

**An Application of Filtering to  
Statistical Process Control**

**A. Thavaneswaran, B.D. Macpherson,  
and B. Abraham**  
*University of Waterloo*

**RR-98-02**

February 1998

# An Application of Filtering to Statistical Process Control

A. Thavaneswaran, B.D. Macpherson, and \*B. Abraham

*Department of Statistics, University of Manitoba, Winnipeg, Manitoba, Canada R3T 2N2*

*\*Institute for Improvement in Quality and Productivity, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1*

**Abstract:** There has been growing interest in the Kalman filter as an estimation technique in statistical process control. In cases where prior information about the process is available, procedures based on the 'optimal' [Godambe (1985)] smoother can be superior to the classical procedures like Shewhart and CUSUM control charts. We also discuss the relationship among EWMA, Kalman filtering and the 'optimal' smoother. This smoother and its applications are also illustrated through an example.

**Keywords and phrases:** ARIMA processes, estimating function, EWMA, Kalman filter, optimal filter, optimal smoother

---

## 1 Introduction

Recently there has been growing interest in the general theory of statistical process control through the use of exponentially weighted moving average (EWMA) charts [Hunter (1986), Montgomery and Mastrangelo (1991)], for autocorrelated data. In this paper an 'optimal' smoother is proposed; it is optimal in the sense that the smoother is a solution of the optimal estimating function, [see for example, Thavaneswaran and Abraham (1988), Godambe, (1985)]. This smoother essentially incorporates the correlation structure of the underlying process, and leads to a control chart with better properties than the EWMA chart.

In the literature two different methods for constructing control charts are proposed for correlated data. In the first method, the basic idea is to model the autocorrelation structure in the original process using an autoregressive integrated moving average (ARIMA) model and apply control charts to the residuals. The second method uses a control chart based on the EWMA statistic, a function of the one step ahead forecast errors. The exponentially weighted moving average statistic gives a procedure which is optimal (in the minimum

mean square error (MSE) sense) for a limited class of ARIMA (p,d,q) processes with  $(p, d, q) = (0, 1, 1)$  [see for example, Abraham and Ledolter (1986)].

Shewhart Control Charts and other Statistical Process Control (SPC) techniques are very useful in industry for process improvement, estimation of process parameters and determination of process capabilities. The assumption of uncorrelated observations is fundamental to the use of the Shewhart control charts [Hunter (1986)]. In this situation a simple model that is used for the observations is:

$$X_t = \mu + \epsilon_t$$

where  $\mu$  is the process mean and  $\epsilon_t$  ( $t = 1, 2, \dots$ ) are independent identically distributed (iid) random variables with mean zero and variance  $\sigma^2$ .

The existence of autocorrelated errors violates the conditions of this model and failure to detect, or ignoring autocorrelations can lead to misleading results. Detection of autocorrelation can be accomplished through diagnostic plots or through a formal test. A simple plot of the residuals from the model can be helpful. If the residuals are plotted against time, and unusually large numbers of residuals with the same sign are observed clustered together, then this is an indication that the errors are governed by positive autocorrelation. On the other hand, rapid changes in sign may indicate the presence of negative autocorrelation. Positively correlated errors can lead to substantial underestimation of  $\sigma^2$  and an increase in the frequency of false alarms; in other words, the in control Average Run Length (ARL) is much shorter than it would be for a process with uncorrelated observations. Thus often, the state of control of the process can not be determined from the usual control charts.

In Section 2, the ARIMA (p,d,q) modeling approach to quality control and EWMA are briefly discussed. Section 3 provides the filtering and prediction algorithms based on estimating functions. It is also shown there that Kalman filtering algorithm does not take into account the autocorrelation structure of the observed process. In addition we provide control charts for correlated observations using a smoother, the form of which depends on the autocorrelation structure of the observed process of interest. Section 4 provides some special cases and an application. Section 5 gives some concluding remarks.

---

## 2 Control Charts for Autocorrelated Data

### 2.1 Modeling the autocorrelations using ARIMA (p,d,q) models

The primary approach is to fit an appropriate time series model to the observations and apply a Shewhart control chart to the residuals from this model. A commonly used time series model is the Autoregressive Integrated Moving Average model which is given by

$$\begin{aligned} \phi_p(B)(1-B)^d X_t &= \theta_q(B)\epsilon_t \\ \text{where } \phi_p(B) &= 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \\ \theta_q(B) &= 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q \end{aligned}$$

and the  $\epsilon_t$ 's are iid random variables with mean zero and variance  $\sigma^2$ .

If  $\hat{X}_t$ 's are the predicted values from a fitted ARIMA model, then

$$e_t = X_t - \hat{X}_t, \quad t = 1, 2, \dots, n$$

are the residuals which are considered to be approximately identically distributed independent random variables. The process  $X_t$  will be declared 'out of control' if a mean shift is detected on the control chart applied to the residuals.

### 2.2 An application of the EWMA statistic to autocorrelated data

The EWMA approach was first suggested by Roberts (1959) and has been discussed by several authors [for example, Abraham and Kartha (1978,1979)]. The EWMA statistic  $Z_t$  is defined as

$$Z_t = \lambda X_t + (1 - \lambda)Z_{t-1}, \quad 0 < \lambda < 1.$$

Montgomery and Mastrangelo (1991) and Hunter (1986) have shown that if the observations are uncorrelated, the control limits for the EWMA control chart under steady state conditions are given by :

$$\begin{aligned} LCL &= \bar{X} - 3\sigma\sqrt{\frac{\lambda}{n(2-\lambda)}} \\ UCL &= \bar{X} + 3\sigma\sqrt{\frac{\lambda}{n(2-\lambda)}} \end{aligned}$$

where  $\bar{X}$  is the overall average and  $\sigma$  is the process standard deviation. The EWMA can also be used for autocorrelated data. As an illustration, consider a situation where the data can be modelled as an ARIMA (0,1,1) process:

$$X_t = X_{t-1} + \epsilon_t - \theta\epsilon_{t-1}.$$

When  $\lambda = 1 - \theta$ , the EWMA is the optimal one step ahead forecast (prediction) for this process ( i.e.  $Z_t = \hat{X}_{t+1/t}$ ), the one step ahead forecast of  $X_{t+1}$  made at time  $t$ ). In this case, the one step ahead forecast errors  $\hat{X}_t - \hat{X}_{t/t-1}$ ,  $t = 2, 3, \dots$  are independent with mean zero and standard deviation  $\sigma$  if the fitted ARIMA (0,1,1) model is correct. Thus, we could set up control charts for the one step ahead forecast errors.

Montgomery and Mastrangelo (1991) argue that generally the EWMA, with a suitable  $\lambda$ , will give an “excellent one step ahead forecast” even if the observations from the process are positively autocorrelated or the process mean does not drift too quickly. In addition they indicate that the EWMA provides good forecasts for models which are not exactly ARIMA (0,1,1) and that some processes which follow a slow random walk, can be well represented by the ARIMA(0,1,1) model. We show however that the use of the EWMA in these situations results in a loss of almost 50% efficiency compared to the smoother based on the optimal estimating function.

### 3 Optimal Filter and Smoother

Let us consider a simple time series model,

$$X_t = \phi X_{t-1} + \epsilon_t$$

where  $\{\epsilon_t\}$  is an iid sequence with mean zero and variance  $\sigma^2$  and  $|\phi| < 1$ . We can define two different means and variances. The unconditional mean of  $X_t$  is zero and the unconditional variance of  $X_t$  is  $\frac{\sigma^2}{1-\phi^2}$ . In addition, the conditional mean and variance of  $X_t$  given  $X_{t-1}, X_{t-2}, \dots, X_1$  are given by

$$E[X_t | X_{t-1}, X_{t-2}, \dots, X_1] = \phi X_{t-1}$$

and

$$\text{Var} [X_t | X_{t-1}, X_{t-2}, \dots, X_1] = \sigma^2$$

respectively.

Most of the recent inferences for non-linear time series models are based on conditional moments. In the next section we use the conditional moments to obtain an optimal filter and compare this with the Kalman filter.

#### 3.1 Optimal filter

Consider a process of the form

$$X_t = \theta_t + N_t \quad t = 1, \dots, n, \dots$$

where  $\{N_t\}$  is an autocorrelated sequence with zero mean and known covariance structure. Estimation of  $\theta_t, t = 1, 2, \dots, n$  from  $n$ -observations on  $\{X_t\}$  without making any restrictions on the parameter sequence is an ill-posed problem. In order to obtain a filtered estimate of  $\theta_t$  we assume that  $\theta_t$  follows a random walk

$$\theta_t = \theta_{t-1} + a_t$$

where  $\{a_t\}$  is a white noise sequence with mean zero and variance  $\sigma_a^2$ .

The following theorem on optimal estimation of  $\theta_t$  obtained by identifying the sources of variation and optimally combining elementary estimating functions [see Heyde (1987)] from each of these sources, gives the filtering formula for  $\theta_t$ .

**Theorem 1:** Let  $F_{t-1}^x$  be the  $\sigma$ -field generated by  $X_1, \dots, X_{t-1}$ . The recursive form of the estimator of  $\theta_t$  based on  $X_1, \dots, X_t$  is given by:

$$\hat{\theta}_t = \hat{\theta}_{t/t-1} + P_{t/t-1} (X_t - \hat{X}_{t/t-1}) / r_t$$

where  $P_{t/t-1} = E \left[ (\theta_t - \hat{\theta}_{t/t-1})^2 | F_{t-1}^x \right]$ ,  $\hat{\theta}_{t/t-1}$  is an estimate of  $\theta_t$  based on  $X_{t-1}, \dots, X_1$ ,  $\hat{X}_{t/t-1}$  is the predictor of  $X_t$  based on  $X_{t-1}, \dots, X_1$ , and  $r_t = E \left[ (X_t - \hat{X}_{t/t-1})^2 | F_{t-1}^x \right]$  its mean square error. Moreover,

$$P_{t/t-1} = P_{t-1/t-1} + \sigma_a^2$$

and

$$\frac{1}{P_{t/t}} = \frac{1}{P_{t/t-1}} + \frac{1}{r_t}$$

**Proof:** The elementary estimating function for fixed  $\theta_t$  based on the  $t^{\text{th}}$  observation is

$$h_{1t} = (X_t - \hat{X}_{t/t-1}) / r_t$$

and the corresponding information [see for example, Abraham et al (1997)] associated with  $h_{1t}$  is  $1/r_t$ .

The elementary estimating function for the parameter of interest, say  $\theta$ , is

$$(X_t - \hat{X}_t) \frac{\partial \hat{X}_t}{\partial \theta} / r_t.$$

Now  $h_{2t}$ , the corresponding estimating function for  $\theta_t$  based on prior information is

$$h_{2t} = (\theta - \hat{\theta}_{t/t-1}) / P_{t/t-1}.$$

Combining the estimating functions  $h_{1t}$  and  $h_{2t}$  the Theorem follows.

Under normality assumption on the errors, the Kalman filtering algorithm turns out to be a special case of Theorem 1.

It is of interest to note that the above filtering algorithm, which could be used in a wider context ( for non-Gaussian linear processes as well ), provides the estimator in terms of the first two conditional moments of the observed process. The same argument is also true for Kalman filtering. However, for many industrial situations the observed processes are autocorrelated and the Kalman filtering formula does not incorporate the autocorrelation structure. In Section 1.3.2 we present a smoother which incorporates the autocorrelation structure and is different from the Kalman filter.

### 3.2 Optimal smoother

In Abraham et al. (1997), an optimal smoother based on conditional moments has been proposed for nonlinear time series models. In this section an optimal smoother which is a solution of an optimal estimating function (based on moments of the observed process) is proposed for linear processes. In this context, the control limits of the chart for monitoring the process level (conditional mean) with the optimal smoother are functions of the autocorrelation of the observed process of interest.

For an industrial process  $\{X_t\}$  having mean zero and covariances  $r(t, s)$  the following theorem provides the form of the optimal smoother.

**Theorem 2:** The optimal predictor of  $X_t$ , based on  $X_{t-1}, X_{t-2}, \dots, X_1$  is given by

$$\hat{X}_{t/t-1} = \sum_{s=1}^{t-1} \frac{r(t, s)}{r(s, s)} X_s \quad (1)$$

and the corresponding mean square error,  $v_t = E [X_t - \hat{X}_{t/t-1}]^2$ , is given by

$$v_t = r(t, t) - \sum_{s=1}^{t-1} \frac{r^2(t, s)}{r(s, s)} \quad (2)$$

We can also give the predictions in terms of predication errors  $X_t - \hat{X}_{t/t-1}$  having covariances  $r^*(t, s)$  :

$$\hat{X}_{t/t-1} = \sum_{s=1}^{t-1} \frac{r^*(t, s)}{r^*(s, s)} (X_s - \hat{X}_{t/t-1}) \quad (3)$$

and the corresponding mean square error,  $V_t = E [X_t - \hat{X}_{t/t-1}]^2$ , is given by

$$V_t = r^*(t, t) - \sum_{s=1}^{t-1} \frac{r^{*2}(t, s)}{r^*(s, s)} \quad (4)$$

**Proof:** When the normality assumption is made, the proof follows from a theorem on normal correlations as in Abraham and Thavaneswaran (1990) or as

in Brockwell and Davis (1991) using projection arguments. For the nonnormal case, the result follows by applying the unconditional version of the optimality criterion given in Thavaneswaran and Thompson (1988).

## 4 Applications

### 4.1 ARMA(0,0,1) process or MA(1) process

Consider the recursive smoothing algorithm for an MA(1) process

$$X_t = \epsilon_t - \theta\epsilon_{t-1} \quad (5)$$

where  $\{\epsilon_t\}$  is a white noise process having mean zero and variance  $\sigma^2$ .

Then it can be easily shown that the one step ahead smoother of  $X_{t+1}$  and its mean square error are given by

$$\hat{X}_{t+1/t} = -\theta (X_t - \hat{X}_{t/t-1}) / V_t$$

and

$$V_t = \sigma_a^2 (1 + \theta^2) - \theta^2 / V_{t-1}, \quad \text{respectively}$$

where

$$V_0 = \sigma_a^2 (1 + \theta^2).$$

Based on  $n$  observations,

$$V_n = E (X_{n+1} - \hat{X}_{n+1/n})^2 = \sigma^2 (1 - \theta^{2n+4}) / (1 - \theta^{2n+2})$$

which converges to  $\sigma^2$  as  $n \rightarrow \infty$ , i.e. MSE of the smoother converges to  $\sigma^2$  for an MA(1) process. Now we look at the asymptotic properties of the EWMA statistic for the MA(1) process and compare its mean square error with that of the smoother.

Consider the EWMA,

$$\hat{Z}_t = \lambda X_t + (1 - \lambda) Z_{t-1}.$$

For the constant mean model with iid errors given in Section 1, Hunter (1986) showed that the mean of the EWMA,

$$E[Z_t] \rightarrow \mu$$

and the variance

$$\text{Var} [Z_t] \rightarrow \frac{\lambda}{2 - \lambda} \sigma^2.$$



or in terms of the discount coefficient  $\omega = 1 - \lambda$ ,

$$\text{Var}[Z_t] \cong \sigma^2 \frac{(1 - \omega)}{(1 + \omega)}.$$

For the MA(1) process (5) it can be shown that the asymptotic variance (MSE) of the EWMA statistic is given by

$$\text{Var}[Z_t] = \sigma^2 \frac{(1 - \omega)}{(1 - \omega)} [(1 + \theta^2) - 2\omega\theta].$$

Note that for small values of  $\omega$ , and  $\theta$  close to 1 this variance is as large as  $2\sigma^2$  and we lose about 50% of the efficiency by using EWMA. For different values of  $\lambda$  and  $\theta$ , Table 1 provides the MSE for the EWMA statistic when  $\sigma^2 = 1$ .

Table 1: MSE for the EWMA

$\lambda$	0	.25	.5	.75	1.0
$\theta$					
-1	1.85	2.38	3.10	4.12	5.7
-0.9	1.68	2.15	2.80	3.71	5.14
-0.8	1.51	1.94	2.52	3.34	4.62
-0.7	1.37	1.75	2.27	3.00	4.14
-0.6	1.25	1.59	2.04	2.69	3.69
-0.5	1.14	1.44	1.84	2.40	3.29
-0.4	1.05	1.32	1.66	2.15	2.92
-0.3	0.98	1.22	1.51	1.93	2.59
-0.2	0.93	1.14	1.38	1.74	2.30
-0.1	0.90	1.08	1.28	1.58	2.04
0	0.88	1.04	1.21	1.45	1.83
.1	0.88	1.03	1.16	1.34	1.65
.2	0.90	1.03	1.13	1.27	1.51
.3	0.94	1.06	1.13	1.23	1.41
.4	1.00	1.11	1.15	1.22	1.34
.5	1.08	1.18	1.20	1.23	1.31
.6	1.17	1.27	1.27	1.28	1.32
.7	1.28	1.39	1.37	1.36	1.36
.8	1.41	1.52	1.49	1.46	1.45
.9	1.56	1.68	1.64	1.59	1.57
1.0	1.73	1.85	1.81	1.76	1.72

It follows from Table 1 that the MSE of EWMA is larger than one when  $|\theta|$  is large. For instance when  $\theta = -.8$  and  $\lambda = .25$  the mean square error is 1.94.

#### 4.2 ARMA(1,1) Process

Now we consider an ARMA(1,1) process of the form

$$X_t - \phi X_{t-1} = \epsilon_t - \theta \epsilon_{t-1}.$$

Then it can be shown that the one step ahead smoother of  $X_{t+1}$  and its mean square error are given by

$$\hat{X}_{t+1/t} = \sigma\phi X_{t-1} - \theta (X_t - \hat{X}_{t/t-1}) / V_t$$

and

$$V_t = \sigma^2 (1 + \theta^2) - \sigma^2\theta^2 / V_{t-1}$$

respectively, where

$$V_0 = \sigma^2(1 + 2\theta\phi - \theta)/(1 - \phi^2).$$

As in the previous case it can be shown that  $V_t$  converges to  $\sigma^2$  while the mean square error of the EWMA depends on  $\phi$ ,  $\theta$  and  $\lambda$ . It is of interest to note here that for purely autoregressive processes the optimal linear smoother turns out to be the conditional expectation.

### 4.3 A numerical example

Montgomery and Mastrangelo (1991) used a control chart based on EWMA for a set of data containing 197 observations from a chemical process. This chart (not presented here to save space) identifies observations 4, 5, 9, 33, 44, 59, 65, 108, 148, 173, 174, 183, 192 and 193 as out-of-control points with the optimal value of the smoothing constant. Moreover, examination of the autocorrelation structure leads to an AR(2) model

$$\hat{X}_t = 54.5973 + .4260X_{t-1} + .2538X_{t-2}$$

for these data. The residuals from this model are uncorrelated.

Figure 1 presents the 197 observations from the same process, along with the optimal smoother with two sigma control limits. This chart identifies observations 6, 33, 44, 45, 61, 65, 108, 174, 192 and 193 as out-of-control points but does not detect any out-of-control conditions between observations 108 and 174. This is because different methods are used to estimate  $\sigma^2$  (one step ahead forecast error variance) on the control charts; this can result in observations being classified differently.

As discussed in the paper by Montgomery and Mastrangelo (1991) and Lowry and Montgomery (1995), the EWMA control chart is a very useful procedure that can be used when the data from a process are autocorrelated. However, it should be noted that the EWMA smoother

- (i) is just an approximation for any process other than an ARIMA (0,1,1) and does not have any optimality property if used for other models,
- (ii) has the difficulty of choosing the smoothing constant using an iterative procedure.

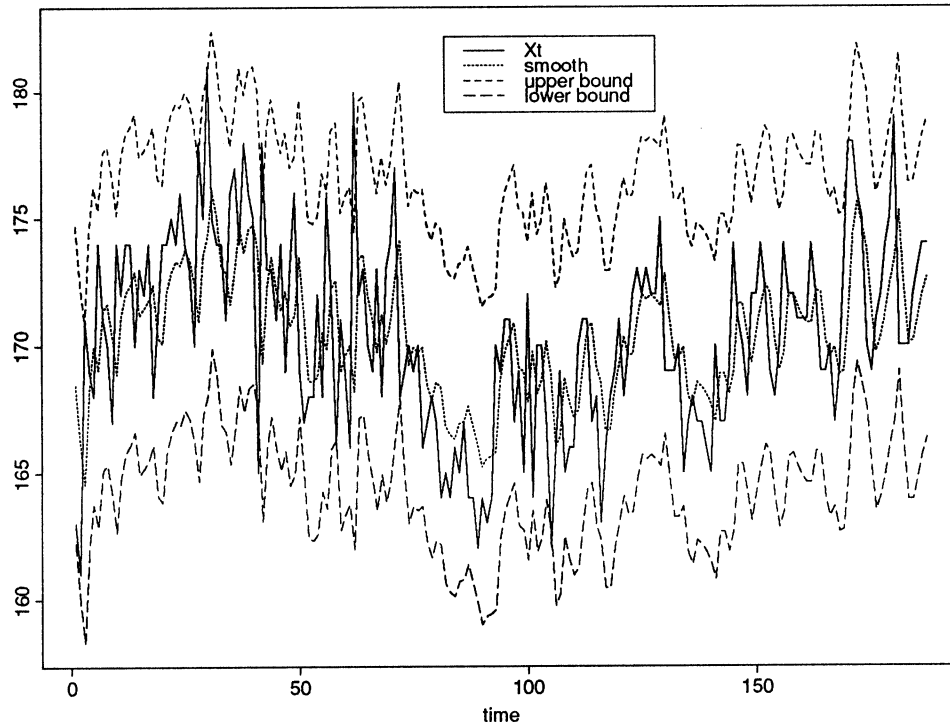


Figure 1. Chemical process observations with the optimal smoother

---

## 5 Summary and Concluding Remarks

Estimating function method is used to construct an optimal smoother for correlated data. As illustrated in Section 4, our examples show that the optimal smoother has advantages over the EWMA for a class of linear processes. For a given set of observations one can easily build a model by examining the autocorrelation structure and obtain the corresponding optimal smoother along with the control limits.

---

## Acknowledgement

B. Abraham was partially supported by a grant from the Natural Sciences and Engineering Research Council of Canada.

---

## References

1. Abraham, B., and Kartha, C.P. (1978). Geometric moving charts and Time Series; *ASQC Technical Conference Transactions*.
2. Abraham, B., and Kartha, C.P. (1979). Forecast stability and control charts. *ASQC Technical Conference Transactions*.
3. Abraham, B. and Ledolter, J. (1986). Forecast functions implied by autoregressive integrated moving average models and other related forecast procedures. *International Statistical Review*, **54(1)**, 5-66.
4. Abraham, B. and Thavaneswaran, A. (1991). A Nonlinear time series model and estimation of missing observations. *Annals of the Institute of Statistical Mathematics* **43**, 493-504.
5. Abraham, B., Thavaneswaran, A. and Peiris, S. (1997). On the prediction for a nonlinear time series models using estimating functions. (to appear in the *Institute of Mathematical Statistics Lecture Notes Series*).
6. Brockwell, P.J. and Davis, R.A. (1991). *Time Series: Theory and Methods*, Second Edition, New York, Springer Verlag.
7. Godambe, V.P. (1985). The foundations of finite sample estimation in Stochastic Processes. *Biometrika*, **72**, 419-429.
8. Heyde, C.C. (1987). On combining quasi-likelihood estimating functions. *Stochastic Processes and Their Applications*, **25**, 281-287.
9. Hunter, J.S. (1986). The Exponentially Weighted Moving Average. *Journal of Quality Technology*, **18**, 203-209.
10. Lowry, C.A. and Montgomery, C.D. (1995). A review of multivariate control charts. *IIE Transactions*, **27**, 800-810.
11. Montgomery, D.C and Mastrangelo, C.M. (1991). Some statistical process control methods for autocorrelated data. *Journal of Quality Technology*, **23**, 179-268.

12. Roberts, S.W. (1959). Control chart tests based on geometric moving averages. *Technometrics*, **1**, 239-251.
13. Thavaneswaran, A. and Abraham, B. (1988). Estimation of nonlinear time series models using Estimating Equations. *Journal of Time Series Analysis*, **9**, 99-108.
14. Thavaneswaran, A. and Thompson, M.E. (1988). A Criterion for filtering in semimartingale models. *Stochastic Processes and Their Applications*, **28**, 259-265.