ENVIRONMETRICS
Environmetrics 2001; **12**: 749–756 (DOI: 10.1002/env.497)

Design of change detection algorithms based on the generalized likelihood ratio test

Giovanna Capizzi*,†

Department of Statistics, University of Padua, Via C. Bottisti 241, 35100 Padova, Italy

SUMMARY

A design procedure for detecting additive changes in a state-space model is proposed. Since the mean of the observations after the change is unknown, detection algorithms based on the generalized likelihood ratio test, GLR, and on window-limited type GLR, are considered. As Lai (1995) pointed out, it is very difficult to find a satisfactory choice of both window size and threshold for these change detection algorithms. The basic idea of this article is to estimate, through the stochastic approximation of Robbins and Monro, the threshold value which satisfies a constraint on the mean between false alarms, for a specified window size. A convenient stopping rule, based on the first passage time of an *F*-statistic below a fixed boundary, is used to terminate the iterative approximation. Then, the window size which produces the most desirable out-of-control ARL, for a fixed value of the in-control ARL, can be selected. These change detection algorithms are applied to detect biases on the measurements of ozone, recorded from one monitoring site of Bologna (Italy). Comparisons of the ARL profiles reveal that the full-GLR scheme provides much more protection than the window-limited GLR schemes against small shifts in the process, but the modified window-limited GLR provides more protection against large shifts.

KEY WORDS: fault detection; generalized likelihood ratio test; false alarms; design procedure; air pollution

1. INTRODUCTION

Suppose $x_1, x_2, \ldots$ are independent observations from p_θ. These values can be either the original observations of a given process or convenient transformations of a dependent sequence. Assume also that before an unknown change time, n_0, the parameter is equal to θ_0 and afterwards the change is $\theta_1 \neq \theta_0$.

Change detection problems can be solved differently according to the various levels of the available information about the parameters θ_0 and θ_1. The case which is carried through this article is concerned with the detection of a change in the mean of the observed sequence, under the hypothesis that the parameter θ_1 after the change is unknown, while the parameter before the change is known. In this situation, change detection algorithms can be based on the generalized likelihood ratio, GLR (Lorden, 1971). Since this ratio depends on the unknown parameters, the change time n_0 and the value of θ_1, the detection algorithm is based on a double maximization

* Correspondence to: G. Capizzi, University of Padua, Via S. Francesco 33, 35100 Padova, Italy
† E-mail: capizzi@stat.unipd.it

Received 25 January 2001
Accepted 20 April 2001

$$g_{na} = \max_{1 \leq j \leq n} \sup_{\theta_1 \in \Theta} S_j^n(\theta_1), \quad n = 1, 2, \ldots \tag{1}$$

where $S_j^n(\theta_1)$ is the log-likelihood ratio for the observations from time j up to time n.

Since the maximization, over $\theta_1 \in \Theta$, is carried out for each possible change time j, between 1 and n, the GLR algorithm can be computationally complex. Moreover, even if in the Gaussian case the double maximization with respect to n_0 and to θ_1 turns out to be explicit, this explicit maximization cannot be obtained in a recursive manner. Modifications of this algorithm are thus of interest. One of the alternative procedures consists of carrying out the maximization of the log-likelihood, over $\theta_1 \in \Theta$, for each possible change time j in a moving window of size M. The underlying idea is to assume that older changes have already been detected or have less and less importance as they get older and older. The 'window-limited' GLR statistic, WGLR, is given by

$$g_{nb} = \max_{n-M+1 \leq j \leq n} \sup_{\theta_1 \in \Theta} S_j^n(\theta_1), \quad n = M, M+1, \ldots. \tag{2}$$

The detection statistic (2) leads to a minimum delay equal to the window size. For stopping when $n - j < M$, this article suggests using a detection rule, denoted by NWGLR, given by

$$g_{nc} = \begin{cases} g_{na} & \text{if } n < M \\ g_{nb} & \text{if } n \geqslant M \end{cases} \tag{3}$$

A lack of control is declared at first n such that the control statistics (1), (2) or (3) are greater than a specified control limit. Performance of the control schemes is usually assessed in terms of the expected number of observations until an alarm is triggered. This function, called average run length (ARL), defines at θ_0 the mean time between false alarms and at θ_1 the mean delay for detection. In the process design one usually tries to derive a scheme that minimizes the mean delay to signal an alarm (out-of-control ARL), given a fixed mean time between false alarms (in-control ARL). As Lai (1995) pointed out, performance of scheme (2) is sensitive to the choice of M and of the other design parameter h. Both values can be chosen on the basis of asymptotical considerations (Lorden, 1971; Lai, 1995). An alternative procedure is to derive a convenient value of the threshold h, conditionally to a fixed value of M (Bordignon and Scagliarini, 2000). Unfortunally, a satisfactory choice of both the window size and the threshold is very difficult in a finite context.

The window size M could be chosen so that, for the same in-control ARL, the most desirable performance overall in terms of out-of-control ARL is obtained. The main idea underlying this article is to estimate sequentially, through the stochastic approximation of Robbins and Monro (1951), the threshold value that satisfies a constraint on the mean time between false alarms, for a given window size. The iterative approximation terminates when the estimate is close to the true threshold with high probability. Since the schemes are designed to have the same in-control ARL, comparisons between the out-of control ARL profiles, for different values of M, allow us to select the more convenient window size. Section 2 illustrates the case of an additive change in the innovation sequence of a linear state-space model. Section 3 describes the design procedure. Section 4 illustrates an application of the design procedure to the ozone concentration data recorded from one monitoring site of Bologna. In Section 5 the ARL's profiles of the three detection schemes are compared.

2. ADDITIVE CHANGES IN LINEAR STATE-SPACE MODELS

Let us assume that an observed sequence $\{Y_n\}$ can be modelled by a linear time-invariant stochastic state-space model

$$\begin{cases} Z_{n+1} = FZ_n + GU_n + W_n \\ Y_n = HZ_n + JU_n + V_n \end{cases} \tag{4}$$

where Z, Y, U are the state, the observation and input vectors, F is the state transition matrix, H the observation matrix, G and J the control matrices and $\{W_n\} \sim \mathrm{N}(0, Q)$, $\{V_n\} \sim \mathrm{N}(0, R)$ are two independent white noise sequences. The change detection problem in a non-independent case can be solved by first using the transformation from observations to innovations.

Given the initial $Z_0 \sim \mathrm{N}(\mu_0, P_0)$, the innovation sequence $\{x_n\}$ can be obtained through the Kalman filter recurtions

$$\begin{cases} \hat{Z}_{n+1} = F\hat{Z}_n + GU_n + FK_n x_n \\ x_n = Y_n - H\hat{Z}_n - JU_n \end{cases} \tag{5}$$

where K_n is the Kalman gain. The recursions (4) and (5) can be re-written so that the innovation sequence does not depend on the input variables, that is

$$\begin{cases} Z_{n+1} - \hat{Z}_{n+1} = (F - FK_n H)(Z_n - \hat{Z}_n) - FK_n V_n + W_n \\ x_n = H(Z_n - \hat{Z}_n) + V_n \end{cases} \tag{6}$$

Additive changes in the state-space model (4) can be expressed in a state-space form as

$$\begin{cases} Z_{n+1} = FZ_n + GU_n + W_n + \Gamma\Upsilon_z(n, n_0) \\ Y_n = HZ_n + JU_n + V_n + \Xi\Upsilon_y(n, n_0) \end{cases} \tag{7}$$

where Γ and Ξ are gain matrices, which account for the change magnitude, while Υ_x and Υ_y are vectors representing the dynamic profile of the changes ($\Upsilon_x(n, n_0) = \Upsilon_y(n, n_0) = 0$, for $n < n_0$). In this article a particular form of (7) is considered: a step-change on the input sequence, given by

$$\begin{cases} Z_{n+1} = FZ_n + GU_n + W_n \\ Y_n = HZ_n + JU_n + V_n + \nu I_{n \geq n_0} \end{cases} \tag{8}$$

where ν is a scalar that corresponds to the unknown change magnitude and $I_{n \geq n_0}$ is the indicator function of the change direction.

It can be shown that a step-change in $\{Y_n\}$ corresponds to a change with dynamic profile on the innovation distribution (Basseville and Nikiforov, 1993). Thus the innovation of the model (8) is of the form

$$x_n = x_n^0 + \nu\rho^*(n, n_0)$$

where x_n^0 is the innovation corresponding to the unchanged model (6) and $\rho^*(n, n_0)$ represents the effect of a change occurring at time $n_0 \leq n$.

The problem addressed in this article is to detect additive changes in the zero mean independent Gaussian sequence $\{x_n\}$. The hypothesis testing problem, in terms of the innovation sequence, is now

$$\begin{array}{lllll} \mathrm{H}_0: & \{x_n\} \sim \mathrm{N}\ (0, \sigma^2) & n = 1, 2, \ldots & & \\ \mathrm{H}_1: & \{x_n\} \sim \mathrm{N}\ (0, \sigma^2) & \text{if} & n \leq n_0 \\ & \{x_n\} \sim \mathrm{N}\ (\nu\rho^*(n, n_0), \sigma^2) & \text{if} & n > n_0 \end{array}$$

The change detection problem, with known θ_0 and dynamic profile of the change, but unknown magnitude, can be solved employing the control statistcs (1), (2) and (3), where $S_j^n(\nu)$ is the log-likelihood ratio of the innovations from x_j to x_n.

The control statistics adapted to the stepwise change are obtained by substituting $\sup_{\theta_1 \in \Theta} S_j^n(\theta_1) = \sup_\nu S_j^n(\nu)$ into (1), (2) and (3), respectively.

Given the Gaussianity of the innovations, we have

$$\sup_\nu S_j^n(\nu) = \hat{\nu}_n(j)\left[\sigma^{-2}\sum_{i=j}^{n}\rho^*(i,j)x_i\right] - \frac{\hat{\nu}_n(j)^2}{2}\left[\sigma^{-2}\sum_{i=j}^{n}\rho^{*2}(i,j)\right]$$

where

$$\hat{\nu}_n(j) = \frac{\sum_{i=j}^n \rho^*(i,j)x_i}{\sum_{i=j}^n \rho^{*2}(i,j)} \tag{9}$$

is the maximum likelihood estimate of the change magnitude at time n, assuming a change at time j. Since, when the steady-state behaviour of the Kalman filter is reached, the signature $\rho^*(n, n_0)$ of the change depends only upon the distance $l = (i - j)$, from the current index i and each supposed change time j, i.e. $\rho^*(1,1) = \ldots = \rho^*(n-1, n-1) = \rho^*(n,n)$, $\rho^*(2,1) = \ldots = \rho^*(n-1, n), \ldots,$ $\rho^*(n-1, 1) = \rho^*(n, 2)$, the computational burden in (1), (2) and (3) can be reduced through the introduction of the function

$$\psi(i-j) = \begin{cases} 1 & \text{if} \quad i - j = 0 \\ 1 - \sum_{s=0}^{i-j-1} HF_\star^s FKI_{i-s-1>0} & \text{if} \quad i - j = 1, 2, \ldots \end{cases}$$

where $F_\star = F(I - KH)$ and K is the steady-state Kalman gain.

Then an explicit expression of $\sup_\nu S_j^n(\nu)$ is given by

$$\sup_\nu S_j^n(\nu) = \frac{\sigma^{-2}[\sum_{i=j}^n \psi(i-j)x_i]^2}{2[\sum_{i=j}^n \psi^2(i-j)]} \tag{10}$$

3. THE DESIGN PROCEDURE

Assume that under the hypothesis of no change: (i) it is possible simulate, $\forall h$, the run lengths

$$RL_a(h^a) = \inf\ \{n : g_{na} > h^a\} \tag{11}$$

$$RL_b(h^b) = \inf\ \{n : g_{nb} > h^b\} \tag{12}$$

$$RL_c(h^c) = \inf\ \{n : g_{nc} > h^c\} \tag{13}$$

with $g_{na},\ g_{nb}$ and g_{nc}, given by (1), (2) and (3), respectively; (ii) $RL_s(h^s) \sim F_{RL_s,h^s}$, with unknown F_{RL_s,h^s}; (iii) the expected value $\mathrm{E}\,(h^s)$ exists; (iv) h^s_* is the unique root of the equation

$$\mathrm{ARL}[0, h^s] - B = 0,$$

where B is a fixed value of the in-control ARL.

Given an initial estimate h^s_1 and a suitable positive constant A, the h^s_* estimate can be sequentially updated through the modified Robbins and Monro process:

$$h^s_{k+1} = h^s_k - \frac{A}{k}\ \bar{n}^s_k, \quad k = 1, 2, \ldots,\ \ s = a, b, c \tag{14}$$

where

$$\bar{n}^s_k = \frac{n_1(h^s_k) + n_2(h^s_k)}{2}$$

is the average of two i.i.d. observations of the random variable $\mathrm{N}\ (h^s_k) = [RL(h^s_k) - B]/B$, with $RL(h^s_k)$ given by (11), (12) and (13), respectively.

The idea is to simulate, for the same value h^s_k, two independent standardized values of the run lengths, so that

$$s^2_k = \frac{\sum_{i=1}^k e^s_i}{k}$$

where $e^s_k = \{n_1(h^s_k) - \bar{n}^s_k\}^2 + \{n_2(h^s_k) - \bar{n}^s_k\}^2$ is an unbiased estimate of $\phi^2 = \mathrm{Var}\{RL(h^s_*)\}$.

Then, a stopping rule, based on the first passage time of an F-statistic, below a fixed boundary w,

$$\tau_q = \inf\left\{k \geq q : u_k(q) = \sum_{i=k-q+1}^{k} \frac{(\bar{n}^s_k)^2}{q s^2_k} < w,\ k = q, q+1, \ldots\right\} \tag{15}$$

is introduced to terminate the iterative estimate (14), (Stroup and Braun, 1982, 1984).

Fixed at a convenient value q, the decision rule, for $k = q, q+1, \ldots,$ is given by

$$\begin{cases} h^s_{k+1} = h^s_k - \frac{A}{k}\ \bar{n}^s_k & \text{if} \quad u_k \geq w \\ \hat{h}^s_* = h^s_k & \text{if} \quad u_k < w \end{cases}$$

It can be shown that, while different choices of A are not crucial, values of q should be chosen between 100 and 500 to have a good balance of the estimate properties and the computing time. Moreover, for large values of q, w can be chosen close to $\mathrm{E}[u_k]$, i.e. $\frac{1}{2}$ (Capizzi and Masarotto,1999).

4. AN APPLICATION TO THE OZONE DATA

The data set consists of the mean hourly average of the ozone, O_3, and hourly nitrogen dioxide concentrations $\{NO_{2_n}\}$ and hourly average temperature values $\{T_n\}$, (see Bordignon and Scagliarini, 2000). The data were measured at one monitoring site of the Bologna urban area network, from June 1993 to December 1996. Using 660 observations, the authors fitted the following linear state-space model:

$$Y_n = [0 \;\; 0 \cdots \cdots 1] Z_n + V_n$$

$$Z_{n+1} = \begin{bmatrix} 0 & 1 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \cdots & 0 & 1 \\ 0.0149 & 0 & 0 & \cdots & -0.3662 & 1.1102 \end{bmatrix} Z_n$$

$$+ \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ 0.3037 & -0.0386 & 0.0636 \end{bmatrix} \begin{bmatrix} T_n^* \\ NO^*_{2_n} \\ NO^*_{2_{n-1}} \end{bmatrix} + W_n$$

where Z_n is the 24×1 state vector, $\{Y_n\} = \log(O_{3_n})$, $\{NO^*_{2_n}\} = \log\,(NO_{2_n})$, $\{T_n^*\} = \log T_n$, V_n a scalar Gaussian white noise with variance $R = 0.0012$, $W_n = [0, \ldots, 0, e_n]'$, a 24×1 Gaussian white noise vector with covariance matrix

$$Q = \begin{bmatrix} 0 & 0 & 0 & \cdots & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \cdots & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0.0185 \end{bmatrix}$$

The $\hat{h}^s_*$ values, for schemes (11), (12) and (13), were just iterated 50 times, for fixed values of B, h_1, A, q, w and M (see Table 1). The mean value of $\hat{h}^s_*$ and the corresponding standard error are listed in Table 2.

Table 1. $\hat{h}^s_*$ values for the GLR, WGLR and NWGLR schemes

	GLR	WGLR	NWGLR
B	250	250	250
h_1	1	1	1
A	1.5	1.5	1.5
q	200	200	200
w	0.5	0.5	0.5
M	∞	4,12,24,48	4,12,24,48

Table 2. Mean values

	GLR M	WGLR M				NWGLR M			
	∞	4	12	24	48	4	12	24	48
h^*	5.657	4.686	5.164	5.295	5.247	4.738	5.273	5.536	5.575
s.e.	0.049	0.040	0.059	0.052	0.042	0.039	0.047	0.047	0.051

5. ARL COMPARISONS

Assume that a change of fixed magnitude, ν, occurred in $n_0 = 0$. For each combination of $\nu = \delta\sigma_{Y|n_0=\infty}$, with $\delta = 0.25, 0.5, 0.75, 1, 1.5, 1.75, 2, 2.5, 3, 3.5, 4, 4.5, 5, 5.5, 6$ and $\sigma_{Y|n_0=\infty} = 0.622$, 3000 values of the out-of-control run lengths were simulated.

Table 3 contains the ARL values for the detection schemes (11), (12) and (13). Examination of these results illustrates that the best out-of-control ARL performance for detecting small shifts in the process is attained for $M = \infty$.

For the other choices of M, the ARL values of the WGLR scheme are larger than the ARLs of the GLR and the NWGLR, for small and large shifts. Moreover the ARL profile of the WGLR scheme shows a minimal delay equal to the moving window size.

On the other hand, if a window-limited approach is performed, the NWGLR scheme (13) seems to be a satisfactory alternative to the WGLR scheme, since it shows improved properties for both small and large shifts. In addition, the ARL values of the NWGLR are smaller than the ARLs of the GLR scheme for detecting large shifts. Finally, the window-size which seems to produce the most desiderable performance is $M = 4$ for the WGLR scheme and $M = 48$ for the NWGLR scheme.

Table 3. Out-of-control ARL values

	GLR M	WGLR M				NWGLR M			
Shift	∞	4	12	24	48	4	12	24	48
0.000	250.00	250.00	250.00	250.00	250.00	250.00	250.00	250.00	250.00
0.155	60.47	156.64	106.06	84.00	86.28	159.43	99.44	78.16	65.40
0.311	21.93	69.44	34.45	32.22	49.52	68.89	30.76	23.67	21.58
0.466	12.82	32.25	17.19	24.64	48.00	30.19	13.86	12.45	12.64
0.622	9.09	15.61	13.07	24.02	48.00	15.68	8.75	8.68	8.99
0.933	5.88	6.86	12.00	24.00	48.00	6.09	5.45	5.71	5.73
1.088	4.97	5.33	12.00	24.00	48.00	4.72	4.67	4.87	4.86
1.244	4.26	4.77	12.00	24.00	48.00	3.92	4.08	4.16	4.23
1.555	3.36	4.20	12.00	24.00	48.00	2.94	3.19	3.31	3.34
1.866	2.77	4.04	12.00	24.00	48.00	2.46	2.62	2.71	2.69
2.177	2.34	4.00	12.00	24.00	48.00	2.16	2.29	2.36	2.33
2.488	2.13	4.00	12.00	24.00	48.00	1.98	2.06	2.10	2.10
2.799	1.99	4.00	12.00	24.00	48.00	1.90	1.95	1.97	1.98
3.110	1.91	4.00	12.00	24.00	48.00	1.85	1.88	1.91	1.90
3.421	1.88	4.00	12.00	24.00	48.00	1.83	1.86	1.87	1.86
3.732	1.86	4.00	12.00	24.00	48.00	1.80	1.86	1.85	1.85

6. SUMMARY AND CONCLUSIONS

The process design of change detection algorithms, based on the generalized likelihood ratio, is discussed. If the run length of these change detection algorithms can be simulated, this article suggests a completely automatic design procedure to estimate the value of the control limit that satisfies a constraint on the mean between false alarms. Then, the out-of control ARL values of the GLR algorithm can be compared to those of window-limited GLR algorithms, for a given in-control ARL and different choices of the window sizes. Although the computational burden in the GLR scheme is reduced by carrying out only a fixed number of maximizations, the corresponding detection algorithm is much less sensitive to small and large shifts. However, an additional component can be included in the stopping rule of the window-limited GLR so that the resulting scheme provides more protection at least against large shifts. Since the NWGLR scheme seems to provide more protection against small shifts in the process as the value of M increases, the performance of these schemes should investigated for a wider range of values of M.

ACKNOWLEDGEMENT

Financial support by MURST is gratefully acknowledged.

REFERENCES

Basseville M, Nikiforov IV. 1993. *Detection of Abrupt Changes: Theory and Applications*. Prentice Hall: Englewood Cliffs.

Bordignon S, Scagliarini M. 2000. Monitoring algorithms for detecting changes in the ozone concentrations. *Environmetrics*. **11**: 125–137.

Capizzi G, Masarotto G. 1999. Calibrazione di una carta di controllo mediante approssimazione stocastica. Atti della XXXIX Riunione Scientifica della Società Italiana di Statistica, Sorrento 14–17 Aprile 1998, in *Supplemento alla Rivista di Scritti di Statistica Economica*, 809–815.

Lai TL. 1995. Sequential change point detection in quality control and dynamic systems, *J. R. Statistical Society B* **57**: 613–658.

Lorden G. 1971. Procedures for reacting to a change in distribution, *The Annals of Mathematical Statistics* **42**: 1897–1908.

Robbins H, Monro S. 1951. A stochastic approximation method, *The Annals of Mathematical Statistics* **22**: 400–407.

Stroup DF, Braun HI. 1982. On a new stopping rule for stochastic approximation, *Z. Wahrsch. Verw. Gebiete* **60**: 535–554.

Stroup DF, Braun HI. 1984. Correction to 'On a new stopping rule for stochastic approximation'. *Z. Wahrsch. Verw. Gebiete* **67**: 237.

