EEST 442 - Revision 3

Comparison of sequential monitoring methods for environmental time series with stochastic cycle

Carlo Grillenzoni

Abstract The paper compares recursive methods for detecting change points in environmental time series. Timely identification of peaks and troughs is important for planning defense actions and preventing risks. We consider linear nonparametric methods, such as time-varying coefficients, double exponential smoothers and prediction error statistics. These methods are often used in surveillance, forecasting and control, and their common features are sequential computation and exponential weighting of data. The innovative approach is to select their coefficients by maximizing the difference between subsequent peaks and troughs detected on past data. We compare the methods with applications to meteorological, astronomical and ecological data, and Monte-Carlo simulations.

Keywords EWMA statistics · Forecast errors · Recursive estimators · Shiryaev-Robert · Varying parameters.

Acknowledgments The author thanks the reviewers for their helpful comments, in particular Referee 1 for detailed remarks.

Department of Planning, University IUAV,
S. Croce 1957, 30135 Venice, Italy
e-mail: carlog@iuav.it

1 Introduction

An important topic in monitoring environmental time series is the detection of turning points, i.e. sequential identification of periods where the slope of a series changes sign. This problem is different from forecasting, nevertheless it is crucial for planning defense actions. For example, in meteorology, knowing the beginning of a period of dry weather leads to rationing water resources, delaying sowing or anticipating harvesting. In urban areas, timely identification of turning points in temperature is important to calibrate heating systems and reduce power consumption. Recent interest in climate changes also urges the correct identification of turning points in the pattern of ecological time series. This is useful for checking size and causes of the changes (e.g. Piao et al. 2011), although data are not easily available.

The study of turning points in cyclical time series was mainly developed in econometrics, where it has followed three approaches. In the first method the series are smoothed with adaptive filters, then first and second difference of the estimated signals are analyzed as the order conditions of continuous functions (e.g. Wildi and Elmer 2008). The second approach is typical of change-point problems and uses sequential tests to detect changes in the mean level. Common statistics are moving averages and cumulative sums (CUSUM, e.g. Manly and MacKenzie 2000), which are equivalent to the likelihood ratio under Gaussianity. The third method estimates models with time-varying parameter (TVP) that are related to the slope of the series. Switching regression models assume a finite number of states, whereas recursive estimators treat continuous parameters (e.g. Ljung 1999).

The main disadvantages of these solutions are as follows: Smoothing methods have problems of accuracy at the borders of the series and delay in the detection (e.g. McAleer and Chan 2006). Change-point methods are suitable for locally stationary processes, but not for cyclical time series; e.g. Chin and Apley (2008) show the detection problems that test statistics have in the presence of nonstationarity. Finally, the efficacy of TVP methods can be hindered by the assumptions made on the parameters. For example, stochastic coefficients and hidden Markov models require complex estimators (e.g. Achcar et al. 2011).

To avoid these problems, this paper deals with sequential exponential methods. In particular, we use the double exponential smoother (DES) to estimate the trend of the series; we focus on the exponentially weighted moving average (EWMA) of forecast errors as control statistic; finally, we apply the exponentially weighted least squares (EWLS) to estimate TVP models. The common feature of these methods is their adaptive and nonparametric nature; however, they deal with different moments of the series, such as mean level and auto-covariances.

Exponential methods involve smoothing parameters that must be properly designed. These coefficients are usually selected with forecasting and control criteria which, however, may not be suitable for point detection; in addition, the sequential monitoring also involves the coefficients of the alarm limits. As in econometrics, this paper jointly selects smoothing and alert coefficients through the maximization of the height difference between detected peaks and troughs (e.g. Bock et al. 2008). Since the maximum difference occurs between the actual turning points, it follows that the proposed solution pursues unbiased detection.

The scheme of the work is as follows: Section 2 introduces models representation and estimation, and defines the detection rules for turning points. Section 3 deals with the selection of smoothing and alert coefficients. Section 4 applies the methods to real and simulated data and compares their performance.

2 Detection models and methods

2.1 Smoothing methods

Given a non-stationary time series X_t, the representation which is used in many smoothing applications is given by

$$X_t = \mu_t + y_t\,, \quad y_t \sim f_y(0, \sigma_t^2)\,, \qquad t = 1, 2 \ldots T, \tag{1}$$

where $\mu_t = \mathrm{E}(X_t)$ is the trend, and y_t is not in general independent and stationary. However, the autocorrelation and heteroskedasticity of y_t do not affect the bias of nonparametric smoothers (see Beran and Feng 2001).

In the econometric literature turning points are usually defined on X_t, or its realizations (see Zellner et al. 1991). However, this approach is problematic, X_t being a stochastic process. Since μ_t is deterministic, it enables a rigorous definition of turning points as local *troughs* (r_i) and local *peaks* (s_i) of the function itself:

$$\begin{aligned} \text{troughs} \quad r_i \;&:\; \mu_{r_i-d} \geq \ldots \geq \mu_{r_i-1} > \mu_{r_i} < \mu_{r_i+1} < \ldots < \mu_{r_i+d}\,, \\ \text{peaks} \quad s_i \;&:\; \mu_{s_i-d} \leq \ldots \leq \mu_{s_i-1} < \mu_{s_i} > \mu_{s_i+1} > \ldots > \mu_{s_i+d}\,, \end{aligned}$$

for some $d \gg 1$. Given the interval [1,T], we assume that the sequence $\{r_i, s_i\}$ contains $n \ll T/2$ pairs, which can be ordered as $1 \leq r_1 < s_1 < r_2 < \ldots < s_n \leq T$. In this sub-section we identify the periods $\{r_i, s_i\}$ by estimating μ_t with smoothing methods, and then finding its local minima and maxima.

The simplest one-sided smoother is the EWMA, which assumes that X_t fluctuates around a constant mean. In the presence of trend components, the DES model provides a suitable estension. Brown (1963 p. 130) defined it as

$$\begin{aligned} \text{simple} \quad \hat{S}_t &= \lambda\, \hat{S}_{t-1} + (1-\lambda)\, X_t\,, \qquad \hat{S}_0 = c_1, \\ \text{double} \quad \hat{\mu}_t &= \lambda\, \hat{\mu}_{t-1} + (1-\lambda)\, \hat{S}_t\,, \qquad \hat{\mu}_0 = c_2, \end{aligned} \tag{2}$$

where $\lambda \in (0, 1]$ is a weighting factor which gives more weight to recent data, and c_1, c_2 are fixed initial conditions.

The algorithm (2) is mainly used in forecasting. As in the linear trend model, the forecast function is the sum of a level and a slope component. In the Brown's approach these components can be estimated from (2) as

$$\begin{aligned} \text{level} \quad \hat{A}_t &= \left(2\,\hat{S}_t - \hat{\mu}_t\right), \\ \text{slope} \quad \hat{B}_t &= \left(\hat{S}_t - \hat{\mu}_t\right)(1-\lambda)/\lambda\,, \\ \hat{X}_{t+k} &= \hat{A}_t + \hat{B}_t\, k\,, \end{aligned} \tag{3}$$

which corresponds to the Holt-Winters algorithm (see Chatfield et al. 2001). As concerned the estimation of the model (1), one can use either $\hat{\mu}_t$ or $\hat{A}_t$, although the latter is less smooth. For the sake of parsimony, the equations of (2) use the same coefficient λ; in Section 3 we will discuss its selection.

The detection of turning points in X_t, can just be defined on the local minima and maxima of the function $\hat{\mu}_t$. Since the estimates are affected by noise, a tolerance value $0 \leq \kappa < \infty$ must be introduced to reduce the number of false alarms. The decision rule, which is necessarily delayed by one lag, then becomes

$$\begin{aligned} \text{troughs} \quad r_i \; : \quad & \Big(\hat{\mu}_{r_i+1} > \hat{\mu}_{r_i} + \kappa \;\Big|\; \hat{\mu}_{r_i} < \hat{\mu}_{r_i-1} - \kappa \Big), \\ \text{peaks} \quad s_i \; : \quad & \Big(\hat{\mu}_{s_i+1} < \hat{\mu}_{s_i} - \kappa \;\Big|\; \hat{\mu}_{s_i} > \hat{\mu}_{s_i-1} + \kappa \Big). \end{aligned} \tag{4}$$

To better understand the meaning of (4), one can rewrite the inequalities in terms of $(\hat{\mu}_t - \hat{\mu}_{t-1})$. With respect to detection rules defined on first and second differences, the advantage of (4) is to emphasize the role of the coefficient $\kappa > 0$, which can hinder false alarms.

2.2 Time-varying parameters

As for smoothing, there is a wide literature on time-varying parameter models (e.g. Grillenzoni 1996 or Ljung 1999). In this paper we consider two semi-parametric schemes which include trend and autoregressive components

$$X_t = \alpha_t + \beta_t \, t + e_{1t} \,, \qquad e_{1t} \sim \text{IN}(0, \sigma^2_{1t}), \tag{5}$$

$$X_t = \phi_t \, X_{t-1} + e_{2t} \,, \qquad e_{2t} \sim \text{IN}(0, \sigma^2_{2t}), \tag{6}$$

their adaptivity can provide residuals which are nearly independent (IN).

In the models (5)-(6), the coefficients $\{\alpha_t, \, \beta_t, \, \phi_t \,, \sigma_t\}$ are deterministic sequences which wander about finite time-average values $|\, \bar{\alpha}, \, \bar{\beta}, \, \bar{\phi}, \, \bar{\sigma} \,| < \infty$, where $\bar{\alpha} = \lim_{T \to \infty} (\, T^{-1} \sum_{t=1}^{T} \, \alpha_t \,)$, etc.. This assumption is called "quasi stationarity" (see Ljung 1999) and allows suitable statistical properties to parameter estimates. If the average values are $\bar{\beta} = 0$, $\bar{\phi} = 1$, then the joint model for X_t is the random-walk plus drift scheme, which is widely used in tests for unit-root (e.g. Fuller 1996).

As for μ_t in the model (1), the parameters of (5)-(6) are non-stochastic, but not necessarily smooth. Therefore, their fluctuations can determine complex patterns in X_t, such as positive and negative trends, local stationarity, structural breaks, and so on. The relationship between varying parameters and turning points derives from

the fact that β_t and ϕ_t determine the local slope of X_t. Thus, if $\beta_t > 0$ or $\phi_t > 1$ we have a positive local trend, whereas a negative trend occurs in the opposite case. This can be checked by investigating the function $\mu_t = \mathrm{E}(X_t)$ of the models (5) and (6), with fixed initial condition $X_0 = c_0$ (a constant)

$$\mu_{1t} = \alpha_t + \beta_t\, t\,, \qquad \mu_{2t} = \prod_{i=1}^{t} \phi_i\, c_0.$$

In the continuous time, differentiation of μ_{1t} gives $\mu'_{1t} = \beta_t$; hence, peaks and troughs occur in correspondence of $\beta_t = 0$. For the model (6), by assuming the step function $\phi_t = \phi_1 > 1$ for $t < \tau$, and $\phi_t = \phi_2 < 1$ elsewhere, the trend function becomes $\mu_{2t} = c_0\, \phi_1^{\tau-1}\, \phi_2^{t-\tau}$, which has a peak at $t = \tau$. Thus, turning points of X_t tend to occur where the parameters $(\beta_t,\, \phi_t)$ cross the thresholds $(0,\, 1)$ respectively. It follows that detection statistics for the peaks and troughs of X_t can be provided by recursive estimates of the regression parameters, namely $(\hat\beta_t,\, \hat\phi_t)$.

Consistently with the semiparametric nature of models (5)-(6), we use the EWLS algorithm. By defining the vectors $\mathbf{z}'_t = [\,1,\, t\,]$ and $\boldsymbol{\theta}'_t = [\,\alpha_t\,,\, \beta_t\,]$, the model (5) can be rewritten as $X_t = \boldsymbol{\theta}'_t\, \mathbf{z}_t + e_t$, and the recursive estimator becomes

$$\begin{aligned}
\hat e_t &= X_t - \hat{\boldsymbol{\theta}}'_{t-1}\mathbf{z}_t\,, \qquad t = 1, 2 \ldots \\
\mathbf{R}_t &= \lambda\, \mathbf{R}_{t-1} + \mathbf{z}_t\, \mathbf{z}'_t\,, \qquad \mathbf{R}_0 = \mathbf{C}_1, \\
\hat{\boldsymbol{\theta}}_t &= \hat{\boldsymbol{\theta}}_{t-1} + \mathbf{R}_t^{-1}\, \mathbf{z}_t\, \hat e_t\,, \qquad \hat{\boldsymbol{\theta}}_0 = \mathbf{c}_2, \\
\hat\sigma_t^2 &= \lambda\, \hat\sigma_{t-1}^2 + (1-\lambda)\, \hat e_t^2\,, \qquad \hat\sigma_0^2 = c_3,
\end{aligned} \tag{7}$$

(see Grillenzoni 1996), where $\lambda \in (0, 1]$, $\hat e_t$ are prediction errors, $\mathbf{R}_t$ is the sum of squared regressors and $\mathbf{C}_1$, $\mathbf{c}_2$, c_3 are fixed initial conditions. The algorithm (7) can also be applied to the model (6) by letting $\mathbf{z}_t = X_{t-1}$ and $\boldsymbol{\theta}_t = \phi_t$.

The initial values $\hat{\boldsymbol{\theta}}_0, \mathbf{R}_0, \hat\sigma_0^2$ of (7) can be estimated with ordinary least squares (OLS) on an initial stretch of data, and are asymptotically negligible when $\lambda < 1$. The properties of EWLS have been well investigated in the constant parameter case, where the consistency holds as $\lambda \to 1$ (see Ljung 1999). For time-varying parameters, λ should be designed according the rate of non-stationarity, as defined by prediction errors and the function $Q_T = \sum_t \hat e_t^2(\lambda)$. Minimizing $Q_T(\lambda)$ is similar to the cross-validation selection of the bandwidth in kernel smoothers.

As stated above, the sequential detection of turning points of X_t can be based on the recursive estimates (7). Specifically, if $\hat{\beta}_t$ *or* $\hat{\phi}_t$ cross their thresholds 0, 1 at time $t = \tau$, then a turning point is detected at τ. Since the estimates are affected by sampling errors, or may cross the thresholds too frequently, tolerance limits $0 \leq (\kappa_1, \kappa_2) < \infty$ must be introduced to reduce the number of weak and false alarms. As in (4) the detection rule becomes

$$\begin{aligned} \text{troughs} \quad r_i \quad &: \quad \left(\hat{\beta}_{r_i} > \kappa_1 \,\middle|\, \hat{\beta}_{r_i-1} < \kappa_1\right) \quad \text{or} \quad \left(\hat{\phi}_{r_i} > 1+\kappa_2 \,\middle|\, \hat{\phi}_{r_i-1} < 1+\kappa_2\right), \quad (8) \\ \text{peaks} \quad s_i \quad &: \quad \left(\hat{\beta}_{s_i} < -\kappa_1 \,\middle|\, \hat{\beta}_{s_i-1} > -\kappa_1\right) \quad \text{or} \quad \left(\hat{\phi}_{s_i} < 1-\kappa_2 \,\middle|\, \hat{\phi}_{s_i-1} > 1-\kappa_2\right), \end{aligned}$$

where we have omitted 0 in the first rule and $\kappa_1 \neq \kappa_2$ for the different scale of β_t and ϕ_t. It is worth noting that the strategy (8) deals with the separate estimation of models (5) and (6), so as to avoid interactions between $\hat{\beta}_t$ and $\hat{\phi}_t$ that would keep them away from the thresholds 0, 1. The main advantage of (8) with respect to (4) is timeliness, because the detection in (4) is delayed at least by one lag.

2.3 Change-point statistics

Monitoring methods based on control statistics arise in change-point problems, where a process X_t is subject to a shift δ in the mean μ at an unknown time $t = \tau$. Extending this approach to the model (1), the mean μ_t becomes a step-function with possible ramps, and y_t is the innovation sequence (e.g. Lai 2001)

$$\begin{aligned} X_t = \mu_t + y_t\,, \qquad \mu_\tau \quad &= \mu_{\tau-1} + \delta, \\ \mu_1 \leq \mu_2 \leq \ldots \leq \mu_{\tau-1} < \quad \mu_\tau \quad &\geq \mu_{\tau+1} \geq \ldots \geq \mu_t\,. \end{aligned}$$

The typical detection strategy is to use test statistics concerned with mean shifts, such as the likelihood ratio (LR). This requires knowledge of the density $f(X_t)$, before and after the change point τ. Under Gaussianity, the optimal LR tends to be equivalent to CUSUM statistics, which are linear functions of the data (e.g. Luo et al. 2012). A serious problem is the autocorrelation of the series, which reduces the power of tests (see Mei 2006). A common remedy is to fit X_t with an autoregressive model and then monitoring its residuals; this coincides with adaptive control techniques used in industrial production (see Box et al. 2009).

As an example, we consider the Shiryaev-Robert (SR) statistic which is based on the sequential LR test. If $f_0(X_t)$ and $f_\delta(X_t)$ are pre-shift and post-shift distributions, the test statistic and its stopping rule are given by

$$R_t(\delta) = \sum_{j=1}^{t} \prod_{i=j}^{t} \frac{f_\delta(X_i)}{f_0(X_i)}, \qquad \hat{\tau} = \min\{\, t : R_t(\delta) > \kappa \,\}.$$

Under mild conditions, this can be extended to dependent series by using conditional densities. In particular, assuming that distributions are continuous and their shape do not depend on the change point, the recursive SR statistic becomes

$$R_t = (1 + R_{t-1})\, \Lambda_t \,, \quad \text{with} \quad \Lambda_t = \frac{f_\delta(\, X_t \,|\, X_{t-1}, X_{t-2} \ldots)}{f_0(\, X_t \,|\, X_{t-1}, X_{t-2} \ldots)}.$$

Further, assuming Gaussian densities with pre-shift mean $\mu = 0$, it can be seen that the SR statistic becomes a function of prediction errors

$$\begin{aligned} e_t &= X_t - \mathrm{E}(\, X_t \,|\, X_{t-1}, X_{t-2} \ldots), \\ R_t &= (1 + R_{t-1})\, \exp\left(\delta\, e_t/\sigma - \delta^2/2\right), \end{aligned}$$

which is linearizable with logarithm. Apart from the a-priori selection of the shift δ, a drawback of LR tests is that they assume only positive values; therefore, they do not enable to identify the type (peak or trough) of a change.

Linear statistics of e_t do not have these problems since they preserve the sign of errors. It can be easily seen that positive errors occur in correspondence of troughs, whereas negative e_t take place after a peak. Since turning points generate patches of errors, CUSUM and EWMA statistics of e_t are more robust indicators. Prediction errors can be estimated efficiently by merging the models (5) and (6) as

$$X_t = \alpha_t + \beta_t\, t + \phi_t\, X_{t-1} + e_t \,, \qquad e_t \sim \mathrm{IN}(0, \sigma_t^2), \tag{9}$$

and using the algorithm (7) with $\mathbf{z}_t' = [\,1,\, t,\, X_{t-1}]$. To reduce false alarms caused by the heteroskedasticity, the EWMA statistic must use standardized errors as

$$\hat{Z}_t = \lambda\, \hat{Z}_{t-1} + (1 - \lambda)\, (\hat{e}_t/\hat{\sigma}_{t-1}) \,, \qquad \hat{Z}_0 = 0. \tag{10}$$

This expression can be inserted in the algorithm (7), by using the same smoothing coefficient λ. Given the relationship between prediction errors and turning points,

the detection rule based on the tolerance limit $\kappa > 0$ becomes

$$\begin{aligned} \text{troughs} \quad r_i \quad &: \quad \Big(\hat{Z}_{r_i} > +\kappa \;\Big|\; \hat{Z}_{r_i-1} < +\kappa\Big), \\ \text{peaks} \quad s_i \quad &: \quad \Big(\hat{Z}_{s_i} < -\kappa \;\Big|\; \hat{Z}_{s_i-1} > -\kappa\Big). \end{aligned} \tag{11}$$

As an alternative to (10) one can use the Shewhart statistic $\hat{Z}_t^1 = \hat{e}_t/\hat{\sigma}_{t-1}$, which monitors individual errors. Vander Wiel (1996) showed that in random-walk processes it has the same signaling performance as optimal statistics. In particular, the average run length (ARL), which measures the expected number of periods between the change point and the alarm signal: $\mathrm{E}(\,\hat{\tau} - \tau \,|\, \delta\,)$, is similar for Shewhart, CUSUM, EWMA and LR, for any value of the shift $\delta \neq 0$. In summary, for non-stationary series the crucial issue is *not* the type of monitoring statistic, rather the kind of model used to generate prediction errors. Trend models as (1) are rigid and ignore autocorrelation; instead, regression models as (9) are adaptive.

3 Design of tuning coefficients

In the previous sections we have discussed various approaches to detection. Their statistics and decision rules contain the coefficients λ and κ that must be properly designed. Traditional selection follows operative criteria; for example, in the exponential smoother (3), λ has to optimize the forecasting performance, and in the recursive estimator (7), λ must minimize $\sum_t \hat{e}_t^2$. In LR-SR statistics, the mean shift δ is selected on the basis of the normal wandering of X_t, and the threshold κ has to attain the desired ARL level. Typically, $\delta \in [.25\,\sigma_e; 1.5\,\sigma_e]$ (hence $\delta \approx 1$ if e_t are standardized) and $\mathrm{ARL}(\kappa \,|\, \delta)$ is computed with simulations (see Chin and Apley 2008). It is difficult, however, to extend this approach to κ of the schemes (4),(8) and (11) because the path of the trend function μ_t and of its turning points are complex. In the following, we present a data-driven approach to the joint selection of λ and κ which is adopted in econometrics (see Grillenzoni 2012).

3.1 Maximum amplitude selection

As in Section 2.1, we assume that the function μ_t has n pairs of troughs/peaks

$\{ r_i, s_i\}$ in the interval $[1, T]$, where $\mu_{r_i} < \mu_{s_i}$ and $r_i < s_i$ for $i = 1, 2 \ldots n$. We also assume that the realizations of X_t have turning points which are close to $\{ r_i, s_i\}$. This means that the variance of the detrended process y_t is small compared to that of X_t. We define the *total fluctuation* of X_t in the interval $[1, T]$ as the sum of the differences in height between subsequent peaks and troughs, that is

$$D_T(\mathbf{r}, \mathbf{s}) = \sum_{i=1}^{n} \left(X_{s_i} - X_{r_i} \right), \qquad n \ll T/2. \tag{12}$$

Its expected value is $\mathrm{E}(D_T) = \sum_i (\mu_{s_i} - \mu_{r_i}) > 0$, and is maximum on [1,T].

Let $\hat{r}_i$ and $\hat{s}_i$ be the turning points identified with the statistics $\hat{\mu}_t, \hat{\beta}_t$ and $\hat{Z}_t$ and the decision rules (4),(8) and (11). They depend on the design coefficients λ and κ by maximizing the function D_T, that is

$$(\hat{\lambda}, \hat{\kappa}) = \arg \max_{\lambda, \kappa} D_T \Big[\hat{\mathbf{r}}(\lambda, \kappa), \, \hat{\mathbf{s}}(\lambda, \kappa) \Big]. \tag{13}$$

In econometrics, this approach corresponds to a strategy of maximum profitability in a sequence of buy and sell actions. In electrical engineering it provides the maximum gain to an amplified signal. The important fact is that (13) is automatic and allows for timely, hence minimum delay, detection of turning points.

The solution (13) may yield an excessive number of turning points, i.e. a value of n much greater than that expected from the visual analysis of the series. Thus, it may be preferable to optimize the mean value D_T/n, or the *penalized* function $P_T(\lambda, \kappa) = \Big[D_T(\lambda, \kappa) - \gamma \, n \Big]$, where $0 \leq \gamma < \infty$ is selected so that the resulting $\hat{n}$ is nearly equal to the number n^* of peaks of X_t observed on $[1, T]$. Since the choice of γ is not simple, one can build the objective function only on the first $n^* < n$ greatest differences $D_i = (Y_{s_i} - Y_{r_i})$, and maximize the partial sum

$$D_T^*(\lambda, \kappa) = \sum_{i=1}^{n^*} D_{[i]}(\lambda, \kappa) \, , \qquad D_{[i]} \geq D_{[i+1]}, \tag{14}$$

where n^* is the number peaks (troughs) observed, or selected, on $[1, T]$.

3.2 Computational aspects

The functions D_T and D_T^* may be non-smooth and may have several local maxima. It is possible to identify the global optima by exploring their surface on a grid of

values for $(\lambda\,,\kappa)$; subsequently, numerical optimization can refine the solution. For testing purposes, the available data-set is split into two segments $T = T_1 + T_2$, where T_1 is the in-sample (or estimation) period and T_2 is the out-of-sample (or forecasting) period. We carry out the selection (13) only on the first T_1 observations; next we compute the measure D_{T_2} on the remaining T_2 data.

Algorithms (2),(7) and (10) are recursive and require initial values; these influence the performance of the methods at the beginning. We solve this problem by assigning reasonable values to $\hat{S}_0, \hat{\mu}_0$; $\hat{\boldsymbol{\theta}}_0, \mathbf{R}_0, \hat{\sigma}_0^2$ and $\hat{Z}_0$, and by adjusting them on an initial stretch of data. Specifically, we add at the beginning of the series, a sub-sample of size $N \ll T_1$, just rescaled to the level of the first observation:

$$X^*_{-j} = X_{N-j} - \Big(X_N - X_1\Big)\,, \qquad j = 1, 2 \ldots N \ll T_1\,. \tag{15}$$

The size of N is not very important for (2), but for (7) it should be large enough.

We can summarize the detection strategy with the following steps:

1) Define the interval $T_1 \ll T$ and the grid of values $\{\lambda_i\}_{i=1}^{n_1}$ and $\{\kappa_j\}_{j=1}^{n_2}$;
2) Run the algorithms of $\hat{\mu}_t$, $\hat{\beta}_t, \hat{\phi}_t, \hat{e}_t$ and $\hat{Z}_t$ for every $\lambda_i,\ i = 1, 2, ..., n_1$;
3) Apply the detection rules (4),(8) and (11) for every $\kappa_j,\ j = 1, 2, ..., n_2$;
4) Sort the sequences of detected points $\{\hat{r}_k\}$, $\{\hat{s}_h\}$, to obtain $\{\hat{r}_l < \hat{s}_l\}$;
5) Compute the function $D_{T_1}(\lambda_i, \kappa_j)$ and find its maximum (λ_0, κ_0);
6) Improve the values of (λ_0, κ_0) with numerical algorithms on D_{T_1};
7) Evaluate the out-of-sample statistic $D_{T_2}(\hat{\lambda}_0, \hat{\kappa}_0)$, on $T_2 = T - T_1$.

As regards Step 4, it should be noted that in the presence of two or more consecutive troughs (peaks), only the first one is retained for reasons of timeliness. However, the criterion $\max D_{T_1}$ will finally select the most effective.

4 Applications to real and simulated data

4.1 Barometric pressure

The first case study deals with the barometric pressure, a key factor in the behavior of meteorological phenomena. It influences temperature, humidity, wind speed and rainfall; hence, it also determines the quality of air in urban areas. The process has

a cyclical pattern with stochastic amplitude and period.

Data come from the station of Chicago Meigs (`www.soils.wisc.edu`) consist of hourly averages for the period Jan. 1 - Feb. 11, 2010, a total of T=1000 time points. Visual inspection of the series shows about 12 peaks on the whole period, see Fig. 1c. Letting T_1=700 the in-sample period, Fig. 1a and Fig. 1b show the functions (13) and (14) with n_1^*=9, for the method (4) based on DES. Since the optimum point of Fig. 1a yields a large number of turning points (n=63), we consider the maximum of Fig. 1b, which is (λ, κ)=(0.754,0.0096) and yields n_1=8 and D_{T_1}=83. The implied estimates $\hat{\mu}_t$ and the peaks are displayed in Fig. 1c. The performance on T_1 and T_2 is relatively good, although two main peaks are missed.

Applying the TVP rule (8) to the AR model (6) and the function (14) with n^*=9, we obtain (λ, κ)=(0.695,0.00028) which yield n_1=8 and D_{T_1}=89. Fig. 2 shows recursive estimates $\hat{\phi}_t$ and detected points; the results are significantly better than those in Fig. 1c. The performance of the regression model (5) is slightly inferior, especially in the out-of-sample period, and is not reported here.

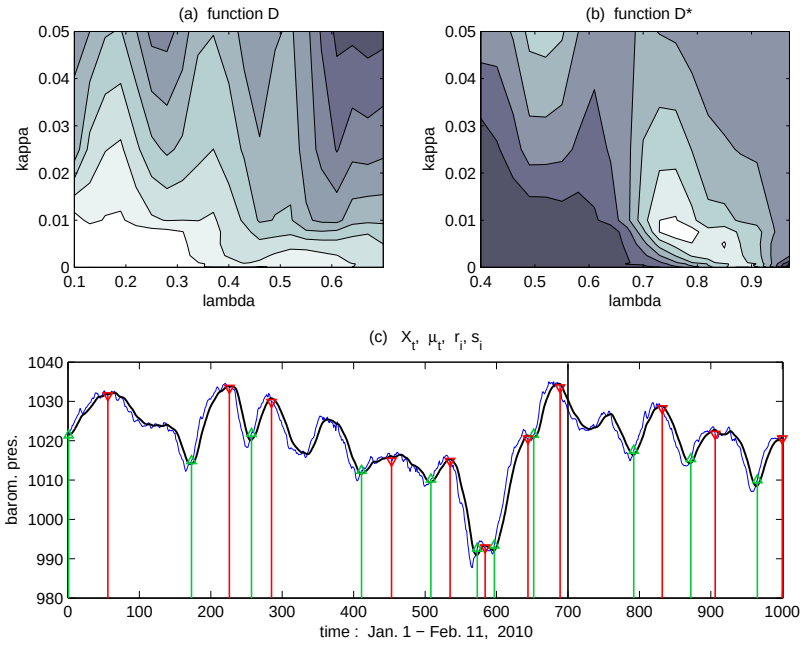

Fig. 1 Application of the method (4) to barometric pressure data: (a),(b) Surfaces of (12) and (14) with n^*=9, evaluated on the sub-sample T_1=700. (c) Series X_t (solid blue), trend estimate $\hat{\mu}_t$ (black), troughs r_i ($\triangle$ green) and peaks s_i ($\triangledown$ red).

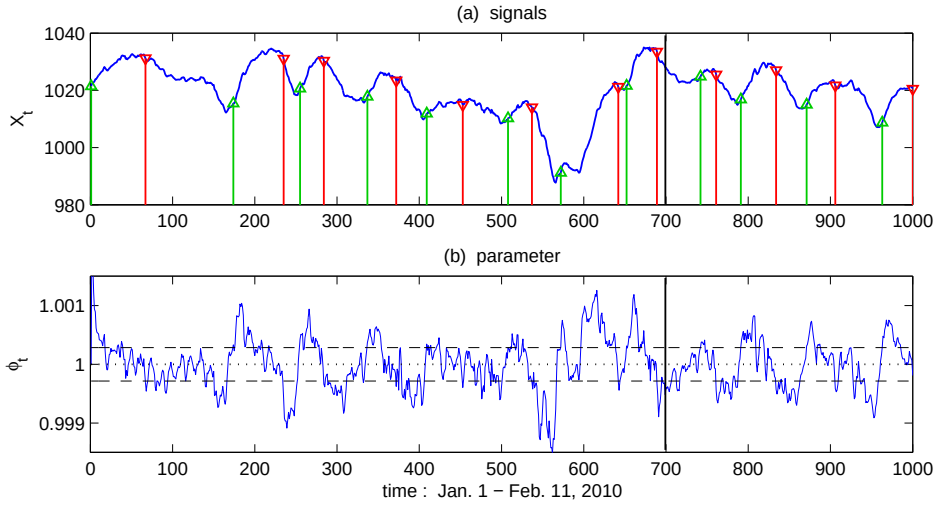

Fig. 2 Application of the method (8) to the hourly barometric pressure: (a) Series X_t (solid blue), detected troughs r_i ($\triangle$ green) and peaks s_i ($\triangledown$ red); (b) Recursive estimates $\hat{\phi}_t$ (solid) and alarm limits $1\pm\kappa$ (dashed).

Finally, we apply the method (11) based on EWMA statistics of prediction errors. The best model for generating white-noise $\hat{e}_t$ is the autoregression (6). Independence is a necessary condition for unbiased detection (e.g. Mei 2006). Selection with criterion (14) provides (λ, κ)=(0.825,0.494), n_1=9 and D_{T_1}=91. As the estimates are not very accurate at the beginning, the detection of the first peak is wrong, but the out-of-sample performance is very good.

Table 1 summarizes main numerical results of the methods discussed so far. We can state that the best method on T_1 is TVP, as it detects the right number of peaks n_1=8, whereas EWMA of prediction errors is the best on T_2.

Table 1 Results of detection methods applied to barometric data: λ, κ are smoothing and alarm coefficients; T_1=700, T_2=300 are in- and out-of-sample periods; D_T is the sum of level difference between peaks and troughs; n is the number of peaks.

Method	Model	Stat.	Rule	$\hat{\lambda}$	$\hat{\kappa}$	D_{T_1}	n_1	D_{T_2}	n_2
DES	(2)	$\hat{\mu}_t$	(4)	0.754	0.0096	83	8	27	3
TVP	(6)	$\hat{\phi}_t$	(8)	0.695	0.00028	**89**	**8**	30	4
EWMA	(6)	$\hat{Z}_t$	(11)	0.825	0.494	91	9	**34**	**4**

4.2 Sunspot number

Sunspot number is the daily count of dark spots on the sun surface which is visible from earth. Sunspots are caused by intense magnetic activity which locally inhibits convection and reduces temperature; hence, their number is an indicator of the solar activity. The series has been collected for about 300 years; despite its randomness, it shows a cyclical pattern with an average period of about 10.5 years. We consider the monthly average from Jan. 1924 to Dec. 2010, a total of T=1032 observations (`http://sidc.oma.be`), and compare the detection methods.

Performance of the smoother (4) is poor as the troughs and peaks are identified with considerable delay. We improve the method with the so-called *oscillation* approach used in the technical analysis of finance. It consists of monitoring crossings of one-sided moving averages of different size: if a short-term average exceeds the long-term one, then a trough is detected. In this paper, we just compare simple and double smoothers $(\hat{S}_t,\, \hat{\mu}_t)$ of the system (2), by following the rule

$$\begin{aligned} \text{troughs} \quad r_i \quad &: \quad \hat{S}_{r_i} > \hat{\mu}_{r_i} + \kappa\,, \\ \text{peaks} \quad s_i \quad &: \quad \hat{S}_{s_i} < \hat{\mu}_{s_i} - \kappa\,, \end{aligned} \tag{16}$$

where κ is a tolerance value. Functions (13) and (14) with n_1^*=5, provide similar results for the selection of the coefficients λ and κ using the first T_1=600 observations. Numerical results are reported in Table 2 and graphs are shown in Fig. 3.

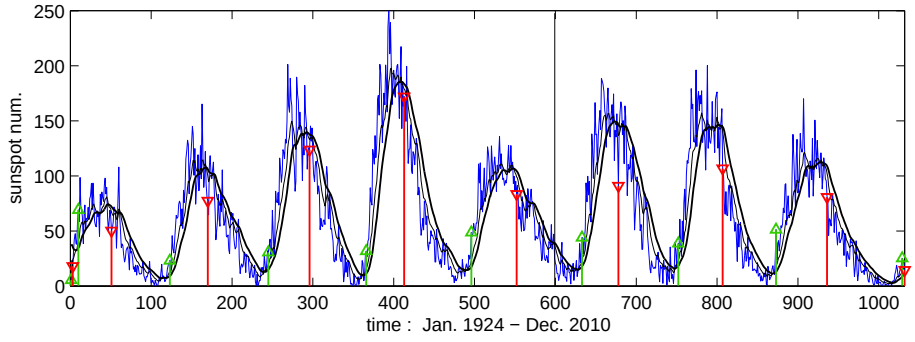

Fig. 3 Application of the method (16) to the monthly sunspot number: Data X_t (blue), estimates $\hat{S}_t$ (black thin), $\hat{\mu}_t$ (black bold), detected troughs r_i ($\triangle$ green) and peaks s_i ($\triangledown$ red).

As regards time-varying parameters, the method based on the model (6) cannot be applied because X_t does not contain a unit-root; hence $\hat{\phi}_t$ cannot be compared with the threshold 1. Instead, the regression model (5) with the rule (8) has provided results which improve those of (16); see the second row of Table 2.

In applying the method (11), we face two problems: the first is the choice of the model for X_t, the second is the inertia of the EWMA statistic. Modeling the sunspot series is a challenging issue; however, we have checked that the mixed model (9) provides acceptable results. Since its innovations are mildly autocorrelated, we solve the loss of efficiency by resetting the statistic to zero whenever it exceeds the alarm limits. Using the indicator function $I(\cdot)$ the modified EWMA statistic is

$$\tilde{Z}_t = \lambda\, \tilde{Z}_{t-1}\, I\Big(|\tilde{Z}_{t-1}| < \kappa\Big) + (1-\lambda)\, \hat{e}_t/\hat{\sigma}_{t-1}\,. \tag{17}$$

The application of (17) to the sunspot series provides the results in Fig. 4 and Table 2. They significantly outperform the others in terms of the statistics D_T.

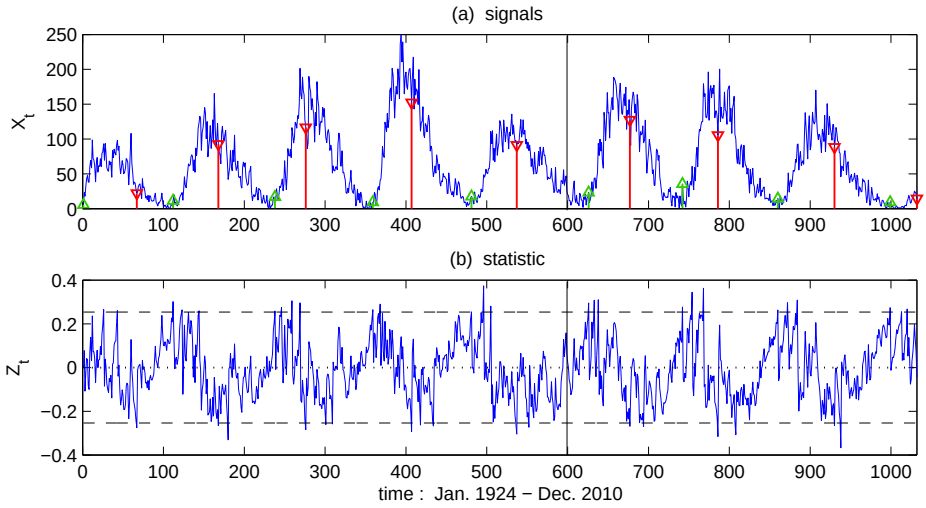

Fig. 4 Application of the method (17)-(11) to monthly sunspot numbers: (a) Data X_t (solid blue) and detected troughs r_i ($\triangle$ green) and peaks s_i ($\triangledown$ red); (b) Statistics (17) $\tilde{Z}_t$ (solid) and alarm bands $\pm\kappa$ (dashed).

Table 2 Results of detection methods applied to the sunspot data. Symbols are explained in the caption of Table 1, except for T_1=600 and T_2=432.

Method	Model	Stat.	Rule	$\hat{\lambda}$	$\hat{\kappa}$	D_{T_1}	n_1	D_{T_2}	n_2
DES	(2)	$\hat{S}_t - \hat{\mu}_t$	(16)	0.85	4.1	316	6	134	4
TVP	(5)	$\hat{\beta}_t$	(8)	0.855	0.63	367	5	158	4
EWMA	(9)	$\tilde{Z}_t$	(11)	0.93	0.195	**421**	**5**	**271**	**4**

4.3 Cosmic rays

Cosmic rays are energy charged subatomic particles originating from the outer space. As atomic radiations, they are dangerous for life forms, but fortunately, they are largely deflected by the magnetic field of earth. Their flux to the earth surface also depends on the solar wind, i.e. the magnetized plasma generated by the sun, which decelerates the incoming particles. As for sunspots, the intensity of solar wind has an average period of 10-11 years; hence, cosmic rays enjoy the inverse cycle.

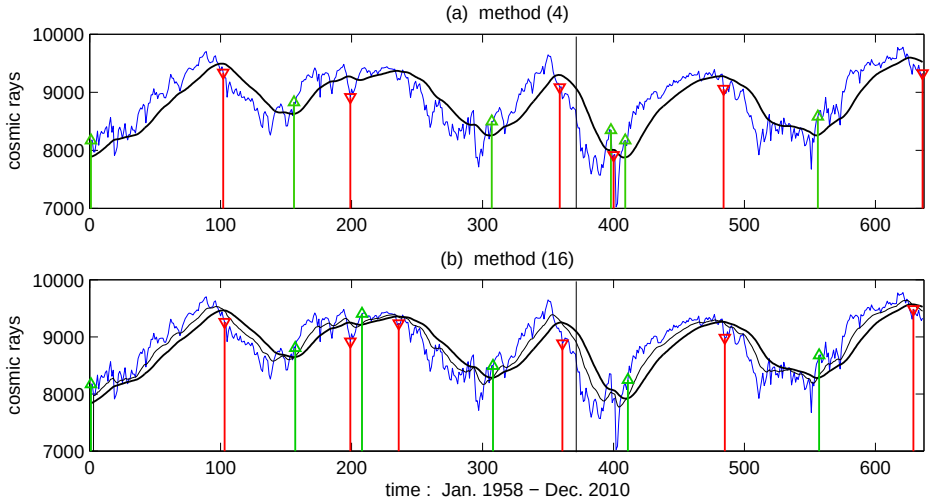

Fig. 5 Application of the methods (4) and (16) to cosmic ray data: (a) Series X_t (solid blue), estimates $\hat{\mu}_t$ (black) and detected troughs r_i ($\triangle$ green) and peaks s_i ($\triangledown$ red); (b) Series X_t (blue), estimates $\hat{S}_t$ (black thin), $\hat{\mu}_t$ (black bold).

We consider the monthly average collected in Moscow from Jan. 1958 to Dec. 2010, a total of T=636 observations (`www.climate4you.com`). In applying detection

methods, the training period for computing the functions (13) and (14) with n_1^*=3, is defined as T_1=372. Fig. 5 shows the results provided by the smoothing methods (4) and (16); apart from a couple of false alarms they are acceptable. Results of the regression methods (8) and (11) are slightly inferior and are summarized in Table 3. The bad out-of-sample performance of EWMA may be caused by outliers and jumps which are present in the series X_t and strongly affect $\hat{Z}_t$ when $\lambda < 0.9$.

Table 3 Results of detection methods applied to cosmic rays data. Symbols are explained in the caption of Table 1, except for T_1=372 and T_2=264.

Method	Model	Stat.	Rule	$\hat{\lambda}$	$\hat{\kappa}$	D_{T_1}	n_1	D_{T_2}	n_2
DES	(2)	$\hat{\mu}_t$	(4)	0.872	0.261	**1841**	**3**	1203	3
DES	(2)	$\hat{S}_t - \hat{\mu}_t$	(16)	0.890	6.41	1425	4	**1541**	**2**
TVP	(5)	$\hat{\beta}_t$	(8)	0.880	2.70	1405	4	1367	2
TVP	(6)	$\hat{\phi}_t$	(8)	0.913	0.0022	1165	3	1196	3
EWMA	(9)	$\tilde{Z}_t$	(11)	0.873	0.321	1404	3	845	3

4.4 Ocean temperature

El Niño southern oscillation (ENSO) is the cyclical behavior of sea surface temperature in the south-tropical part of eastern Pacific Ocean. Given the size of the area, its anomalies are held responsible for extreme climate events around the world, including local droughts and floods. ENSO period and amplitude are also indicators of global climate changes (`www.esrl.noaa.gov/psd/enso/mei`).

We consider the monthly variation of temperature from Jan. 1951 to Dec. 2010, a total of T=720 observations. Training period is defined as T_1=450; given the numerous turning points the best criterion for selecting the coefficients is the function (14) with n_1^*=11. The best detection method is EWMA of the prediction errors of model (9); the results are displayed in Fig. 6 and Table 4.

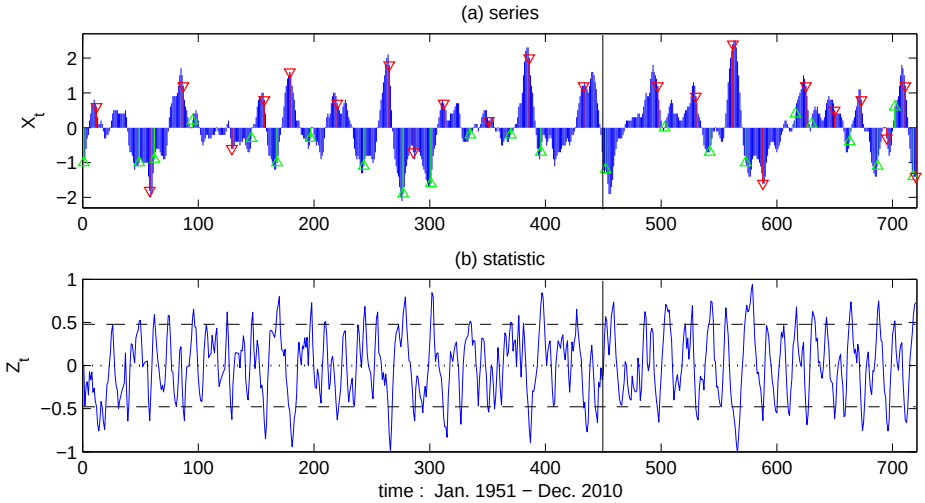

Fig. 6 Application of the method (11) to the ocean temperature variation: (a) Series X_t (solid blue) and detected troughs r_i ($\triangle$ green) and peaks s_i ($\triangledown$ red); (b) Statistics (11) $\hat{Z}_t$ (solid) and alarm bands $\pm\kappa$ (dashed).

Table 4 Results of detection methods for the ocean temperature variation. Symbols are explained in the caption of Table 1, except for T_1=450 and T_2=270.

Method	Model	Stat.	Rule	$\hat{\lambda}$	$\hat{\kappa}$	D_{T_1}	n_1	D_{T_2}	n_2
DES	(2)	$\hat{\mu}_t$	(4)	0.537	0.0091	12.2	13	7.1	6
TVP	(5)	$\hat{\beta}_t$	(8)	0.119	0.203	11.1	10	7.2	7
TVP	(6)	$\hat{\phi}_t$	(8)	0.088	0.069	13.2	13	8.1	9
EWMA	(9)	$\hat{Z}_t$	(11)	0.776	0.478	**17.8**	**13**	**9.7**	**10**

4.5 Simulation experiments

In-depth evaluation of the detection methods requires simulation experiments. We consider a process with a deterministic trend-cycle which is blurred by a random walk; namely $X_t = \mu_t + \omega_t + y_t$, where $\mu_t = 1 + 0.01\,t + 0.0000075\,t^2$ is the trend, $\omega_t = 1 + \sin(t/35)$ is the cycle, and $y_t = y_{t-1} + a_t$ with $a_t \sim \text{IN}(0, 0.05)$ Gaussian. We generate N=500 realizations of length T=1000 (an example is given in Fig. 7), and we apply the procedure of Section 3.2 with T_1=600 and T_2=400.

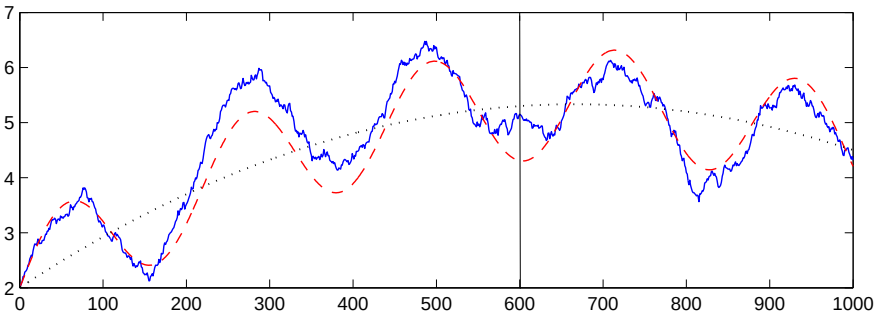

Fig. 7 Typical realization of the simulated process: trend μ_t (dotted), cycle ω_t (dashed), series X_t (solid).

Table 5 Results of detection methods applied to simulated data. The statistics are average values of the estimates over N=500 replications; T_1=600, T_2=400 and d is the mean detection delay over the turning points of μ_t.

Method	Model	Stat.	Rule	$\bar{\lambda}$	$\bar{\kappa}$	$\bar{D}_{T_1}$	$\bar{n}_1$	$\bar{d}_1$	$\bar{D}_{T_2}$	$\bar{n}_2$	$\bar{d}_2$
DES	(2)	$\hat{\mu}_t$	(4)	0.846	0.0005	5.21	2.02	40.6	1.27	1.42	42.1
DES	(2)	$\hat{S}_t - \hat{\mu}_t$	(16)	0.821	0.0286	5.87	2.81	16.2	2.86	2.38	16.4
TVP	(5)	$\hat{\beta}_t$	(8)	0.785	0.0095	**5.98**	**2.85**	**14.3**	**2.98**	**2.44**	**14.7**
TVP	(6)	$\hat{\phi}_t$	(8)	0.833	0.0114	5.32	3.10	18.2	2.11	2.02	26.0
EWMA	(9)	$\hat{Z}_t$	(11)	0.926	0.243	5.87	3.09	29.3	2.79	1.98	31.5

The coefficients (λ, κ) are selected with the criterion function D_{T_1}/n_1 because it provides the best performance overall. Average values of the estimates are reported in Table 5. It includes the mean detection delay $d = (2\,n)^{-1} \sum_{i=1}^{n} [(\hat{r}_i - r_i) + (\hat{s}_i - s_i)]$, where r_i and s_i are the turning points of $(\mu_t + \omega_t)$. On the basis of these results, the best detection method is (8), based on the TVP model (5), because it has max D and min d. The worst one is DES (4) in view of the biggest values of d.

5 Conclusions

In this article we have compared detection methods based on exponential weighting of observations, and applied them to various environmental data sets, such as mete-

orological, astronomical, climatological and ecological. Suitable evaluation approach is found to be the one based on out-of-sample statistics n_2 and D_{T_2}; i.e. the number of detected peaks and the level difference between paired peaks and troughs. The target value for n_2 is the number of peaks observable in the period T_2; however, in time series with stochastic cycles, definition and count of turning points may be difficult (see Fig. 6a). Indicator D_{T_2} is preferable because it deals with the location of the points and, therefore, with the unbiasedness of the methods.

Evaluation of the statistics D_{T_2} in Tables 1-5 does not allow one to conclude about the overall best method. EWMA approach outperforms the others in 3 cases, and could be further improved by increasing the order of the model (9), e.g. by including terms as X_{t-k}, $k > 1$. An important contribution to the performance of EWMA is given by the adaptive estimator (7); especially for the treatment of the heteroscedasticity with $\hat{e}_t/\hat{\sigma}_{t-1}$. In our framework, EWMA solution is intrinsically related to TVP methods, which have a reasonably good performance.

The attempt to improve the results by combining various methods may be interesting, but difficult to follow. In principle, the different sequences $\{\hat{r}_i, \hat{s}_i\}_j$ (where j is the index of methods) could be pooled and ordered; however, this surely increases the number of detections n, but may not improve the statistics D_T. Alternatively, one may simply use the best methods in *parallel*, i.e. take a decision only if it is signaled by at least two methods. This approach, however, tends to drastically reduce the number of detections.

The fundamental step in the proposed methodology is the selection of smoothing and alarm coefficients. This is carried out by maximizing $D_{T_1}(\lambda, \kappa)$ on the initial period T_1. An important aspect which needs further investigation is the stability of the selected coefficients in the subsequent period T_2. In situations of fast evolution and structural changes (such as those induced by volcano eruptions or wildfires), the coefficients should be updated as and when new observations become available. At the same time, old data should be discarded from the estimation sample and suitable design of T_1 must be addressed.

References

Achcar JA, Rodrigues ER, Tzintzun G (2011) Using stochastic volatility models to analyse weekly ozone averages in Mexico City. Environ Ecol Stat 18:271-290

Beran J, Feng Y (2001) Local polynomial estimation with a FARIMA-GARCH error process. Bernoulli 7:733-750

Bock D, Andersson E, Frisén M (2008) The relation between statistical surveillance and technical analysis in finance. In: Frisén M, Financial Surveillance. Wiley, London, pp. 69-92

Box GEP, Luceño A, Del Carmen Paniagua-Quinones M (2009) Statistical Control by Monitoring and Adjustment, 2nd Edition. Wiley, New York

Brown RG (1963) Smoothing, forecasting and prediction of discrete time series. Prentice-Hall, Englewood Cliffs NJ

Chatfield C, Koehler B, Ord K, Snyder D (2001) A new look at models for exponential smoothing. The Statistician 50:147-159

Chin CH, Apley DW (2008) Performance and robustness of control charting methods for autocorrelated data. J Korean Inst Ind Eng 34:122-139

Fuller WA (1996) Introduction to statistical times series. Wiley, New York

Grillenzoni C (1996) Testing for causality in real time. J Econometrics 73:355-376

Grillenzoni C (2012) Sequential estimation and control of time-varying unit-root processes. Seq Analysis 31:22-39

Lai TL (2001) Sequential analysis: some classical problems and new challenges. Stat Sinica 11:303-408

Ljung L (1999) System identification: theory for the user. Prentice Hall, Upper Saddle River NJ

Luo P, DeVol TA, Sharp JL (2012) CUSUM analyses of time-interval data for online radiation monitoring. Health Physics 102:637645

Manly BFJ, MacKenzie D (2000) A cumulative sum type method for environmental monitoring. Environmetrics 11:151-166

McAleer M, Chan F (2006) Modelling trends and volatility in atmospheric carbon dioxide concentrations. Environ Modell Softw 21:1273-1279

Mei Y (2006) Suboptimal properties of Page's CUSUM and Shiryayev-Roberts procedures with dependent observations. Stat Sinica 16:883-897

Piao S, Wang X, Ciais P, Zhu B, Wang T, Liu J (2011) Changes in satellite-derived vegetation growth trend in temperate and boreal Eurasia from 1982 to 2006. Global Change Biology 1-12

Vander Wiel SA (1996) Monitoring processes that wander using integrated moving average models. Technometrics 38:139-151

Wildi M, Elmer S (2008) Real-time filtering and turning-point detection. Available at the URL www.idp.zhaw.ch

Zellner A, Hong C, Min C-K (1991) Forecasting turning points in international output growth rates. J Econometrics 49:275-304