Journal of Statistical Planning and Inference 141 (2011) 2556–2575

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Parameter estimation in first-order autoregressive model for statistical process monitoring in the presence of data autocorrelation

Thaung Lwin

CSIRO Mathematical and Information Sciences, Private Bag 10, Clayton, Rosebank MDC, Vic. 3169, Australia

ARTICLE INFO

Article history:
Received 29 August 2008
Received in revised form
24 January 2011
Accepted 8 February 2011
Available online 16 February 2011

Keywords:
Autocorrelative model
Kalman filter
ANOVA for time series
Variogram modelling
Empirical Bayes
EWMAST

ABSTRACT

The present paper considers an application of the first-order autoregressive (*AR(1)*) model to realizations, $(\lambda_1, \ldots, \lambda_n)$ of an unobservable variable, λ, representing a quality characteristic of a process monitored at a sequence of 'time' intervals in mineral processing or manufacturing production. At time $t_i (i=1, \ldots, n)$ a set of m_i observations $x_i^T = (x_{i1}, \ldots, x_{im_i})$ are made on the realization λ_i. The unknown realizations are observed subject to errors, implying an errors-in-variables model, for the observed sequence of data. The model, referred to as an autocorrelative model, has a reasonably wide range of applications in process monitoring with autocorrelated data.

Observed variograms of data in many production processes very often exhibit empirical forms such as a linear or an exponential model. Variation structure based on these models is consistent with an assumption of an autocorrelative model for the original sequence of observations and justifies a detailed study of efficient parameter estimation for the model. Application of such a model to process data requires both estimation of the unobservables, $(\lambda_1, \ldots, \lambda_n)$, in constructing one-step-ahead predictions and also estimation of all the underlying model parameters.

For given values of the underlying model parameters, estimation of the unobservables can be carried out most efficiently by Kalman-filter technique. Estimation of the model parameters can be handled by a number of techniques. Specific contributions of the present paper are: (i) a parametric approach comprising a comprehensive development of the full maximum likelihood technique for estimation of the model parameters in the presence of random effects, the number of which increases with the number of observations, (ii) a semi-parametric approach combining a direct or indirect fitting of a variogram combined with the method of moments, and minimum prediction error sum of squares techniques for estimation of model parameters and (iii) a modified procedure for developing an EWMA statistical control chart or process monitoring in the presence of data autocorrelation based on efficient parameter estimates together with its average-run-length properties.

1. Introduction

A first-order autoregressive process with errors in variables (*AR(1)–EIV*) model may be defined as follows:

Suppose that successive sets of observations are taken at 'time' points $\{t_1, \ldots, t_n\}$ to monitor a process. The set of observations at time point t_i is denoted by

$$\mathbf{x}_i = (x_{i1}, \ldots, x_{im_i})^T \quad i = 1, \ldots, n \tag{1.1}$$

E-mail address: thaung.lwin@csiro.au

0378-3758/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2011.02.005

In Eq. (1.1), x_{ij} is the jth observation on an underlying process characteristic λ_i measured with error ε_{ij}. The variation due to the measuring process at time t_i is represented by the model

$$x_{ij} = \lambda_i + \varepsilon_{ij} \tag{1.2}$$

where ε_{ij} is a random variable with mean zero and constant variance σ_ε^2.

Assume that the underlying quality characteristic λ_i is an unobservable random variable generated by *AR(1)* model

$$\lambda_i - \mu_0 = \rho(\lambda_{i-1} - \mu_0) + \eta_i \quad i = 1,\ldots,n \tag{1.3}$$

where η_i is a random variable with mean 0 and constant variance σ_η^2 representing the process variation. The initial state variable, λ_0, is assumed to have an 'a priori' distribution with mean μ_0 and variance σ_0^2.

The quantity ρ in Eq. (1.3) is an unknown constant which takes a value in the interval $0 \le \rho \le 1$. Two sub-cases have been considered in detail in the literature; (i) the case when $\rho = 1$, which is also known as *the steady model* (see e.g. Bather, 1963; Harrison, 1967), and (ii) the case when $0 < \rho < 1$ which is used as a model for statistical process control (SPC) in the presence of data autocorrelation (see e.g. Vasilopoulos and Stamboulis, 1978; Lu and Reynolds, 1999). The parameters $(\mu_0, \sigma_0^2, \rho, \sigma_\eta^2, \sigma_\varepsilon^2)$ are called the model parameters, as compared to the quantities λ_i's which are the 'unobservables' of the dynamic system defined by Eqs. (1.2) and (1.3).

The *AR(1)–EIV* model will henceforth be referred to as "*the autocorrelative model*" in line with the terminology of the SPC literature.

It should be noted that the lower-boundary case of $\rho = 0$ is the standard independent and identically distributed (IID) process for the λ_i's in Eq. (1.3). For an IID case with x_{ij} having a Poisson data distribution with mean λ_i at time t_i, Hoadley (1981) developed a quality measurement plan based on Bayesian approach.

On the other hand, the upper-boundary case of $\rho = 1$ is essentially the same as the model of discrete Brownian motion for the quality characteristic employed by Bather (1963).

The problem of quality monitoring is to obtain estimates of the series of "unobservables", λ_i's, free from noise. The "unobservables" are also known as "random effects" in the analysis of variance (ANOVA) literature. Due to a temporal nature of the series, in application, the desired estimation method is a sequential one with the estimate at the ith stage being updated at the $(i+1)$th stage when the new data $\boldsymbol{x}_{i+1}$ is observed. This immediately suggests an application of Kalman filter techniques for updating. It also suggests an estimation of the dynamic model parameters based on the data cumulated up to the ith period. Furthermore, it is required to estimate the uncertainty of estimates of the quality characteristic λ_i to construct more informative control limits for the observations. Using the set up of dynamic models, the uncertainty of the estimates of λ_i's can best be provided by extending the approach of Hoadley (1981) which employs, for the IID case, the posterior variance of λ_i for each given time t_i.

The model defined by Eqs. (1.2) and (1.3) is one special case of a dynamic linear model (DLM) which can be handled, in principle, by Kalman filter techniques (see West and Harrison, 1989). The latter techniques have been extensively used in forecasting problems. The main purpose here is not so much to provide forecasts or interval estimates of the quality characteristic for prediction, but to provide means of developing alternatives to control charting procedures with emphasis on: (i) using more efficient and/or robust estimates of the model parameters when parametric assumptions can or cannot be made and (ii) to relate the model parameter estimates to all underlying variation components. Details are given in Sections 2 and 3.

In applying the steady model or autocorrelative model to the observed data, two estimation problems need to be attended to: (a) estimation of the unobserved realizations λ_i's and (b) estimation of model parameters $\boldsymbol{\psi} = (\mu_0, \sigma_0^2, \sigma_\eta^2, \sigma_\varepsilon^2)$ for the steady model or $\boldsymbol{\psi} = (\mu_0, \sigma_0^2, \rho, \sigma_\eta^2, \sigma_\varepsilon^2)$ for the autocorrelative model.

Estimation of unobservables, λ_i's, can be readily achieved by using a special case of Kalman filter, which also provides a one-step-ahead predictor for the same quantities. On the other hand, estimation of the model parameters can be carried out in a number of ways as explained in Section 4.

Use of adaptive Kalman filtering in statistical process control, in regard to the estimation of the 'unobservables', is reasonably well established. A very early application appeared in Bather (1963), who employed the steady model to arrive at a special form of Kalman Filter procedure. Its use in the area was proposed formally by Phadke (1982), who considered a random walk model which is a slightly more general version of the steady model. In more recent works, the steady model was advocated again by Crowder (1986). The relationship between the steady model and its extension, linear growth model, and also the exponentially weighted moving average (EWMA) type models was explored by Kirkendall (1986).

The paper by Bather (1963) or Kirkendall (1986) did not explicitly discuss the estimation of the parameters for the steady model, but the papers by Phadke (1982), Crowder (1986) and Crowder and Eshleman (2001) addressed this problem using variants of maximum likelihood approach. Generally, the model parameter estimation methods may be grouped into two broad categories as follows.

1.1. Parametric approach

The errors ε_{ij} and the process fluctuations, η_i's, are assumed to be distributed as Gaussian random variables and the likelihood function of the data is employed to construct maximum likelihood estimators of the model parameters $\boldsymbol{\psi}$. A number of previous works dealt with variants of the full likelihood approach for the steady model. Phadke (1982)

and Crowder (1986) considered an estimation of the model parameters by the method of discounted likelihood and a modified maximum likelihood, respectively. Crowder and Eshleman (2001) did consider a special case of the full likelihood approach for the steady model, through a recasting of the problem in the framework of an "Empirical Bayes" approach, when each of the group sizes, m_i's, is unity. A similar development for the autocorrelative model with an unknown ρ and unequal m_i seems lacking in the literature. Also lacking is a development for the general case when the probability distributions F and G for the random components ε_{ij} and η_i are not necessarily Gaussian, though with known parametric forms.

The development in the present paper addresses all these issues. It unifies the results for both the steady model and the autocorrelative model using the full likelihood method.

1.2. Semi-parametric approach

The parametric assumptions may be relaxed in constructing estimates of model parameters that are reasonable in terms of bias and precision. In particular, the method of moments can be introduced if a simple and realistic empirical model of the variogram of the given data sequence can be constructed. For data that can be represented by the autocorrelative model, a good working model of the underlying theoretical variogram is found to be an exponential variogram. For the special case when the parameter ρ becomes unity, the underlying model becomes the steady model, for which the corresponding theoretical variogram is a linear variogram. In this case, the exponential variogram deduced for the autocorrelative model also approaches a linear variogram as a limiting form, when $\rho \to 1$. An early semi-parametric application of the variogram method for steady model was given by Harrison (1967). The variogram method has been extensively developed in particulate sampling theory (Gy, 1979) to study various sources of variation in sequences of data, indexed by time, mass, etc. The present paper unifies relevant aspects of the variogram method with those of forecasting techniques to provide a sound semi-parametric approach. This approach is simple and evidence has been found that reasonably robust estimates of the model parameters can be derived through this approach at a substantially reduced amount of computation. The methods can be applied effectively when only weak parametric assumptions can be made for the underlying probability distributions of the data and can be upgraded through iteration to derive the maximum likelihood estimators of the parametric approach when strong parametric distributional assumptions can be made for the data. A practical example is given in Section 5 where a data set from Montgomery and Mastrangelo (1991) is analysed to illustrate additional insights gained by using the present approach.

2. The autocorrelative model

We now consider in detail the autocorrelative model as given by Eqs. (1.2) and (1.3). First, the model is recast in the following form below to investigate its variational properties

$$x_{ij} = \mu_0 + \alpha_i + \varepsilon_{ij} \tag{2.1}$$

$$\alpha_i = \rho\alpha_{i-1} + \eta_i \quad i = 1,\ldots,n \tag{2.2}$$

where $\alpha_i (= \lambda_i - \mu_0)$'s are the unobservables with α_0 being a random variable having zero mean and variance σ_0^2, the random errors ε_{ij}'s and η_i's are independent random variables with zero means and variances σ_ε^2 and σ_η^2, respectively. The model parameters are $(\mu_0, \sigma_0^2, \rho, \sigma_\eta^2, \sigma_\varepsilon^2)$.

When $\rho = 1$, the above model reduces to the steady model. However, $\rho = 1$ is the boundary case for Eq. (3.2), where we assume $|\rho| < 1$ for stationarity of the distribution of the unobservables, α_i's. Thus the results for the steady model do not always follow by making a simple substitution of unity in place of ρ in the results of the present section. The steady model is treated in more detail in a separate paper (Lwin, 2005).

2.1. Theoretical variogram of the autocorrelative model

Define the mean of ith group as

$$\bar{x}_i = \sum_{j=1}^{m_i} x_{ij}/m_i; \quad i = 1,\ldots,n \tag{2.3}$$

The observed variogram of the group means is given by $\{V(1),\ldots,V(n-1)\}$ where

$$V(i) = \sum_{j=1}^{n-i} (\bar{x}_{j+i} - \bar{x}_j)^2/\{2(n-i)\}; \quad i = 1,\ldots,n \tag{2.4}$$

We need to evaluate the expected value of the ith component of the variogram to see which of the model parameters may be estimable by the quantities that constitute the observed variogram.

Consider the difference between $\bar{x}_{j+i}$ and $\bar{x}_j$, the two observations corresponding to the two time points t_{j+i} and t_j. We have from Eq. (2.1)

$$\bar{x}_{j+i} - \bar{x}_j = \bar{\varepsilon}_{j+i} - \bar{\varepsilon}_j + \alpha_{j+i} - \alpha_j$$

so that

$$(1/2)E(\bar{x}_{j+i}-\bar{x}_j)^2=(1/2)\sigma_\varepsilon^2(1/m_{j+i}+m_j)+V_\alpha(i) \tag{2.5}$$

where

$$V_\alpha(i)=(1/2)E(\alpha_{j+i}-\alpha_j)^2$$

Now

$$\alpha_{j+i}-\alpha_j=(\rho^{j+i}-\rho^j)\alpha_0+(\rho^{j+i-1}-\rho^{j-1})\eta_1+(\rho^{j+i-2}-\rho^{j-2})\eta_2+\cdots +(\rho^{i-1}-\rho^{j-i})\eta_j+\rho^{i-1}\eta_{j+1}+\rho^{i-2}\eta_{j+2}+\cdots+\eta_{j+i}$$

Hence

$$V_\alpha(i)=\{2(1-\rho^2)\}^{-1}\rho^{2j}(1-\rho^i)^2\sigma_0^2+[(1-\rho^i)^2(1-\rho^{2j})+(1-\rho^{2i})]\sigma_\eta^2 \tag{2.6}$$

When $j\to\infty$, we have

$$V_\alpha(i)=\gamma^2(1-\rho^i) \tag{2.7}$$

where

$$\gamma^2=(1-\rho^2)^{-1}\sigma_\eta^2 \tag{2.8}$$

This is the theoretical variogram of the autocorrelative process when it becomes stable. Hence when the process settles down, after start up, the underlying variogram of the data is an exponential variogram.

3. Estimation of the quality signal by Bayesian (Kalman filter) approach

Estimation of the unobservables α_i's by Kalman filter approach is a straightforward application of Kalman recursive algorithm (Kalman, 1960), for fixed values of the autocorrelative model parameters $(\mu_0,\sigma_0^2,\rho,\sigma_\eta^2,\sigma_\varepsilon^2)$. The resulting estimators also provide Bayes estimators of the quality signal, λ_i's if the joint distribution of α_i's is regarded as a 'prior' distribution.

The initial state λ_0 is also an unobservable and is assumed to have a prior distribution with mean μ_0 and variance σ_0^2. We consider the class of estimators for the unobservable λ_i that are linear in the mean x_{-i}.

Let $\mathbf{X}^{(i)}=\{\mathbf{x}_1,\cdots,\mathbf{x}_i\}$ denote the cumulative data of observations up to and including time t_i. The linear Bayes estimator of λ_i and its posterior variance given the data $\mathbf{X}^{(i)}$ are obtained once the corresponding results for α_i's are constructed. Let the linear Bayes estimator of α_i and its posterior variance based on the data, $\mathbf{X}^{(i)}$, up to time t_i be

$$\hat{\alpha}_i^*=E(\alpha_i|\mathbf{X}^{(i)}) \tag{3.1}$$

$$\hat{Q}_i^*=Var(\alpha_i|\mathbf{X}^{(i)}) \tag{3.2}$$

Then we have the updating sequences for Eqs. (3.1) and (3.2) as follows, see e.g. West and Harrison (1989, pp. 107–111)

$$\hat{\alpha}_i^*=\rho\hat{\alpha}_{i-1}^*+w_i(\bar{x}_i-\mu_0-\rho\hat{\alpha}_{i-1}^*) \tag{3.3}$$

where

$$d_i^2=\sigma_\varepsilon^2/m_i,\quad T_i=\rho^2\hat{Q}_{i-1}^*+\sigma_\eta^2,\quad w_i=T_i(d_i^2+T_i)^{-1}$$

and the quantities $\hat{Q}_i^*$ are calculated from the recursive equation

$$\hat{Q}_i^*=(\rho^2\hat{Q}_{i-1}^*+\sigma_\eta^2)d_i^2/(\rho^2\hat{Q}_{i-1}^*+\sigma_\eta^2+d_i^2) \tag{3.4}$$

The expression (3.3) can be used to obtain the updating sequence for $\hat{\lambda}_i^*$ as

$$\hat{\lambda}_i^*=\mu_0+\hat{\alpha}_i^* \tag{3.5}$$

The above iterative procedure can be started with $\hat{\alpha}_0^*=0$ and $\hat{Q}_0^*=\sigma_0^2$. The need to estimate the parameter σ_0^2 is obviated by using the limiting form of $\hat{Q}_i^*$ as $i\to\infty$. It can be readily shown that the required limiting form is given by

$$Q_i^*=(\sigma_\eta^2/2)H_i \tag{3.6}$$

where

$$H_i=\rho^{-2}[\{R_i(1-\rho^2)+1\}^2+4\rho^2R_i]^{1/2}-\rho^{-2}\{R_i(1-\rho^2)+1\} \tag{3.7}$$

with

$$R_i=d_i^2/\sigma_\eta^2 \tag{3.8}$$

If only the asymptotic expression of $\hat{Q}_i^*$ is used, the weights in Eq. (3.3) are given by

$$w_i = (2+\rho^2 H_i)/(2+\rho^2 H_i+2R_i) \tag{3.9}$$

When $\rho=0$, the quantities, w_i, $\hat{\lambda}_i^*$, and $\hat{Q}_i^*$ become

$$w_i = (1+R_i)^{-1}, \quad \hat{\lambda}_i^* = (1-w_i)\mu_0 + w_i\bar{x}_i, \quad Q_i^* = \sigma_\eta^2 d_i^2/(\sigma_\eta^2+d_i^2) \tag{3.10}$$

as is expected for an IID sequence of λ_i's.

When $\rho=1$, the corresponding expressions are given by

$$\hat{\lambda}_i^* = (1-w_i)\hat{\lambda}_{i-1}^* + w_i\bar{x}_i \tag{3.11}$$

where

$$w_i = \{1+(1+4R_i)^{1/2}\}/\{1+(1+4R_i)^{1/2}+2R_i\} \tag{3.12}$$

and the posterior variance is

$$Q_i^* = (\sigma_\eta^2/2)\{(1+4R_i)^{1/2}-1\} \tag{3.13}$$

in agreement with the expressions for the steady model as given by Bather (1963) and Crowder (1986).

A remark now is in order. The estimator for the ith unknown λ_i, given by the expression (3.5), is of the same form as the geometric moving average estimator of Roberts (1959) in that it is a weighted average of the estimator of the previous stage estimate, $\hat{\lambda}_{i-1}^*$, and the current data $\bar{x}_i$. Although the information on the time order of the data sequence was explicitly considered, Roberts' development assumed an IID sequence, without including the random effects in the model.

The present development assumes a non-iid probabilistic structure for the data and uses a Bayes formulation to construct the estimator. In the process, the weights are determined as a function of the variational parameters of the probabilistic model. Thus the 'weight' is not a fixed value such as 2/5 recommended by Roberts (1959) for the case $m_i=m$, a constant, as a choice of wide appeal.

4. Estimation of the autocorrelative model parameters

4.1. Semi-parametric approach

This approach does not make use of any assumption on parametric form of the distribution of the random quantities $\mathbf{X}^{(n)} = \{\mathbf{x}_1,\ldots,\mathbf{x}_n\}$. It consists of three components: (a) variogram method, (b) method of moments and (c) minimum prediction sum of squares (*PRESS*) method combined. The techniques are simple and readily implementable. Estimates obtained by this approach can also be used as initial values in starting the maximum likelihood method if a parametric approach is to be adopted. Although less efficient than the corresponding ML estimates, the semi-parametric solution can be quite close to the parametric solution if the number of groups, n, is large.

4.1.1. Variogram method

The expected value of the observed variogram is given by (2.10) which can be rewritten as

$$EV(i) = \sigma_\varepsilon^2 A_i + \gamma^2\{1-\exp(-vi)\} \tag{4.1}$$

where

$$v = -\ln(\rho) \tag{4.2}$$

For the autocorrelated data sequence, information on the parameters $(\rho,\sigma_\eta^2,\sigma_\varepsilon^2)$ can be obtained from the variogram and the set of statistics $\{V(1),\ldots,V(n-1)\}$ can be employed to estimate the parameters $(v,\gamma^2,\sigma_\varepsilon^2)$ when n is a reasonably large number. The usual analysis based on the variogram method considers a fitting of a theoretical model to an observed variogram up to $N=[(n-1)/2]$ lags of the data.

Suppose the observed variogram is modelled by the theoretical variogram (4.1) in the region of the first N lags. Then the model can be fitted to the data by using a weighted least squares approach (see Cressie, 1991). This approach minimizes the quantity

$$\sum_{i=1}^{n-i}[V(i)/E\{V(i)\}-1]^2 \tag{4.3}$$

with respect to $(v,\gamma^2,\sigma_\varepsilon^2)$. The estimates of γ^2 and v and Eqs. (2.8) and (4.2) then provide corresponding estimates of (ρ,σ_η^2).

It should be noted here that σ_ε^2 may be estimated independently from the replicate data for the case when $m_i>1$ for all values of i. Based on the within-group mean square S_W^2, an independent estimate is obtained as

$$\hat{\sigma}_\varepsilon^2 = S_W^2 = \sum_{i=1}^{n}\sum_{j=1}^{m_i}(x_{ij}-\bar{x}_i)^2/\left(\sum_{i=1}^{n} m_i - n\right) \tag{4.4}$$

In such a case, the sum of squares in Eq. (4.3) can be minimized with respect to (ρ,σ_η^2) only, while the value of σ_ε^2 is fixed at $\hat{\sigma}_\varepsilon^2$.

There still remains the problem of estimating μ_0. The variogram on its own does not contain information on this remaining parameter. However prior mean μ_0 is usually available as a prior guess, or a standard target value to be aimed at or the grand mean of the observations up to a start up period. With this modification, the observed series and the variogram provide enough information for the estimation of model parameters.

4.1.2. The method of moments and PRESS method

The method assumes that an independent estimate of σ_ε^2, such as Eq. (4.4), is available. It first assumes a value of ρ, and constructs an estimate of σ_η^2 by using a weighted between-group variance defined as follows:

$$S_B^2 = \sum_{i=1}^{n} p_i(\bar{x}_i - \bar{\bar{x}})^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} p_i p_j(\bar{x}_i - \bar{x}_j)^2 \tag{4.5}$$

where $p_i = m_i / \sum_{j=1}^{n} m_j$ is the proportion of ith the group size to the total sample size of all n groups and

$$\bar{\bar{x}} = \sum_{i=1}^{n} p_i \bar{x}_i \tag{4.6}$$

The expected value of the weighted between-group variance can be evaluated in terms of the parameters of the theoretical variogram $E\{V(i)\}$ of Eq. (4.1).

Using the expressions (4.1) and (4.5), the expected value of S_B^2 can be shown to be the quantity

$$E(S_B^2) = K\sigma_\varepsilon^2 + \sigma_\eta^2\tau(\rho) \tag{4.7}$$

where K is defined as

$$K = \sum_{i=1}^{n}\sum_{j=i}^{n} p_i p_j(1/m_i + 1/m_j) \tag{4.8}$$

and

$$\tau(\rho) = (1-\rho^2)^{-1}\sum_{i=1}^{n}\sum_{j\neq i}^{n} p_i p_j(1-\rho^{|t_i - t_j|}) \tag{4.9}$$

When m_i's are equal and t_i's are equidistant with $t_i = i$, we take $p_i = 1/n$ and arrive at the result

$$E(S_B^2) = m^{-1}(1-1/n)\sigma_\varepsilon^2 + \sigma_\eta^2\tau(\rho) \tag{4.10}$$

where $\tau(\rho)$ is simplified to

$$\tau(\rho) = (1-\rho^2)^{-1}[(1-1/n) - 2n^{-2}\{n(\rho-\rho^n)/(1-\rho) - (\rho-\rho^n)/(1-\rho^2) - (n-1)\rho^n/(1-\rho)\}] \tag{4.11}$$

For a given value of ρ, an estimate of the long-term variance is given by

$$\hat{\sigma}_\eta^2 = (S_B^2 - K\hat{\sigma}_\varepsilon^2)/\tau(\rho) \tag{4.12}$$

where $\hat{\sigma}_\varepsilon^2$ is an independent estimate of σ_ε^2 as given in Eq. (4.4).

When $\rho = 0$, the expression in Eq. (4.10) provides the usual estimate of the 'between-group' variance in ANOVA. An analogous expression for the case of $\rho = 1$, however, cannot be deduced by substitution in Eq. (4.10). It can however be deduced from Eq. (4.10) by using L'Hospital's rule as

$$\tau(\rho = 1) = (1/2)\sum_{i=1}^{n}\sum_{j\neq i}^{n} p_i p_j |t_i - t_j|) \tag{4.13}$$

Suppose that $\hat{\lambda}_i^+$ is a one-step-ahead prediction for λ_i, defined by the mean of posterior predictive distribution of a new observation at the $(i-1)$th stage. Then from Eqs. (4.5) and (4.6)

$$\hat{\lambda}_i^+ = \mu_0 + \hat{\alpha}_{i-1}^* \tag{4.14}$$

and variance

$$\hat{Q}_i^+ = \hat{Q}_{i-1}^* \tag{4.15}$$

Thus based on the predictive distribution of the ith stage observation a weighted deviation sum of squares criterion can be constructed. This criterion is to minimize the quantity termed "Prediction Error Sum of Squares" (*PRESS*), defined as

$$PRESS = \sum_{i=1}^{n}(\bar{x}_i - \hat{\lambda}_i^+)^2/\hat{Q}_i^+ \tag{4.16}$$

Expressions for $\hat{\lambda}_i^+$ and $\hat{Q}_i^+$ in Eqs. (4.14) and (4.15) can be reparameterised as a function of ρ only by replacing μ_0 by $\bar{\bar{x}}$, setting σ_ε^2 fixed at $\hat{\sigma}_\varepsilon^2$, and replacing σ_η^2 by the expression (4.12). The quantity *PRESS* can then be minimized as a function of ρ to obtain an "optimal" estimate $\hat{\rho}$.

The above discussion leads us to the following iterative procedure for estimating the model parameters.

Step 1: Obtain an estimate of σ_ε^2 either from replicate data or an independent experiment.
Step 2: Choose an initial estimate of ρ by investigating the variogram.
Step 3: Calculate an estimate of σ_η^2 using Eq. (4.12) with estimates of ρ and σ_ε^2
Step 4: Calculate the one-step-ahead predictor $\hat{\lambda}_i^+$ using the estimates of the parameters in Steps 1–3.
Step 5: Update the estimate $\hat{\rho}$ by minimizing *PRESS* with respect to ρ.

The Eqs. (4.4), (4.7) and (4.9) (or (4.13)) provide the ANOVA Table for the data generated according to the autocorrelative (or steady) model as noted in Lwin (2005).

4.2. Parametric approach

4.2.1. The maximum likelihood method

The full maximum likelihood applicable for general parametric distributions F and G, with probability density functions (pdfs) f and g, is described in detail in Appendix A. In the present section, a simplified version is given for the case when a Gaussian assumption is made for both pdfs, f and g, and also when $m_i=m$ for all group sizes.

The likelihood function can be factored into two parts. The first part, L_1, depends only on the parameter σ_ε^2, while the second part, L_2, depends on all four parameters $(\mu_0,\rho,\sigma_\eta^2,\sigma_\varepsilon^2)$. Information on σ_ε^2 is contained mainly in L_1, although L_2 provides some information on it through the ratio

$$\phi = \gamma^2/\sigma_\varepsilon^2 \tag{4.17}$$

Thus a reasonable approximation to the full likelihood approach is obtained by maximising the likelihood function in two stages. In the first stage, L_1, which is essentially the marginal distribution of the within-group mean square, S_W^2, is maximized with respect to σ_ε^2 only, leading to Eq. (4.4).

Next we note that, L_2 can be replaced by the joint probability density function of $(\bar{x}_1,\ldots,\bar{x}_n)$ derived from the integrated multivariate pdf

$$f(\bar{x}_1,\cdots,\bar{x}_n|\boldsymbol{\psi}) = \int \{\Pi_{i=1}^n f(\bar{x}_i|\lambda_i,\boldsymbol{\kappa}_i)\} g(\boldsymbol{\lambda}|\boldsymbol{\xi}) d\boldsymbol{\lambda} \tag{4.18}$$

For the multivariate Gaussian (MVG) case, this joint pdf is obtained from the distribution

$$\bar{\mathbf{x}} \stackrel{d}{=} MVG[\mu_0\mathbf{1},\sigma_\varepsilon^2 P(\phi,\rho)] \tag{4.19}$$

where $\boldsymbol{D}$ is an $n\times n$ diagonal matrix with the ith diagonal element

$$d_{ii} = 1/m_i \tag{4.20}$$

$$P(\phi,\rho) = \mathbf{D} + \phi\,\boldsymbol{\Gamma} \tag{4.21}$$

In the expression for $P(\phi,\rho)$.

In the second stage, the likelihood function of the group means $(\bar{x}_1,\cdots,\bar{x}_n)$ is maximized, for fixed σ_ε^2, with respect to the remaining three parameters.

Thus the approximate likelihood approach given above leads to the following procedure:

Step 1: Obtain an estimate of σ_ε^2 as in Eq. (4.4) using the within group mean square S_W^2.
Step 2: Obtain an initial estimates of ρ and $\phi=\sigma_\eta^2/\{\sigma_\varepsilon^2(1-\rho^2)\}$. These can be based on the estimate derived from the semi-parametric approach.
Step 3: Compute the inverse of the matrix $P(\phi,\rho)$ using an explicit expression for the case when $m_i=m$ (see the Appendix) or by application of Trench (1964) algorithm for the case of unequal group sizes.
Step 4: Compute the estimates of μ_0 by using the relation

$$\hat{\mu_0} = [\mathbf{1}^T\{P(\phi,\rho)\}^{-1}\mathbf{1}]^{-1}[\mathbf{1}^T\{P(\phi,\rho)\}^{-1}\bar{\mathbf{x}}] \tag{4.22}$$

Step 5: Minimize the quantity

$$C = \ln\{|P(\phi,\rho)|\} + \sigma_\varepsilon^{-2}\{(\bar{\mathbf{x}}-\mu_0\mathbf{1})^T\{P(\phi,\rho)\}^{-1}(\bar{\mathbf{x}}-\mu_0\mathbf{1})\} \tag{4.23}$$

with respect to ϕ and ρ. The minimum achieved provides the required approximate maximum likelihood estimates of ϕ and ρ. The approximate ML estimate of σ_η^2 is calculated by using the equation

$$\sigma_\eta^2 = \phi\{\sigma_\varepsilon^2(1-\rho^2)\} \tag{4.24}$$

Step 6: Using new estimates of ϕ and ρ recompute the estimate of μ_0 as in ***Step 2*** and also compute the residual mean square

$$\sigma^2_{RES} = n^{-1}\{(\bar{\mathbf{x}}-\mu_0\mathbf{1})^T\{P(\phi,\rho)\}^{-1}(\bar{\mathbf{x}}-\mu_0\mathbf{1})\} \tag{4.25}$$

The above procedure is also an empirical Bayes (EB) type procedure in the sense that the unknown model parameters $(\mu_0,\rho,\sigma^2_\eta)$ are estimated from the likelihood function derived from the marginal joint distribution of the observed means $(\bar{x}_1,\ldots,\bar{x}_n)$, for a fixed value of σ^2_ε. All the unobservables $(\lambda_1,\ldots,\lambda_n)$ are estimated by the Bayes estimator, namely, the posterior mean (3.5).

The above approximate likelihood procedure is an adaptation, to the autocorrelative model, of the procedure proposed for the steady model by Crowder and Eshleman (2001). The procedure can be modified to obtain a two-stage weighted least squares method by replacing Eq. (4.23) in Step 5 by the residual mean square Eq. (4.25). Gaussian assumption can then be relaxed. The above algorithm based on ***Step1–Step 5*** can be employed for the steady model, provided Eq. (4.26) is replaced by the expression

$$P(\phi,\rho) = \mathbf{D}+\phi_0\mathbf{U}$$

where $\mathbf{U}$ is an $n\times n$ matrix with the (i,j)th element

$$u_{ij} = \min(i,j) \tag{4.26}$$

This has been shown by Crowder and Eshleman (2001).

5. Application to process control

5.1. EWMAST for AR(1) model

The process model (2.1) and (2.2) has been considered in the context of statistical process control by Lu and Reynolds (1999), who suggested two procedures for constructing control limits using EWMA chart based on (i) original observations and (ii) residuals. Alternatively, Zhang (1998) developed a method of constructing a statistical control chart for stationary process data (EWMAST), and the present model (2.1) and (2.2) is covered by Zhang's method for the case when $|\rho|<1$. To consider both approaches simultaneously, it is assumed that the group sizes are equal: $m_i=m$; $i=1,\ldots,n$.

For a pre-assigned target value τ for the process, the EWMA chart advocated by Lu and Reynolds (1999) employs the control statistic

$$Y_i = (1-w)Y_i + w\bar{x}_{i-1} \tag{5.1}$$

together with the control limits

$$\tau \pm c\{w/(2-w)\}^{1/2}(\hat{\sigma^2_\varepsilon}/m+\hat{\gamma^2})^{1/2} \tag{5.2}$$

Where w is a smoothing constant and the quantities (w,c) are determined to obtain the desired behaviour of steady state Average Run Length (ARL) for pre-assigned values of the parameters $(\rho,\zeta=\gamma^2/(\gamma^2+\sigma^2_\varepsilon/m))$. Lu and Reynolds (1999) provide extensive tables of the ARL as a function of the four parameters (ρ,ζ,w,c) to enable a robust design of the EWMA chart. The methodology of Section 4 provides efficient estimates of not only the parameters (ρ,ζ), but also that of w (see Eq. (3.9)). Thus it gives an option of data-based estimate of the weight function.

On the other hand, Zhang's EWMAST control limits are given as

$$\tau \pm L[\{w/(2-w)\}C(0)]^{1/2}\left[1+2\sum\nolimits_{k=1}^{M}\{C(k)/C(0)\}(1-w)^k\{1-(1-w)^{2(M-k)}\}\right]^{1/2} \tag{5.3}$$

where L is a Gaussian distribution factor for an assigned probability level (taken as 3 for the specific Tables of the ARL provided) and M is a reasonably large number and the quantities $C(k)$ are the process covariances and can be expressed for stationary case of the model (2.1) and (2.2) as

$$C(|i-j|) = Cov(\bar{x}_i,\bar{x_j}) = \begin{cases} \sigma^2_\varepsilon/m+\gamma^2 \text{ for } |i-j|=0 \\ \gamma^2\rho^k \text{ for } |i-j|=k \end{cases} \tag{5.4}$$

By taking the limiting expression of (5.3) as $M\to\infty$ and applying the stationary covariances in Eq. (5.4), Eq. (5.3) becomes

$$\tau \pm L[\{w/(2-w)\}(\sigma^2_\varepsilon/m+\gamma^2)]^{1/2}[1+2\{\gamma^2/(\sigma^2_\varepsilon/m+\gamma^2)\}\{\rho(1-w)\}\{1-\rho(1-w)\}^{-1}]^{1/2} \tag{5.5}$$

In the following the approach based on the control limits in Eq. (5.2) will be referred to as Lu–Reynolds approach, while that based on Eq. (5.3) will be referred to as Zhang approach. The two approaches, when specialised to the AR(1) model of (5.4), are practically the same for situations when ρ is small and/or w is large and the quantity c in Eq. (5.2) is set to be the same as the quantity L in Eq. (5.3). Hence we focus our discussion on the application of EWMAST limits in Eq. (5.5)

combining with the estimation techniques of Section 4.2. This modified approach will be termed a "Parameterised EWMAST". It has a number of added advantages. First, more efficient parameter estimates based on the data are available, in contrast to Zhang's original (non-parametric) estimates based on sample auto-covariances in Eq. (5.3). Second, it enables one to conduct a sensitivity analysis by enabling the construction of a robust EWMAST design by assessing the effect of varying the parameter estimates, within the limits of the corresponding uncertainties, on the performance of the control limits.

5.2. A modified monitoring procedure for AR(1) model

The "Parameterised EWMAST" is proposed here as comprising of not only the application EWMAST in Eq. (5.5) using model based parameter estimates, but also supplementing it with two additional sets of control limits for individual observations based on the partitioning of the total variation into three components: (i) short-term (replication) variation, (ii) medium-term (lag-one) variation and (iii) overall variation. The control limits in Eq. (5.5) represent the overall variation. Those representing the first two types of variation are constructed by using the formulae:

$$\text{Control limits (Short-term)} = \tau \pm 3[\{w/(2-w)\}(\sigma_{\varepsilon}^2/m)]^{1/2} \tag{5.6}$$

$$\text{Control limits (Medium-term)} = \tau \pm 3[\{w/(2-w)\}(\sigma_{PRED}^2)]^{1/2} \tag{5.7}$$

In Eq. (5.7), the quantity σ_{PRED}^2 is the prediction variance (see e.g. Davis and Jones (1968) for its definition and computation). For the *AR(1)–EIV* model it can be shown that

$$\sigma_{PRED}^2 = [\{\sigma_{\eta}^2+\sigma_{\varepsilon}^2(1+\rho^2)\}/2](1+[\{\sigma_{\eta}^2+\sigma_{\varepsilon}^2(1-\rho^2)\}/\{\sigma_{\eta}^2+\sigma_{\varepsilon}^2(1+\rho^2)\}]^{1/2}) \tag{5.8}$$

The short-term control limits employ the replication variance only. The medium-term control limits employ the magnitude of variation employed for one-step-ahead prediction limits for individual observations, in constructing the EWMAST limits. The first and second types of limits are useful as early warnings especially when a systematic trend, signalling a shift in the basic process mean, is suspected but the trend signal is not strong enough to be detected by the overall EWMAST control limits of Eq. (5.5).

5.3. Performance of the modified monitoring procedure

We now discuss an assessment of the performance of proposed monitoring procedure, "Parameterised EWMAST". An accepted property for assessing the performance of control limits such as Eq. (5.5) is to compute the "Average Run Length (ARL)" of the control procedure. It is the average number of runs (or time steps) required till a signal is given that a shift in the underlying level of the process mean has occurred. Such a signal is triggered at observation t if Y_t lies outside the limits in Eq. (5.5). The ARL value is evaluated for given values of the process parameters as well as for various choices of the control limit parameters (L,w) in Eq. (5.5). The behaviour of the change in ARL is then studied with respect to the change in the underlying process parameters as well as the control limit parameters in designing an EWMAST scheme.

Tables of ARL for the control limits in (5.2) and (5.5) have been constructed by Lu and Reynolds (1999) and Zhang (1998), respectively. Their results are directly applicable to assessment of the limits (5.5). With the process mean and process variance parameters standardised, the ARL of Eq. (5.5) depends on the possible true level of ρ, the magnitude δ of the shift in the mean and also the choice of w.

The "Parameterised EWMAST" approach based on Eqs. (5.5) and (5.7) is demonstrated for the data of the motivating example in Section 6. For this example, it will be seen that the ARL tables of Zhang (1998, particularly Table 4) need to be complemented by an extension corresponding to the choice of $w$in the vicinity of 0.9 in order to assess the performance of both Eqs. (5.6) and (5.8). The ARL values for the range of parameter choices of ρ=(0.25,0.35,0.45,0.55,0.65,0.75,0.85), w=(0.2,0.3,0.4,0.9) and δ=(0.0,0.5,1.0,2.0,3.0) are computed by using an approximate procedure assuming the steady-state conditions and are described in Appendix C, Table 1. The design implications of the results in Table 1 (Appendix C) are in agreement with the conclusions arrived at by the studies of Lu and Reynolds (1999) and Zhang (1998).

In general, the choice of a low value for w (say below 0.4) results in EWMAST which detects small shifts in the mean better than is the case for the choice of higher value for w, especially when the level of autocorrelation is low. On the other hand, for a choice of a high value w, though generally effective in detecting larger shifts, the performance of EWMAST is enhanced more when the level of autocorrelation is high. In short, EWMAST charts are to be designed according to the behaviour of ARL which is a function of the autocorrelation parameter and the nature of the drift in the mean that is aimed to be detected. This means that the parameter w is chosen from the ARL tables (for design) once ρ is determined and its uncertainty is assessed. In this context, the full advantage of the "Parameterised EWMAST" technique becomes more obvious. From the construction of the optimal Bayes predictor in Eq. (3.5), estimates of both ρ and w as well as their uncertainties are already suggested by the data to enable the process operator to make a more informed decision on the design of the EWMAST.

Next, we discuss the performance of the supplementary control limits (5.6) and (5.7).

The ARL behaviour of the short-term control limits (5.6) as well as the medium term control limits is completely described by the work of Crowder (1987, 1989), who considered for IID process model. Hence, one can conclude that the

small shifts in the trend above short-term variation will be detected by these limits faster than the overall EWMAST. On the other hand, the ARL behaviour of Eq. (5.7) can be assessed, approximately, by the ARL Tables provided in Table1of Appendix C, corresponding to the level of autocorrelation given by the relation

$$\rho = \rho^* = \{1-\sigma_\eta^2/(\sigma_{PRED}^2-\sigma_\varepsilon^2)\}^{1/2}$$

Comparison of Eqs. (5.2) and (5.7) shows that, for $c=3$, using Eq. (5.7) instead of Eq. (5.2) is the same as acting as though the value of ρ is at a different value $\rho=\rho^*$. Thus, "Parameterised EWMAST" can be designed to detect the early signal of a systematic trend by using the medium-term control limits as a supplement to provide a first warning signal, with the overall limits (5.5) providing the final warning for action.

6. An example

An example is extracted from Montgomery and Mastrangelo (1991), by employing a data set produced by reading the graph of Example 1 from this paper.

The observed variogram is calculated and plotted in Fig. 1. It is seen that the variogram is locally of an exponential type up to lag 30, although the overall shape of the variogram is linear. Montgomery and Mastrangelo (1991) fitted the model by a first-order autoregressive model of the form

$$x_i = 13.04 + 0.847\, x_{i-1} \tag{6.1}$$

The residual mean square error of the above fitted model can be calculated as

$$\sigma_{RES}^2 = 13.7867 \tag{6.2}$$

After inspecting the residuals from this model, Montgomery and Mastrangelo (1991) suggested that this model is an adequate fit to the data.

An implication of this result is that in the frame work of a DLM, an appropriate model for the data should be the autocorrelative model of Section 4. Consequently an exponential variogram model should be a better representation of the underlying variogram of the data than a linear variogram model.

In analysing the data two approaches are adopted. First, an application of the semi-parametric approach is made with respective applications of the variogram method and the *PRESS* method. Second, an application of the parametric approach is made using the likelihood method based on MVG distributional assumptions.

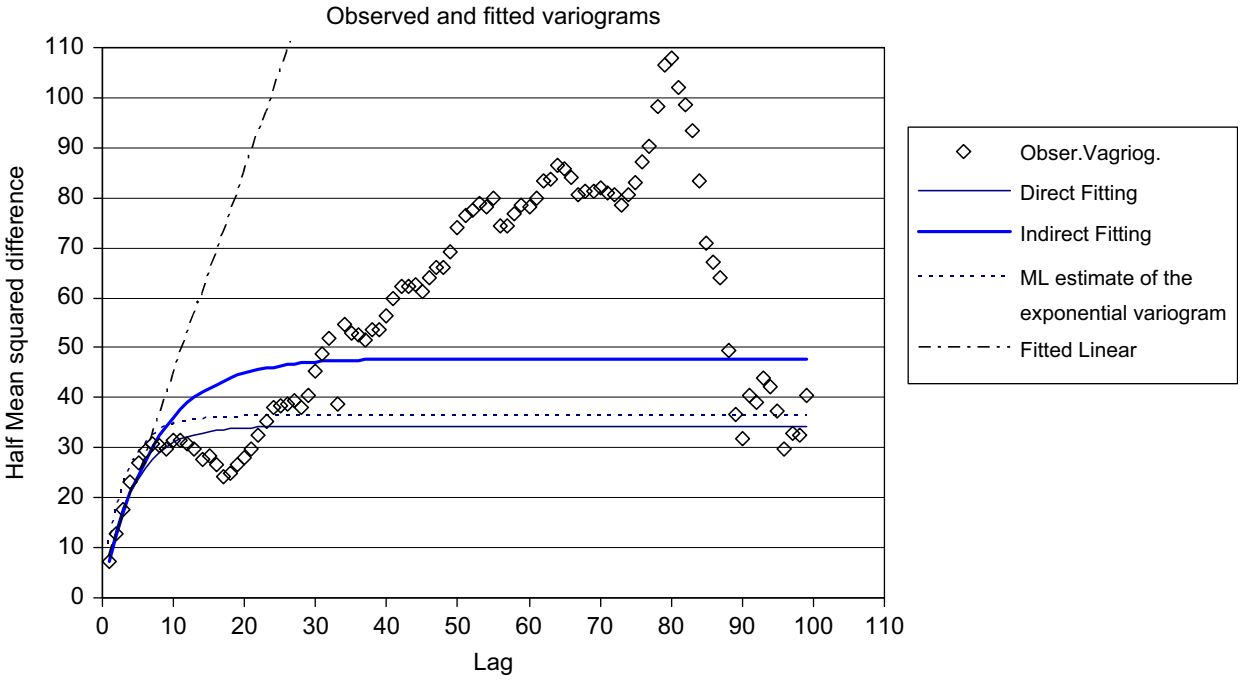

Fig. 1. An empirical modelling of the observed variogram.

6.1. Semi-parametric approach

6.1.1. Variogram method

The observed variogram exhibits an exponential behaviour for the first 30 lags of the data. A small cyclic component of variation is detected by the variogram but is not as substantial as the dominant behaviour of the first section. The weighted least square criterion (4.3) was used with maximum lag 30 to fit the variogram model (4.1), with A_i=1, to the observed variogram. The parameter estimates are obtained as

$$\sigma_\varepsilon^2 = 1.1508, \quad \gamma^2 = 33.1235, \quad v = 0.2334 \tag{6.3}$$

The above set of values for (γ^2, v) will lead to estimates of the model parameters as

$$\sigma_\varepsilon^2 = 1.1508, \quad \sigma_\eta^2 = 12.3553, \quad \rho = 0.7918 \tag{6.4}$$

The estimate of ρ suggests a significant departure of the data from the steady model assumption.

The set of parameter estimates in Eq. (6.4) for the autocorrelative model is used to obtain (i) the exact weights in Eq. (3.3) and (ii) the asymptotic weights in Eq. (3.9). The exact weights converge to the asymptotic weight (w=0.9) after a few time steps. In Fig. 1, the asymptotic weights are used to construct estimates of λ_i's. The results are depicted as one-step-ahead predictions, to demonstrate the adequacy of the model.

Next we assume that an independent estimate of σ_ε^2 is the same as the estimate obtained by direct fitting of the exponential as above, estimate of σ_η^2, by using the relationship (4.10) and that of ρ by minimising (4.3), which is computed up to 30 lags. The estimates are obtained as

$$\sigma_\varepsilon^2 = 1.1508, \quad \sigma_\eta^2 = 11.7424, \quad \rho = 0.8714 \tag{6.5}$$

A comparison of Eqs. (6.4) and (6.5) shows that the direct and indirect methods of fitting the exponential variogram provide similar results. The indirect method attempts to capture the variation beyond the 30 lags, used for the direct method to restrict it to the region of dominant exponential behaviour. In Figs. 1 and 2, the estimates of Eqs. (6.4) and (6.5) are used to construct the fitted variograms.

6.1.2. The PRESS method

The method of minimum prediction error sum of squares (Section 5.1) is used to estimate σ_η^2 and ρ for fixed value of σ_ε^2 as obtained in Eq. (6.4). The resulting estimates are

$$\sigma_\varepsilon^2 = 1.1508, \quad \sigma_\eta^2 = 8.86, \quad \rho = 0.9253 \tag{6.6}$$

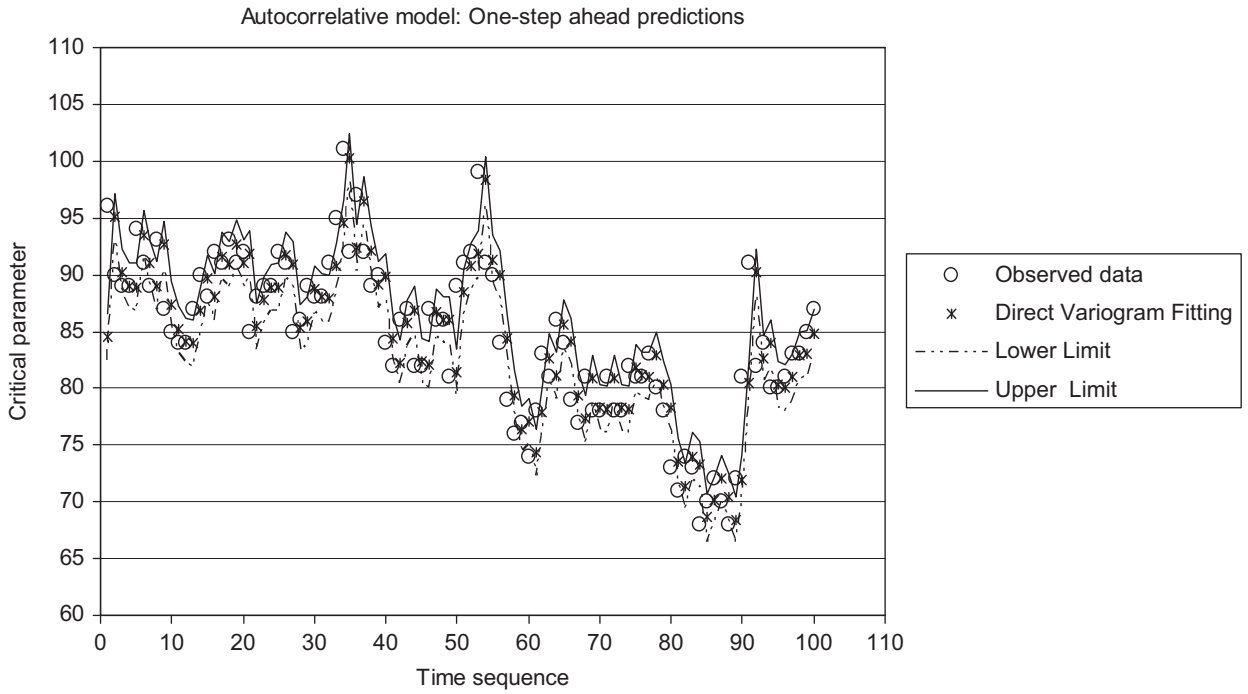

Fig. 2. Application of autocorrelative model by variogram fitting method.

The estimates for σ_η^2 and ρ are different from the corresponding estimates based on the variogram method. This can be explained by the fact that the overall pattern of the observed variogram supports the steady model and the *PRESS* method takes into account the behaviour of the whole sequence, not just a section of the variogram.

6.2. Parametric approach

The estimates of the three parameters obtained by the variogram approach are used as the starting values for applying the approximate maximum likelihood approach of Section 5.2.2. The results are obtained as

$$\sigma_\varepsilon^2 = 1.1508, \quad \sigma_\eta^2 = 11.7044, \quad \rho = 0.8693 \tag{6.7}$$

The ML solution was obtained by direct optimization of the likelihood for fixed value of σ_ε^2 at the estimate derived by the variogram method. Thus the estimates of σ_η^2 and ρ are calculated on the assumption that σ_ε^2 is known with no sampling error.

Since $m_i = m = 1$, $i = 1, \ldots, n$, for the present example, the approximate likelihood method is also identical to the full likelihood method. It should also be noted that, strictly speaking, the ML method in this case does not provide an estimate for σ_ε^2. It provides only the residual mean square which may be taken as a solution of the likelihood equation. The residual mean square of the autocorrelative model fit by the likelihood method is found to be

$$\sigma_{RES}^2 = 1.1505 \tag{6.8}$$

The convergent value of the residual mean square is virtually the same as the variogram estimate of σ_ε^2. Using an approximate χ^2 distribution for the latter the standard error of both estimates may be taken as

$$S.E.(\sigma_{RES}^2) = 0.1627 \tag{6.9}$$

These results indicate that the model fit is in agreement with the data up to the estimated replicate error as suggested by the variogram intercept. The estimates for the remaining two parameters are different from those obtained by the *PRESS* method, but are closer to the estimates obtained by the variogram method. In Fig. 3, the estimates of Eq. (5.6) are used together with the asymptotic weight ($w = 0.92$) (4.10), to construct estimates of λ_i's. The results are again depicted as one-step-ahead predictions.

Finally, an approximate covariance matrix of the ML estimates is computed by using the formulae for the elements of the information matrix (see Appendix A). The elements of the covariance matrix are given below. It should be noted that if replicate observations are available the values of the standard errors should be adjusted for the uncertainty in the

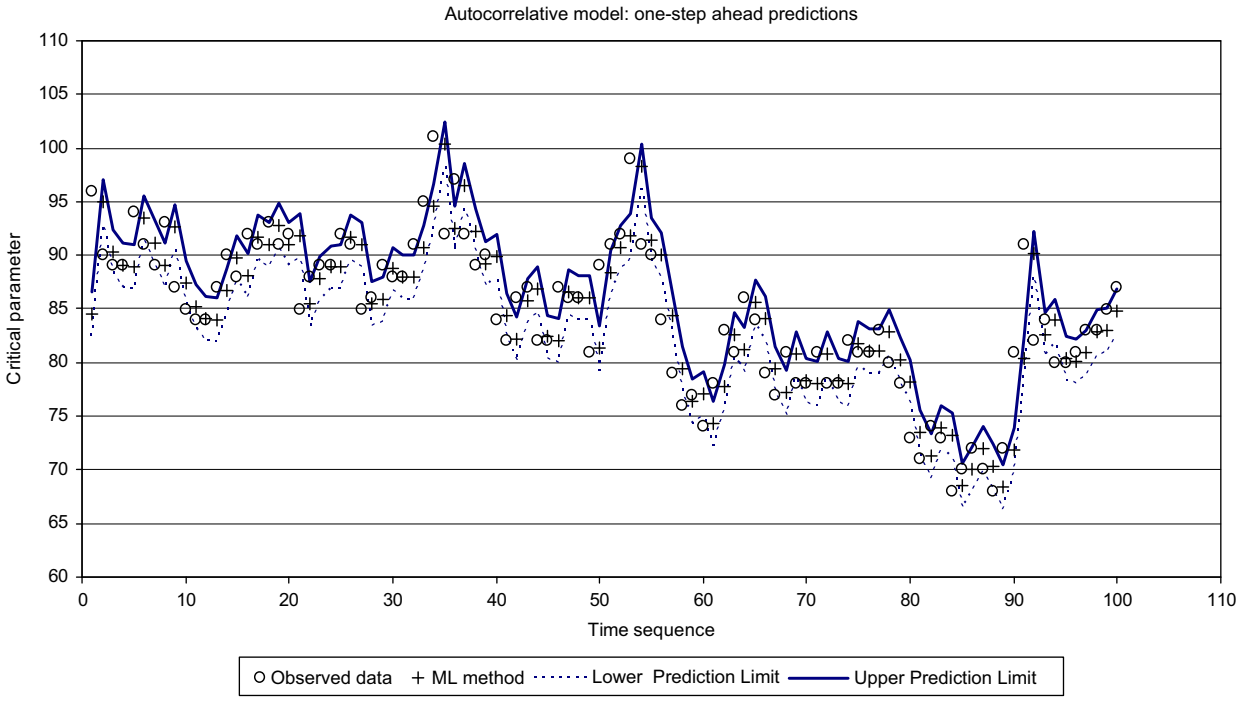

Fig. 3. Application of autocorrelative model by maximum likelihood method.

estimation of σ_ε^2

$$Var(\hat{\mu}_0) = 5.2724, \quad Var(\hat{\gamma^2}) = 5.3571, \quad Var(\hat{\rho}) = 0.000673 \tag{6.10}$$

$$Cov(\hat{\mu}_0, \hat{\gamma^2}) = 0, \quad Cov(\hat{\mu}_0, \hat{\rho}) = 0, \quad Cov(\hat{\gamma^2}, \hat{\rho}) = 0.0154 \tag{6.11}$$

Using the values of the variances of the estimates, the standard errors of the estimates are calculated as follows:

$$S.E(\hat{\mu}_0) = 2.2962, \quad S.E.(\hat{\sigma_\eta^2}) = 2.0888, \quad S.E.(\hat{\rho}) = 0.0259 \tag{6.12}$$

6.3. Control limits based on the autocorrelative model

Let τ be a target value for the underlying quality characteristic for a given monitoring problem. For the current example, we set it to be the process mean 84.52 and employ the estimates ($\hat{\sigma_\varepsilon^2} = 1.1508, \hat{\gamma^2} = 47.91, \hat{\rho} = 0.87, \hat{w} = 0.92$) in Eq. (6.7) to obtain the control limits via Lu and Reynolds (5.2) or Parameterised EWMAST (5.5) approach. The results of the upper and lower EWMA control limits as follows:

$$\text{Control limits (Lu and Reynolds)} = 84.52 \pm 19.31 = (65.2, 103.8)$$

$$\text{Control limits (Parameterised EWMAST)} = 84.52 \pm 20.75 = (63.8,\ 105.3) \tag{6.13}$$

As expected both approaches provide very similar limits and we display only the EWMAST limits in Fig. 4. It is seen that, based on the model estimates of the whole data set as the training data set, the EWMAST control limits suggest that the process is in-control as Montgomery and Mastrangelo (1991) concluded. Nevertheless, in practice the control limits could be based on only a part of the data, such as the first half. The model estimates would then be different and a substantial number of the lower points in the second half of the data would lie outside the control limits.

In general, it is thus necessary to see if the data exhibits a systematic pattern in terms of a run of observations above or below limits based on the basic short-term as well as medium-term variation. Using Eqs. (5.6) and (5.7) these control limits are computed as follows:

$$\text{Control limits (Short\_term)} = 84.52 \pm 2.96 = (81.6, 87.5) \tag{6.14}$$

$$\text{Control limits (Medium\_term)} = 84.52 \pm 10.05 = (74.5,\ 94.6) \tag{6.15}$$

A systematic run of individual observations outside the short-term limits should serve as an initial *"alert"*. Similarly a systematic run of individual observations outside medium-term limits should be regarded as a *second alert* (first warning) of a possible drift in the data to enable the process control operator to prepare for a possible action for correction. As seen in Fig. 4, the process was drifting into a *second alert* regime only briefly between observation number 30 and 60 before

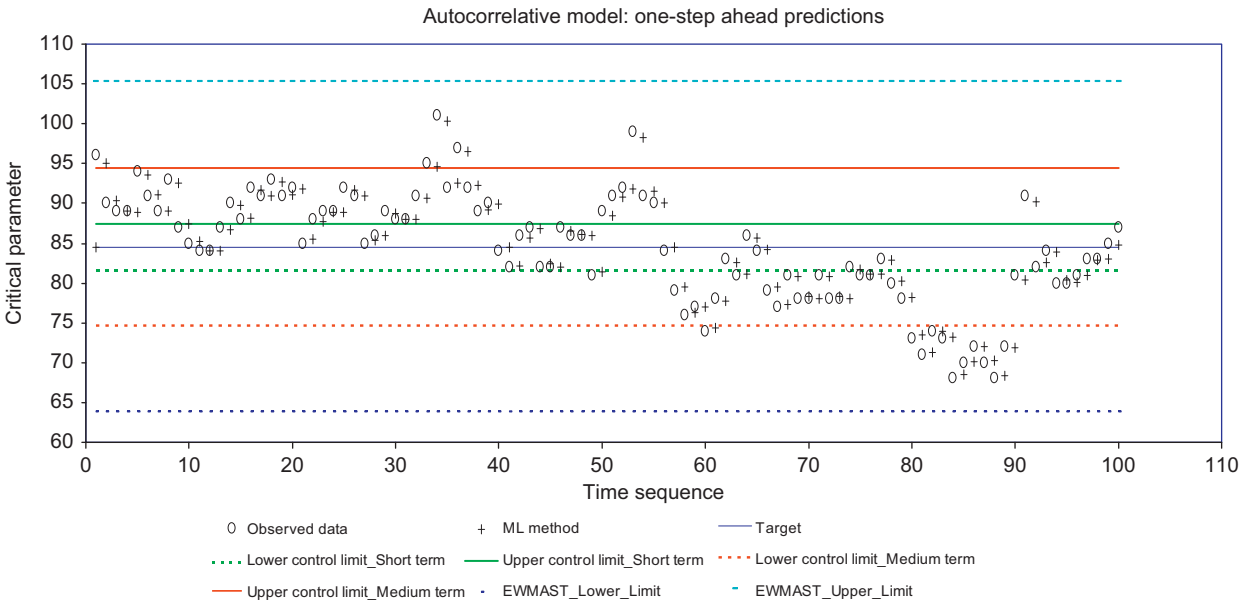

Fig. 4. Control limits based on autocorrelative model by maximum likelihood method.

stabilising between observation number 60 and 80. A systematic downward drift occurred between observation number 80 and 90 which would have signalled a *second alert* (first warning) in a practical monitoring situation. Thus depending on the tightness of the specification limits the operator would start preparation for a necessary action to stabilise the process after the trend signal at observation 80.

Assuming that the complete data set is a training data set, the steady-state ARL values for the EWMAST control limits (6.13) and (6.14) are computed and listed in an ARL- Table 1 in Appendix C, It is seen that for the choice of the parameter $w=0.9$, application of the control limits (6.14) based on medium-term variation (approximately equivalent to assuming $\rho=0.20$ in Table 1 of Appendix C) results in (in-control) ARL value of 371 runs as compared to the application of the control limits (6.13) based on overall process variation (equivalent to assuming $\rho=0.85$ in Table 1 of Appendix C) which results in (in-control) ARL value of 520. On the other hand, compared to Eq. (6.13), the use of Eq. (6.14) gives a reduction in ARL values of about 45% to 55% for signalling a shift in the mean of half to one standard deviation and of about 65% to 85% for signalling a shift of one to three standard deviations. Thus application of Eq. (6.14) has a higher power of detecting real shifts at the expanse of about 30% increase in false positive signals in medium-term.

7. Discussion

The term *statistical process control* is used by way of following the tradition in the area. The word *control* in the term is not regarded as strictly correct for the purpose of the methodology developed here and instead, the word *monitoring* is suggested as a more suitable substitute.

The EWMA approach was initially introduced by Roberts (1959) for the purpose of constructing a control chart for a sequence of independently and identically distributed (IID) observations, in the absence of random effects, as an alternative to Shewhart chart. However even in early applications of these control charts (see e.g. discussions in Vasilopoulos and Stamboulis, 1978), it was widely appreciated that many practical data sets exhibit features indicating the presence of autocorrelation.

There are two main existing approaches to dealing with a sequence of autocorrelated data.

The first approach is concerned with finding adjustments for existing control limits from EWMA approach to account for the data correlation as explained by a time series probability model. Commonly used time series models are, in general, members of the class of Auto-Regressive Integrated Moving Average (*ARIMA*) models. An adjustment needs to be worked out for each chosen model and such calculations had been made only for a few models in the *ARIMA* class. The choice of an appropriate model is an issue and its resolution requires a large sample size among others.

The second approach employs a time series model also from the class of *ARIMA* models. Once an appropriate member of this class is identified, the approach calculates the residuals of the observed data from the fitted model and applies Shewhart chart to the sequence of residuals. The main issues for this approach are the same as those for the first approach. In addition, an accurate theory is needed for constructing efficient control limits for the residual observations from a chosen time-series model.

A compromise between the above two major approaches is given by Montgomery and Mastrangelo (1991), who use the EWMA for the control statistic with weights determined by minimization of one-step ahead prediction error. The resulting control statistic was then used as a one-step-ahead predictor to calculate the residuals of the observations. A Shewhart chart was constructed to the sequence of residuals to obtain control limits, which are then translated into control limits for original observations. An advantage of this approach is that there is not a necessity to choose a specific model in constructing the control statistic and the method is reasonably simple for application. However, there are a number of issues raised by this approach.

First the method does not have a well defined procedure that identifies various sources of variation and subsequently evaluates the corresponding magnitudes.

The choice of the smoothing constant, based on one-step-ahead prediction error, is not interpretable in terms of the various magnitudes of variation. Since the controlled limits are constructed for the residuals by using Shewhart procedure, the prediction intervals are not based on an explicit modelling of process variation.

The approach of the present paper is along the lines of the first approach with emphasis on model checking and efficient model parameter estimation for the case of the *AR(1)–EIV* (*autocorrelative*) model. Issues relating to evaluation of various sources and magnitudes of variation as well as the connection between the smoothing constant and the different magnitudes of variation are resolved. The selection of such a model is also aided by a model check procedure based on an inspection of the variogram as well as an analysis of the actual residuals of the model fit. Linear and exponential type variograms are frequently observed in process monitoring data (Gy, 1979) justifying a detailed study of the steady and autocorrelative models.

The autocorrelative model is seen to be closely connected to the Kalman filter technique, through a Bayesian development, in constructing one-step-ahead predictors in a recursive formula. It is reduced to the steady model when the serial correlation parameter is unity. It also contains the classical IID model as a special case when the serial correlation parameter is zero. The results for the IID model can readily be deduced from those of the autocorrelative model. However, the steady model was treated separately, since the results for it cannot always be readily deduced by a simple substitution in those of the autocorrelative model.

The Bayes type predictor for a given time t was obtained as a function of the current data as well as the parameters of the dynamic model. These parameters are estimated from the previous data by simple method of moment procedures as well as the maximum likelihood procedures. The empirical Bayes technique was demonstrated as an appropriate method of model parameter estimation.

The methodology of the present paper leads to estimators of the form similar to those of EWMA. While the EWMA estimators were originally developed for the IID sequence of observations, the present approach allows for the incorporation of autocorrelation in the sequence of the data. Furthermore the weights necessary for the construction of the estimators are now shown to be directly related to the variational parameters of the underlying probability model.

The Bayesian development for the IID case was considered by Hoadley (1981), with Poisson variation for the observations. The steady model has been considered by Crowder (1986) and Kirkendall (1986) using Gaussian distribution assumption for the observations and a modified likelihood function. Crowder and Eshleman (2001) considered the full likelihood function of the steady model for the case when all group sizes are unity, i.e. $m_i=1$, ($i=1, \ldots, n$). The present paper extends this approach to the autocorrelative model without restriction on the group sizes. It also provides alternative methods which do not depend on a Gaussian distributional assumption of the observations.

Application of the autocorrelative model to the analysis of a data set published in Montgomery and Mastrangelo (1991) brings out, more clearly, the issue of realistic estimation of a short-term variance. Neither replicate data nor data generated independently to provide short-term variation are available for this example. Thus only the extrapolation from a fitted variogram model was used as an estimate of σ_ε^2.

Application of the approximate ML method, which is the same as the full ML method for $m_i=1$, ($i=1, \ldots n$), was carried out under the Gaussian distribution assumption of the errors. The ML approach also provides an alternative estimate of σ_ε^2 as a function of the residual sum of the squares of the model fit. The alternative estimate (1.1505) was found to be practically the same as the value (1.1508) obtained by the variogram method suggesting that the ML analysis supports the autocorrelative model. The ML estimates of the remaining parameters were also found to be in reasonable agreement with those obtained by the variogram method.

These estimates enable a more efficient construction of relevant control limits based on estimates of short-term as well as long-term variation, in the presence of data autocorrelation. Use of these control limits, relative to the grand mean as a target value, for the particular example (see Fig. 4) clearly demonstrates that a significant drift of the process mean exists in the second half of the data sequence as compared to the first. Thus, while accepting the conclusion arrived at by the analysis of Montgomery and Mastrangelo (1991) for the same case tentatively, the present approach draws the process operator's attention to the possibility of a significant drift caused by a longer-term variation of the observations.

Finally, the suggested Parametric EWMAST approach is summarised as follows:

Step 1: Specify a training data set under stationary conditions for the underlying process and check the suitability of the autocorrelative model by using the variogram. If the steady model or an autocorrelative model is reasonable proceed to ***Step 2***.

Step 2: Determine the parameter estimates by ML method. Based on the uncertainty of the estimates, respecify the size of the training data set to achieve the desired level of uncertainty of the estimate of long-term variation.

Step 2: Using the estimates of ρ and w construct the Parametric EWMAST control limits based on short-term and medium-term variation using the formulae (5.7) and (5.8) and also based on over-all process variation using the formula (5.6).

Step 3: If the process wanders outside the control limits based on medium-term variation, issue a warning to make preparations for taking action and increase the degree of monitoring activity until the process wanders back into control limits based on short-term variation.

Step 4: If the process wanders outside the control limits based on over-all process variation, issue a warning to activate the operations necessary to bring the process level into short-term or medium-term control limits as necessary. If such operations cannot bring the process under control, issue an out-of-control warning to activate critical process review.

Acknowledgments

The author is indebted to Dr. Ron Sandland of CSIRO, Australia for drawing his attention to the work by Harrison (1967) and for an enthusiastic encouragement of the present work in its earlier stages.

Appendix A. Maximum likelihood approach for general parametric error distributions

The likelihood function of the data $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ can be derived by assuming a parametric form for the probability distribution of $\mathbf{x}_i$ given $\lambda_i (i=1, \ldots, n)$ and also for the joint probability distribution of $\{\lambda_1, \ldots, \lambda_n\}$. Specifically the following assumptions are made:

Conditional on $\{\lambda_1, \ldots, \lambda_n\}$, the quantities $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ are independently distributed.

The probability density function (pdf) $f(\mathbf{x}_i|\lambda_i,\boldsymbol{\kappa}_i)$ of the random vector $\mathbf{x}_i$ is given by

$$f(\mathbf{x}_i|\lambda_i,\boldsymbol{\kappa}_i) = \Pi_{j=1}^{m_i} f(x_{ij}|\lambda_i,\boldsymbol{\kappa}_i) \tag{A.1}$$

In the specific application considered below, the pdf $f(x_{ij}|\lambda_i,\boldsymbol{\kappa}_i)$ is assumed to be a univariate Gaussian (UVG) probability density function, with mean λ_i and variance σ_ε^2 defined in Eq. (2.1). Thus $\boldsymbol{\kappa}_i=\{\sigma_\varepsilon^2\}$. The joint distribution of $\boldsymbol{\lambda}=(\lambda_1,\ \ldots,\ \lambda_n)^T$ has a probability density function $g(\boldsymbol{\lambda}|\xi)$, where ξ is a vector of parameters of the pdf $g(.|.)$. In the specific application considered below, the function $g(\boldsymbol{\lambda}|\xi)$ is assumed to be a multivariate Gaussian (MVG) density function with mean vector $\boldsymbol{\mu}=\mu_0\mathbf{1}$ and covariance matrix $\gamma^2\boldsymbol{\Gamma}$, where $\mathbf{1}$ is an $(n\times 1)$ unit vector, γ^2 is as defined in Eq. (2.8) and $\boldsymbol{\Gamma}$ is an $(n\times n)$ matrix with (ij)th element

$$\Gamma_{ij}=\rho^{|i-j|} \tag{A.2}$$

This is the stationary pdf of $\boldsymbol{\lambda}$, when the quantities $\{\eta_1,\ \ldots,\ \eta_n\}$ are IID Gaussian random variables with mean 0 and variance σ_η^2 in the *AR(1)* model (1.3). In this case $\xi=(\mu_0,\sigma_\eta^2,\rho)$ are parameters of $g(.|.)$, which are reparametrized for the purpose of analysis below as $\xi=(\mu_0,\gamma^2,\rho)$.

The above Gaussian assumptions are summarised as

$$x_{ij}:UVG(\lambda_i,\sigma_\varepsilon^2) \tag{A.3}$$

and

$$\lambda:MVG(\xi,\gamma^2\boldsymbol{\Gamma}) \tag{A.4}$$

Now we first consider the maximum likelihood (ML) approach for general pdf's f and g. Let

$$\boldsymbol{\psi}=(\boldsymbol{\kappa}_1,\ldots,\boldsymbol{\kappa}_n,\xi) \tag{A.5}$$

Under the above assumptions, the marginal joint distribution function $H(.|.)$ of the observed data has the pdf, h, given below

$$h(\mathbf{X}|\boldsymbol{\psi})=\int \Pi_{i=1}^{n} f(\mathbf{x}_i|\lambda_i,\boldsymbol{\kappa}_i)g(\boldsymbol{\lambda}|\xi)d\boldsymbol{\lambda} \tag{A.6}$$

Define the quantities

$$l(\mathbf{X}|\boldsymbol{\lambda},\boldsymbol{\kappa}_1,\ldots,\boldsymbol{\kappa}_n)=\Pi_{i=1}^{n} f(\mathbf{x}_i|\lambda_i,\boldsymbol{\kappa}_i) \tag{A.7}$$

and

$$p(\mathbf{X}|\boldsymbol{\psi})=l(\mathbf{X}|\boldsymbol{\lambda},\boldsymbol{\kappa}_1,\ldots,\boldsymbol{\kappa}_n)g(\boldsymbol{\lambda}|\xi)$$

The structure of Eq. (A.6) is suitable for application of established methodology in the Empirical Bayes (EB) approach to the estimation of unknown parameters. One can treat the function $l(\mathbf{X}|\boldsymbol{\lambda},\boldsymbol{\kappa}_1,\cdots,\boldsymbol{\kappa}_n)$ as the data distribution of the observed data, the function $g(\boldsymbol{\lambda}|\xi)$ as the prior distribution of the random effects and construct the "posterior" distribution function, $B(\boldsymbol{\lambda}|\boldsymbol{X},\boldsymbol{\psi})$ with the pdf given by

$$B(\boldsymbol{\lambda}|\mathbf{X},\boldsymbol{\psi})=p(\mathbf{X}|\boldsymbol{\psi})/h(\mathbf{X}|\boldsymbol{\psi}) \tag{A.8}$$

The problem of estimating the model parameters $\boldsymbol{\psi}$, is now the same as the problem of estimating, simultaneously, the unknown fixed parameters of the data distribution and the fixed parameters of the "prior" distribution. If the parametric forms of the "prior" and "data" distributions are known, standard methods can be applied to estimate these parameters as follows (see Maritz and Lwin, 1989, pp. 49–52), for a more detailed explanation of the parametric EB approach.

Let ψ_u and ψ_v be two generic elements of the unknown vector parameter $\boldsymbol{\psi}$. First we note that the marginal log-likelihood function of the data is given by L obtained by taking the logarithm of the pdf $h(.|.)$. Assume that the quantity on the right-hand side of Eq. (A.6) is differentiable under the integral sign. Denote by the operator $E_B(.)$ the expectation operation with respect to the posterior distribution $B(\boldsymbol{\lambda}|\boldsymbol{X},\boldsymbol{\psi})$. Then we can express the first and second derivatives of L with respect to ψ_u and ψ_v as

$$\partial L/\partial\psi_u=E_B(\partial\ln p(\mathbf{X}|\boldsymbol{\psi})/\partial\psi_u=E_B(\partial\ln l/\partial\psi_u)+E_B(\partial\ln g/\partial\psi_u) \tag{A.9}$$

$$\partial^2 L/(\partial\psi_u\partial\psi_v)=E_B\{\partial^2\ln p(\mathbf{X}|\boldsymbol{\psi})/(\partial\psi_u\partial\psi_v)\}=E_B\{\partial\ln l/(\partial\psi_u\partial\psi_v)\}+E_B\{\partial\ln g/(\partial\psi_u\partial\psi_v)\} \tag{A.10}$$

Then the likelihood equation for a generic parameter ψ_u is given by

$$E_B(\partial\ln l/\partial\psi_u)+E_B(\partial\ln g/\partial\psi_u)=0 \tag{A.11}$$

Next, an iterative algorithm based on Newton–Raphson method can be developed to derive solutions to these likelihood equations as follows.

Let $\boldsymbol{\psi}_0$ be an initial estimate of $\boldsymbol{\psi}$, assumed to be of dimension q. Let $\mathbf{S}$ be a q vector whose uth element is given by the quantity (4.25) and also let $\mathbf{I}$ be a $(q\times q)$ matrix with (u,v)th element given by the quantity (4.26), with all expressions being evaluated at $\boldsymbol{\psi}=\boldsymbol{\psi}_0$. Then the scoring method of solving the likelihood equations for parameters can be applied to obtain an iterative solution for an estimate $\boldsymbol{\psi}$ through the equation:

$$\hat{\boldsymbol{\psi}}=\boldsymbol{\psi}_0-\mathbf{I}^{-1}\mathbf{S} \tag{A.12}$$

When the iteration converges, the corresponding asymptotic covariance matrix of the parameter estimates is given by

$$Cov(\hat{\boldsymbol{\psi}}) = -\{E_H(\mathbf{I})\}^{-1} \tag{A.13}$$

For the particular case under consideration, the corresponding quantities **S** and **I** for the unknown parameters, $(\sigma_\varepsilon^2,\mu_0,\rho,\gamma^2)$, can be expressed in terms of posterior expectations and covariances of functions of λ_i.

We now specialise the general likelihood equations, (A.9), to the special case of Gaussian distributions (A.3) and (A.4) for f and g.

Let

$$N = \sum_{i=1}^{n} m_i$$
$$F = \alpha_1^2 + \alpha_n^2; \quad G = \sum_{i=2}^{n-1} \alpha_i^2; \quad H = \sum_{i=1}^{n-1} \alpha_i \alpha_{i+1}$$

First note (see also Glasbey, 1979) that

$$\partial \ln l/\partial \sigma_\varepsilon^2 = -(1/2)[(N/\sigma_\varepsilon^2) + (1/2)\sum_{i=1}^{n}\sum_{j=1}^{m_i}\{(x_{ij}-\lambda_i)^2/\sigma_\varepsilon^4\}]$$

$$\partial \ln g/\partial \mu_0 = 2\mathbf{1}^T\boldsymbol{\Gamma}^{-1}(\boldsymbol{\lambda}-\boldsymbol{\mu})$$
$$\partial \ln g/\partial \gamma^2 = -n(2\gamma^2)^{-1} - (2\gamma^4)^{-1}(\boldsymbol{\lambda}-\boldsymbol{\mu})^T\boldsymbol{\Gamma}^{-1}(\boldsymbol{\lambda}-\boldsymbol{\mu})$$
$$\partial \ln g/\partial \rho = \{\gamma^2(1-\rho^2)^2\}^{-1}\{-(n-1)\rho(1-\rho^2)\gamma^2 + \rho F + 2\rho G - (1+\rho^2)H\}$$

Using the above relations and noting that $\boldsymbol{\mu}=\mu_0\mathbf{1}$, we have the likelihood equations for $(\sigma_\varepsilon^2,\mu_0,\rho,\gamma^2)$ as

$$\hat{\mu_0} = (\mathbf{1}^T\boldsymbol{\Gamma}^{-1}\mathbf{1})^{-1}\{\mathbf{1}^T\boldsymbol{\Gamma}^{-1}E_B(\boldsymbol{\lambda})\} \tag{A.14}$$

$$\hat{\sigma_\varepsilon^2} = N^{-1}\sum_{i=1}^{n} E_B\{(\mathbf{x}_i-\lambda_i\mathbf{1})^T(\mathbf{x}_i-\lambda_i\mathbf{1})\} \tag{A.15}$$

$$\hat{\gamma^2} = n^{-1}tr[\boldsymbol{\Gamma}^{-1}\{E_B(\boldsymbol{\lambda}-\boldsymbol{\mu})(\boldsymbol{\lambda}-\boldsymbol{\mu})^T\}] \tag{A.16}$$

$$\hat{\rho} = \rho - J_1(\rho)/J_2(\rho) \tag{A.17}$$

where

$$J_1(\rho) = \{\gamma^2(1-\rho^2)\}^{-1}\{-(n-1)\gamma^2\rho(1-\rho^2) + \rho E_B(F) + 2\rho E_B(G) - (1+\rho^2)E_B(H)\}$$

$$J_2(\rho) = \{\gamma^2(1-\rho^2)^3\}^{-1}[-(n-1)\gamma^2(1+\rho^2)(1-\rho^2) + \{E_B(F)+2E_B(G)\}(1+3\rho^2) - 2\rho(3+\rho^2)E_B(H)]$$

The above likelihood equations can be used as iterative sequences for constructing sequences of estimates for each of the unknown parameters, by calculating the right-hand side of the equations at an initial set of values for the parameters to get an updated sequence of the parameter estimates on the left hand side. The quantities $E_B(F),E_B(G),E_B(H)$ in the expressions in (A.17) are obtained from the posterior mean and covariance matrix of $\boldsymbol{\lambda}$ derived below.

Let $E_B(\boldsymbol{\lambda})$ and $Cov_B(\boldsymbol{\lambda})$ be the expectation and covariance of $\boldsymbol{\lambda}$ with respect to the distribution $B(\boldsymbol{\lambda}|\boldsymbol{X},\boldsymbol{\psi})$. Let $E_B(\lambda_i)$ and $Var_B(\lambda_i)$ be the corresponding quantities for the ith component λ_i. Also let $\boldsymbol{\Gamma}_\rho$ and $\boldsymbol{\Gamma}_{\rho\rho}$ be $(n\times n)$ matrices with (i,j)th elements $|i-j|\rho^{|i-j|-1}$ and $|i-j|(|i-j|-1)\rho^{|i-j|-2}$, respectively. Then elements of the information matrix for the parameter estimates can be calculated using the following formulae:

$$E_B\{\partial^2 \ln l/\partial\sigma_\varepsilon^4\} = (1/2)[(N/\sigma_\varepsilon^4) - \sum_{i=1}^{n} E_B\{(\mathbf{x}_i-\lambda_i\mathbf{1})^T(\mathbf{x}_i-\lambda_i\mathbf{1})/\sigma_\varepsilon^6\}] \tag{A.18}$$

$$E_B\{\partial^2 \ln g/\partial\mu_0^2\} = -\mathbf{1}^T\boldsymbol{\Gamma}^{-1}\mathbf{1} \tag{A.19}$$

$$E_B\{\partial^2 \ln g/\partial\mu_0\partial\gamma^2\} = -\mathbf{1}^T\boldsymbol{\Gamma}^{-1}\gamma^{-2}E_B(\boldsymbol{\lambda}-\boldsymbol{\mu}) \tag{A.20}$$

$$E_B\{\partial^2 \ln g/\partial\mu_0\partial\rho\} = -\mathbf{1}^T(\boldsymbol{\Gamma}^{-1}\boldsymbol{\Gamma}_\rho\boldsymbol{\Gamma}^{-1})\gamma^{-2}E_B(\boldsymbol{\lambda}-\boldsymbol{\mu}) \tag{A.21}$$

$$E_B\{\partial^2 \ln g/\partial\gamma^2\partial\gamma^2\} = n/(2\gamma^4) - (2/\gamma^6)E_B\{Q_1(\boldsymbol{\lambda})\} \tag{A.22}$$

$$E_B\{\partial^2 \ln g/\partial\gamma^2\partial\rho\} = (2\gamma^4)^{-1}E_B\{Q_2(\boldsymbol{\lambda})\} \tag{A.23}$$

$$E_B\{\partial^2 \ln g/\partial\rho\partial\rho\} = -(1/2)\,trace\{\boldsymbol{\Gamma}^{-1}\boldsymbol{\Gamma}_{\rho\rho} - \boldsymbol{\Gamma}_\rho\boldsymbol{\Gamma}^{-1}\boldsymbol{\Gamma}_\rho\boldsymbol{\Gamma}^{-1}\} + (1/2\gamma^2)[E_B\{Q_4(\boldsymbol{\lambda})\} - \{E_B\{Q_3((\boldsymbol{\lambda}))\}] \tag{A.24}$$

where

$$\sum_{i=1}^{n} E_B\{(\mathbf{x}_i-\lambda_i\mathbf{1})^T(\mathbf{x}_i-\lambda_i\mathbf{1}) = \sum_{i=1}^{n}[\{\mathbf{x}_i-E_B(\lambda_i)\mathbf{1}\}^T\{\mathbf{x}_i-E_B(\lambda_i)\mathbf{1}\} + Var_B(\lambda_i)]$$

$$E_B\{Q_1(\boldsymbol{\lambda})\} = E_B\{(\boldsymbol{\lambda}-\boldsymbol{\mu})^T\boldsymbol{\Gamma}^{-1}(\boldsymbol{\lambda}-\boldsymbol{\mu})\} = \{E_B(\boldsymbol{\lambda})-\boldsymbol{\mu}\}^T\boldsymbol{\Gamma}^{-1}\{E_B(\boldsymbol{\lambda})-\boldsymbol{\mu}\} + trace\{\boldsymbol{\Gamma}^{-1}\,Cov_B(\boldsymbol{\lambda})\}$$
$$E_B\{Q_2(\boldsymbol{\lambda})\} = E_B\{(\boldsymbol{\lambda}-\boldsymbol{\mu})^T\boldsymbol{\Gamma}^{-1}\boldsymbol{\Gamma}_\rho\boldsymbol{\Gamma}^{-1}(\boldsymbol{\lambda}-\boldsymbol{\mu})\} = \{E_B(\boldsymbol{\lambda})-\boldsymbol{\mu}\}^T\boldsymbol{\Gamma}^{-1}\boldsymbol{\Gamma}_\rho\boldsymbol{\Gamma}^{-1}\{E_B(\boldsymbol{\lambda})-\boldsymbol{\mu}\} + trace\{\boldsymbol{\Gamma}^{-1}\boldsymbol{\Gamma}_\rho\boldsymbol{\Gamma}^{-1}\,Cov_B(\boldsymbol{\lambda})\}$$
$$E_B\{Q_3(\boldsymbol{\lambda})\} = E_B\{(\boldsymbol{\lambda}-\boldsymbol{\mu})^T\boldsymbol{\Gamma}^{-1}\boldsymbol{\Gamma}_\rho\boldsymbol{\Gamma}^{-1}\boldsymbol{\Gamma}_\rho(\boldsymbol{\lambda}-\boldsymbol{\mu})\} = \{E_B(\boldsymbol{\lambda})-\boldsymbol{\mu}\}^T\boldsymbol{\Gamma}^{-1}\boldsymbol{\Gamma}_\rho\boldsymbol{\Gamma}^{-1}\boldsymbol{\Gamma}_\rho\{E_B(\boldsymbol{\lambda})-\boldsymbol{\mu}\} + trace[\boldsymbol{\Gamma}^{-1}\boldsymbol{\Gamma}_\rho\boldsymbol{\Gamma}^{-1}\boldsymbol{\Gamma}_\rho\,Cov_B(\boldsymbol{\lambda})\}$$
$$E_B\{Q_4(\boldsymbol{\lambda})\} = E_B\{(\boldsymbol{\lambda}-\boldsymbol{\mu})^T\boldsymbol{\Gamma}^{-1}\boldsymbol{\Gamma}_{\rho\rho}\boldsymbol{\Gamma}^{-1}(\boldsymbol{\lambda}-\boldsymbol{\mu})\} = \{E_B(\boldsymbol{\lambda})-\boldsymbol{\mu}\}^T\boldsymbol{\Gamma}^{-1}\boldsymbol{\Gamma}_{\rho\rho}\boldsymbol{\Gamma}^{-1}\{E_B(\boldsymbol{\lambda})-\boldsymbol{\mu}\} + trace[\boldsymbol{\Gamma}^{-1}\boldsymbol{\Gamma}_{\rho\rho}\boldsymbol{\Gamma}^{-1}\,Cov_B(\boldsymbol{\lambda})\}$$

For Gaussian case the quantities $E_B(\boldsymbol{\lambda})$ and $Cov_B(\boldsymbol{\lambda})$ can be explicitly evaluated as

$$E_B(\boldsymbol{\lambda}) = \{Cov_B(\boldsymbol{\lambda})\}^{-1}(\sigma_\varepsilon^{-2}\mathbf{D}^{-1}\bar{\mathbf{x}}+\gamma^{-2}\boldsymbol{\Gamma}^{-1}\boldsymbol{\mu}) \tag{A.25}$$

$$Cov_B(\boldsymbol{\lambda}) = (\sigma_\varepsilon^{-2}\mathbf{D}^{-1}+\gamma^{-2}\boldsymbol{\Gamma}^{-1})^{-1} \tag{A.26}$$

Where $\mathbf{D}$ is an $(n\times n)$ diagonal matrix with the ith diagonal matrix $1/m_i$ and $\bar{\mathbf{x}}$ is an $(n\times 1)$ vector with ith element $x_{\cdot i}$ and $\boldsymbol{\Gamma}^{-1}$ can be evaluated analytically (see e.g. Galbraith and Galbraith, 1974).

A remark is now in order. Evaluation of these quantities are computationally intensive since it involves inversion of matrices (Gaussian case) or multiple integrals (general distributions) of large dimensions which increase with increasing value of n. Thus maximisation of the likelihood function itself is computationally tedious. For the special case when the group sizes are equal, the explicit formula for a matrix of the form (A.26), can be employed to reduce the computational burden substantially (see Appendix B).

Appendix B. Explicit inverse of a patterned matrix

For the autocorrelative model, we have the matrix

$$\mathbf{P}(\phi,\rho) = \mathbf{D}+\phi\boldsymbol{\Gamma}$$

where $\boldsymbol{D}$ is an $(n\times n)$ diagonal matrix, with ith diagonal element $1/m_i$ and $\boldsymbol{\Gamma}$ is an $(n\times n)$ matrix with the (i,j)th element $\Gamma_{ij}=\rho^{|i-j|}$. In the special case when $m_i=m$, ($i=1,\ \ldots\ n$), the inversion of the matrix $\mathbf{P}(\phi,\rho)$ can be carried out using an analytical expression due to Galbraith and Galbraith, 1974). It substantially reduces the computational burden in applying the ML technique.

Let $\theta=1/(m\phi)$. Also let β be the smaller root of the equation

$$\rho\theta\beta^2-\{(1-\rho^2)+\theta(1+\rho^2)\}\beta+\rho\theta = 0$$

and define the quantity T by

$$T = (1-\beta^2)\{(1-\rho\beta)^2-(\beta-\rho)\beta^{(2n)}\}$$

Let $\mathbf{M}$ be an $(n\times n)$ matrix with (i,j)th element given by

$$m_{rr} = T^{-1}(1-\rho\beta)^2(\beta-\rho)\{1-\beta^{2(r-1)}\}\{1-\beta^{2(n-r)}\} + T^{-1}(1-\beta^2)(1-\rho\beta)^2-(\beta-\rho)^2\{1-\beta^{2(n-1)}\}$$

and for $r<s$

$$m_{rs} = T^{-1}\beta^{(s-r-1)}(1-\rho\beta)(\beta-\rho)\{(1-\rho\beta)-(\beta-\rho)\beta^{2r-1}\}\{(1-\rho\beta)-(\beta-\rho)\beta^{2(n-s)+1}\}$$

Then

$$\mathbf{P}^{-1}(\phi,\rho) = \phi(\theta\mathbf{I}+\boldsymbol{\Gamma}) = \phi(1-\rho\beta)(\rho-\beta)\{\rho(1-\rho^2)\}^{-1}\mathbf{M}$$

Furthermore, the determinant of $\mathbf{P}(\phi,\rho)$ can also be explicitly obtained as
$|\mathbf{P}(\phi,\rho)| = \{\phi^{-1}\rho(1-\rho^2)\}^n\{(1-\rho\beta)(\rho-\beta)\}^{-n}\{(1-\rho^2)(1-\beta^2)\}^{-1}\{(1-\rho\beta)-(\beta-\rho)\beta^{2n}\}$

Appendix C. Calculation of stationary Average run length of Parametric EWMAST based on asymptotic distribution of the process probability model

We adopt the approach of Crowder (1987) to calculate the ARL values of Parametric EWMAST for the autocorrelative model, assuming the measurement error σ_ε^2 is negligible. In this case the process variation model is the same as $AR(1)$ model for which simulated ARL values have been tabulated by Zhang (1998) for selected values of the smoothing parameter not exceeding 0.4, using simulation for the exact model. We extend the results to provide similar calculations for selected values of the smoothing parameter above 0.4 to cover the case of the Example in Section 6. We assume that $\mu=\delta\gamma^2$ and $m=1$. Let $ARL(u)$ be the average run length given that EWMAST starts with $Y_t=u$ after the process reaches steady state. If the stationary distribution of the first observation x_1 after the start of the steady state is denoted by $F(x|u)$ and its density function $F(x|u)$ then following the arguments of Crowder (1987), we have

$$ARL(u) = 1+\int_{|(1-w)u+wx|\le h}\{1+ARL[\{x-(1-\omega)u\}/\omega]f(y|u)dy \tag{C1}$$

Under the Gaussian assumption for the data, the distribution $F(x|u)$ is that of a Gaussian distribution with mean $E(x|u)$ and variance $Var(x|u)$, which are, respectively, evaluated according to the asymptotic formulae given as follows:

$$E(x|u) = \mu + B(u-\mu)$$

$$Var(x|u) = \gamma^2 - B^2 Var\{Y_t\}$$

where $Var(Y_t)$ is the steady state variance of the EWMA whose asymptotic form is given by

$$Var(Y_t) = w(2-w)^{-1}\{1+\rho^2(1-w)\}\{1-\rho^2(1-w)\}^{-1}$$

and B is the ratio

$$B = Cov(x_t, Y_{t-1})/Var(Y_t)$$

where the asymptotic form of the covariance is given by

$$Cov(x_t, Y_{t-1}) = \rho^2(1-w)\gamma^2\{1-\rho^2(1-w)\}^{-1}$$

Table 1
ARL.

Autocorrelation	Shift parameter	Parameter w				
		0.2	0.3	0.4	0.8	0.9
ρ=0.00	0	559.87	465.55	421.16	372.85	370.95
	0.5	44.13	53.16	63.59	119.14	136.55
	1	10.84	11.70	13.35	28.49	35.31
	2	3.80	3.51	3.42	4.42	5.17
	3	2.41	2.15	1.98	1.80	1.87
ρ=0.20	0	581.32	478.67	429.69	374.05	371.51
	0.5	50.01	59.07	69.25	121.74	138.05
	1	12.29	13.18	14.87	29.66	36.14
	2	4.24	3.90	3.79	4.75	5.46
	3	2.65	2.35	2.17	1.95	2.01
ρ=0.25	0	594.18	486.61	434.92	374.87	371.92
	0.5	53.65	62.68	72.67	123.29	138.97
	1	13.19	14.10	15.81	30.36	36.65
	2	4.51	4.14	4.01	4.94	5.63
	3	2.80	2.47	2.28	2.04	2.09
ρ=0.35	0	631.65	509.95	450.51	377.64	373.43
	0.5	64.74	73.43	82.69	127.78	141.70
	1	15.98	16.89	18.62	32.39	38.11
	2	5.34	4.87	4.69	5.49	6.11
	3	3.26	2.86	2.63	2.30	2.32
ρ=0.45	0	690.10	546.85	475.67	382.97	376.62
	0.5	83.34	90.87	98.51	134.73	146.13
	1	20.77	21.60	23.23	35.49	40.41
	2	6.74	6.10	5.82	6.33	6.85
	3	4.04	3.50	3.19	2.69	2.69
ρ=0.55	0	781.11	605.09	516.21	393.23	383.32
	0.5	115.13	119.37	123.49	145.50	153.37
	1	29.26	29.69	30.89	40.25	44.01
	2	9.17	8.20	7.71	7.62	7.95
	3	5.38	4.60	4.13	3.30	3.23
ρ=0.65	0	929.44	701.06	584.26	413.58	397.69
	0.5	172.96	168.41	164.81	162.95	165.82
	1	45.52	44.55	44.40	47.86	49.89
	2	13.75	12.09	11.12	9.69	9.71
	3	7.87	6.60	5.82	4.27	4.07
ρ=0.75	0	1198.67	876.38	710.32	457.05	430.52
	0.5	291.63	262.48	240.53	194.37	189.55
	1	81.50	75.49	71.04	61.35	60.57
	2	23.73	20.29	18.05	13.41	12.82
	3	13.19	10.78	9.23	6.02	5.56
ρ=0.85	0	2250.10	1279.71	1001.97	913.21	519.54
	0.5	636.13	488.95	414.04	468.76	245.87
	1	188.34	157.88	137.39	177.51	84.92
	2	52.88	42.80	36.13	46.83	19.85
	3	28.28	22.07	18.11	22.19	8.93

The integral Eq. (C1) can be solved by using Gaussian quardrature as shown by Crowder (1987). The ARL values are calculated by varying the values of (ρ,w)and shift parameter δ as shown in the Table 1 below.

References

Bather, J., 1963. Control charts and the minimization of costs. J. R. Stat. Soc. Ser. B 25, 49–80.

Cressie, N., 1991. In: Statistics for spatial data. John Wiley, New York.

Crowder, S., 1986. An application of adaptive Kalman Filtering to statistical Process control. In: Keats, J.B., Hubele, N.F. (Eds.), Statistical Process Control in Automated Manufacturing. Mercel Dekker, New York.

Crowder, S., 1987. Average run lengths of exponentially weighted moving average charts. J. Qual. Technol. 19, 161–164.

Crowder, S., 1989. Design of exponentially weighted moving average schemes. J. Qual. Technol. 21, 155–162.

Crowder, S., Eshleman, L., 2001. Small sample properties of an adaptive filter applied to low volume SPC. J. Qual. Technol. 33, 29–46.

Davis, H.T., Jones, R.H., 1968. Estimation of the innovation variance of a stationary time series. J. Am. Stat. Assoc. 63, 141–149.

Glasbey, C.A., 1979. Correlated residuals in non-linear regression applied to growth data. J. R. Stat. Soc. Ser. C Appl. Stat. 28, 251–259.

Galbraith, R.F., Galbraith, J.F., 1974. On the inverse of some patterned matrices arising in the theory of stationary time series. J. Appl. Prob. 11, 63–71.

Gy, P., 1979. In: Sampling Particulate Materials. Elsevier Scientific Publishing Company, Amsterdam.

Harrison, P.J., 1967. Exponential smoothing and short-term sales forecasting. Manage. Sci. 13, 821–842.

Hoadley, B., 1981. The quality measurement plan. Bell Syst. Tech. J. 61, 2081–2107.

Kalman, R.E., 1960. A new approach to linear filtering and prediction theory. J. Basic Eng. 82, 34–45.

Kirkendall, J., 1986. The relationship between certain Kalman filter models and exponential smoothing models. In: Keats, J.B., Hubele, N.F. (Eds.), 'Statistical Process Control in Automated Manufacturing' by. Mercel Dekker, New York.

Lwin, T., 2005. Uncertainty about uncertainty in material sampling—a crucial part of risk assessment. In: Proceedings of the Second World Conference on Sampling and Blending 2005, Australian Institute of Mining and Metallurgy, Melbourne, Australia, pp. 33–38.

Lu, C.W., Reynolds JR., M.R., 1999. EWMA Control charts for monitoring the mean of the autocorrelated process. J. Qual. Technol. 31, 166–188.

Maritz, J.S., Lwin, T., 1989. Empirical Bayes Methods. Chapman and Hall, London.

Montgomery, D.C., Mastrangelo, C.M., 1991. Some statistical process conrol methods for autocorrelated data (with discussion). J. Qual. Technol. 23, 179–204.

Phadke, M.S., 1982. Quality evaluation plan using adaptive Kalman filtering. Bell Syst. Tech. J. 61, 2081–2107.

Roberts, S.W., 1959. Control chart testing based on geometric moving averages. Technometrics 1, 239–251.

Trench, W.F., 1964. An algorithm for the inverse of finite Toeplitz matrices. J. Soc. Indust. Appl. Math. 12, 515–522.

Vasilopoulos, A.V., Stamboulis, A.P., 1978. Modification of control chart limits in the presence of data correlation. J. Qual. Technol. 10, 20–30.

West, M., Harrison, P., 1989. In: Bayesian forecasting and dynamic models. Springer Verlag, New York.

Zhang, N.H., 1998. A statistical control chart for stationary process data. Technometrics 40, 24–38.