distributions of discrete random variables," *Communications in Statistics. Sim & Comput*, 9, 1-18.

Zacks, S. (1983), "Survey of classical and Bayesian approaches to the change-point problem: Fixed sample and sequential procedures of testing and estimation," *Recent advances in statistics*, 245-269.

1992:1	Teräsvirta, T., Tjøstheim, D. & Granger, C.W.J.	Aspects of Modelling Nonlinear Time Series
1992:2	Palaszewski, B.	A conditional stepwise test for deviating parameters
1992::3	Guilbaud, O.	Exact Semiparametric Inference About the Within--Subject Variability in 2 x 2 Crossover Trails
1992:4	Svensson, E. & Holm, S.	Separation of systematic and random errors in ordinal rating scales