Biometrical Journal **46** (2004) 1, 90–102 DOI: 10.1002/bimj.200410007

On the Online Detection of Monotonic Trends in Time Series

R. Fried[*,1] and **M. Imhoff**[2]

[1] Department of Statistics, University of Dortmund, 44221 Dortmund, Germany
[2] Surgical Department, Hospital of Dortmund, 44137 Dortmund, Germany

Received December 2002, revised August 2003, accepted October 2003

Summary

The online detection of a monotonic trend in a time series with a time-varying mean is an important task in medical applications like intensive care monitoring, that is rendered difficult by autocorrelations. Statistical control charts designed for industrial processes are not adequate as they typically rely on a fixed target value, and many detection rules assume a trend to be linear or neglect autocorrelations. We report our experience with the online detection of slow monotonic trends. Our approach is based on a moving time window, and time-varying autocorrelations are estimated online using parametric assumptions. The performance of versions of this approach is investigated in a simulation study. We find that shrinkage estimation of the time-varying mean improves the results.

Key words: Statistical process control; Online monitoring; Change-point detection; Autocorrelation; Shrinkage estimation.

1 Introduction

In many modern applications of statistical data analysis subsequently measured observations need to be analyzed online. The fast and reliable detection of patterns of change in e.g. medical or environmental time series is important since such patterns point at some change in the data generating mechanism. In intensive care for instance, early detection of monotonic trends in physiological time series allows the physician to take some therapeutical intervention before a critical threshold is exceeded.

Many rules for trend detection rely on independent observations, although many time series exhibit strong, possibly time-varying autocorrelations (Endresen and Hill, 1977). Positive autocorrelations cause monotonic sequences in the data similar to deterministic trends. Vice versa, deterministic trends do not only affect the process mean, but they also strongly influence sample autocorrelations, which are needed to standardize test statistics for trend detection. In retrospective analysis often a simple linear trend is fitted for detrending the data and then the autocorrelations are approximated by an autoregressive (AR) model for the residuals (Cochrane and Orcutt, 1949, Bloomfield and Nychka, 1992). AR models constitute a quite flexible model class describing a wide variety of autocorrelation functions and simple algorithms for model fitting exist.

For online trend detection, control charts based on exponential smoothing are frequently recommended (Trigg, 1964; Cembrowski et al., 1975; Endresen and Hill, 1977; Montgomery and Mastrangelo, 1991; Schack and Grieszbach, 1994). However, for any choice of the weighting parameter there are scenarios where the resulting EWMA rule shows poor performance. Monte Carlo comparisons of EWMA and CUSUM charts reveal that none of them has overall optimal performance (Chang and Fricker, 1999). Moreover, most of these charts are designed for industrial processes and assume stationarity in the steady state, i.e. unique model parameters for the whole process, as well as the existence of a fixed target value. In applications such as monitoring pollution in environmental sciences or

[*] Corresponding author: e-mail: fried@statistik.uni-dortmund.de

controlling vital signs in intensive care stationarity can often not be assumed, and it is not possible to specify a target value in advance because of natural fluctuations within the data generating mechanism like seasonality or biorhythms (Endresen and Hill, 1977; Högel, 2000; Schmid and Steland, 2000; Gather et al., 2001; Gather, Imhoff and Fried, 2002).

We adapt Brillinger's (1989) approach for retrospective detection of a monotonic trend to the online-monitoring context by analyzing the data in a moving time window. Moving window techniques are useful to estimate time-varying model parameters and to construct adaptive control limits assuming stationarity to hold only locally (Dahlhaus, 1997). W.r.t. the length of the time window we must look for a compromise since a long time window results in a small variance for the expense of a large bias. Deterministic and stochastic variability can hardly be separated based on a moderate amount of data without any assumptions, particularly if the parameters vary in time. We restrict to the case that the noise can be approximated by an AR model as it is frequently met in practice (Gerodette, 1987; Bloomfield and Nychka, 1992). Several strategies based on linear modelling are compared for automatic detrending of the data when estimating the AR parameters, which are needed to standardize the nonparametric test statistic.

We proceed as follows. Section 2 describes the basic underlying model and the proposed procedure. In Section 3 the reliability of this procedure and of the parameter estimators is checked via simulations. In Section 4 the procedure is applied to physiological data observed in intensive care, before we finish with a discussion of the results.

2 Automatic Trend Detection

We assume that at each time point $t \in \mathbb{Z}$ the measurement Y_t of a deterministic signal μ_t is disturbed by additive autocorrelated random noise E_t,

$$Y_t = \mu_t + E_t\,, \quad t \in \mathbb{Z}\,.$$

In retrospective applications often a linear trend $\mu_t = \beta_1 t + \beta_0$ is assumed (Cochrane and Orcutt, 1949; Bloomfield and Nychka, 1992; Sun and Pantula, 1999) and it is tested whether an estimate $\hat{\beta}_1$ of β_1 is significantly different from zero. However, this means to specify a fixed form of the mean. Trends which are not linear may not be detected this way. This problem becomes even more serious in online monitoring since a procedure needs to function automatically and reliably in a wide variety of situations.

Abelson and Tukey (1963) suggest using a weighted sum $\sum_{t=1}^{N} c_t Y_t$ to test for a monotonic increase of μ_t during a time interval $t = 1, \ldots, N$, i.e. $\mu_1 \le \mu_2 \le \ldots \le \mu_N$ with $\mu_t < \mu_{t+1}$ for at least one $t \in \{1, \ldots, N-1\}$. They restrict the weights $c_1, \ldots, c_N$ to fulfill $\bar{c} = \sum_{t=1}^{N} c_t/N = 0$, such that $\sum c_t \mu_t$ also equals zero in case of a constant mean $\mu_1 = \mu_2 = \ldots = \mu_N$. Then the weights are determined to solve

$$\max_{c}\ \min_{\mu}\ \frac{|\sum (c_t - \bar{c})\,(\mu_t - \bar{\mu})|^2}{\sum (c_t - \bar{c})^2\ \sum (\mu_t - \bar{\mu})^2}\,,$$

where $\bar{\mu} = \sum \mu_t/N$, i.e. to have optimal worst case discriminatory power for an extremely unfavorable trend. This results in

$$c_t = \left[(t-1)\left(1 - \frac{t-1}{N}\right)\right]^{1/2} - \left[t\left(1 - \frac{t}{N}\right)\right]^{1/2}$$

and the corresponding worst case is a single step change. The hypothesis of a constant mean should be rejected in favor of an increasing (decreasing) mean if the standardized test statistic

$$T = \frac{\sum_{t=1}^{N} c_t Y_t}{\hat{\tau}}$$

takes large positive (negative) values, where $\hat{\tau}$ is an estimate of the standard deviation τ of the weighted sum in the numerator. Abelson and Tukey (1963) estimate τ and calculate critical values for T under the assumption that the noise process consists of independent identically distributed variables. Brillinger (1989) generalizes this approach to stationary noise processes applying a nonparametric estimator of τ and large sample asymptotics. However, Monte Carlo experiments show that his rule is too sensitive even in retrospective applications to long time series (Woodward, Bottone and Gray, 1997).

In online monitoring of a locally stationary process where both the mean and the autocorrelations may vary slowly in time we may be interested whether a monotonic trend has occurred during the last n observations. The choice of n will often be guided by the application. Large sample asymptotics like those used by Brillinger in his nonparametric approach are not useful if the window width n is moderate only. We use $n = 60$ observations corresponding to one hour of measurements. In order to separate deterministic and stochastic variability we impose a parametric model for the noise assuming that it can be approximated within each time window by an AR(p) process

$$E_t = \varphi_{1,N} E_{t-1} + \ldots + \varphi_{p,N} E_{t-p} + U_t\,, \quad t = 1, \ldots, n,$$

where we denote the observed $N \geq n$ values by $y_{n-N+1}, \ldots, y_0, y_1, \ldots, y_n$ for notational simplicity, i.e. $y_1, \ldots, y_n$ correspond to the current time window. Here, $\varphi_{1,N}, \ldots, \varphi_{p,N}$ are unknown autoregressive coefficients and $\{U_t, t \in \mathbb{Z}\}$ denotes an unobservable white noise process with uncorrelated, identically distributed disturbances having zero mean and unknown variance σ_N^2. Autoregressive models are commonly used in practice as they allow to approximate a broad variety of autocorrelation functions, and sometimes they can be justified from background knowledge. We note that all parameters may vary in time, i.e. depend on the current time window, and suppress indices representing the time window further on for simplicity. We do not address the problem of choosing the fixed order p from the incoming data as this is difficult in view of time-varying model parameters and possible trends. In many applications p can be selected by analyzing historic data representing a steady state.

In order to standardize the weighted sum $\sum_{t=1}^{n} c_t Y_t$ we need to estimate its variance

$$\tau^2 = \mathrm{Var}\left(\sum_{t=1}^{n} c_t Y_t\right) = \sum_{t=1}^{n} \sum_{s=1}^{n} c_t c_s \gamma(t-s)\,. \tag{1}$$

Hence, we need reliable estimates of the autocovariances $\gamma(0), \gamma(1), \ldots, \gamma(n-1)$ at time lags $0, \ldots, n-1$, or, equivalently, of the AR parameters. A deterministic trend seriously affects the ordinary sample autocovariances as these decay to zero very slowly, irrespective of the true values. Preliminary experiments told us that nonparametric detrending by a running mean used by Brillinger (1989) can hardly be done if the window width n is moderate since the number of observations included in the running mean must be sufficiently large to reduce the noise, but it must be small in comparison to n. Instead we use a parametric approach for detrending the data in a preliminary step. The basic idea is to approximate a deterministic trend within the current time window by a linear model $\mu_t = \boldsymbol{f}(t)' \boldsymbol{\beta}$ and to estimate the autocovariances from the residuals. In retrospective applications often a simple linear trend $\mu_t = \beta_0 + \beta_1 t$ is assumed. In the following we investigate some modifications of this idea in an automatic online application.

Fitting a simple linear trend for detrending is not completely satisfactory as we want reliable estimates for all time windows, also for those where only some of the observations are influenced by a trend, e.g. its beginning. For more flexibility we add further trend functions in the regression getting $\boldsymbol{f}(t) = (f_0(t), \ldots, f_k(t))'$ with $k + 1 \geq 2$. The functions $f_i(t)$ could be chosen as low order polynomials, but then extrapolation of a trend is difficult because of the strong curvature of polynomials of order higher than one. Instead we use ramp functions

$$g_s(t) = \begin{cases} 0, & t < s \\ t - s, & t \geq s \end{cases} \quad s \in \{0, \ldots, n-1\}\,,$$

describing a linear trend starting at time point s. Then $\mu_t = \boldsymbol{f}(t)' \boldsymbol{\beta}$ is a piecewise linear function. For $n = 60$ we take $k = 3$ with $f_1 = g_0, f_2 = g_{20}, f_3 = g_{40}$, and f_0 being constant equal to 1. This allows a rough approximation of any trend. We estimate $\boldsymbol{\beta}$ using

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{V}} = \left(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{y}\,,$$

where

$$\boldsymbol{X} = \begin{pmatrix} f_0(1) & \dots & f_k(1) \\ f_0(2) & \dots & f_k(2) \\ \vdots & & \vdots \\ f_0(n) & \dots & f_k(n) \end{pmatrix}$$

and $\boldsymbol{y} = (y_1, \dots, y_n)'$. For $\boldsymbol{V} = \boldsymbol{I}$, the $(n \times n)$-identity matrix, we get the ordinary least squares estimate $\tilde{\boldsymbol{\beta}}_k = \hat{\boldsymbol{\beta}}_{\boldsymbol{I}}$, while for $\boldsymbol{V} = \boldsymbol{\Sigma} = (\gamma(i-j))_{1 \le i, j \le n}$, the $(n \times n)$ covariance matrix of $E_1, \dots, E_n$, we get the generalized least squares estimate. While the latter is not feasible as $\boldsymbol{\Sigma}$ is unknown, the former is based on independence. For any reasonable choice of $\boldsymbol{V}$ subtracting $\boldsymbol{f}(t)' \hat{\boldsymbol{\beta}}_{\boldsymbol{V}}$ from y_t reduces the impact of a trend and we obtain estimates $\hat{\gamma}_{\boldsymbol{V}}(h)$, $h = 0, \dots, n-1$, of the autocovariances from the residuals $\hat{Z}_t = Y_t - \boldsymbol{f}(t)' \hat{\boldsymbol{\beta}}_{\boldsymbol{V}}$, $t = 1, \dots, n$.

In the following we restrict attention to AR(1) models, which are frequently used in practice, particularly for short time series. Some simplifications are possible then. We call the single autoregressive parameter φ. Since $\gamma(0) = \sigma^2/(1-\varphi^2)$ and $\gamma(h) = \varphi^h \gamma(0)$, $h \ge 1$, the autocovariances can be estimated by inserting estimates of σ^2 and φ into these equations. If the model $\boldsymbol{\mu}_t = \boldsymbol{f}(t)' \boldsymbol{\beta}$ holds exactly we get consistent estimates of φ and σ using

$$\tilde{\varphi}_k = \frac{\sum_{t=2}^{n} (y_t - \boldsymbol{f}(t)' \tilde{\boldsymbol{\beta}}_k)\,(y_{t-1} - \boldsymbol{f}(t-1)' \tilde{\boldsymbol{\beta}}_k)}{\sum_{t=1}^{n-1} (y_t - \boldsymbol{f}(t)' \tilde{\boldsymbol{\beta}}_k)^2}\,, \tag{2}$$

$$\tilde{\sigma}_k^2 = \frac{1}{n-k-1} \sum_{t=2}^{n} [(y_t - \boldsymbol{f}(t)' \tilde{\boldsymbol{\beta}}_k) - \tilde{\varphi}_k (y_{t-1} - \boldsymbol{f}(t-1)' \tilde{\boldsymbol{\beta}}_k)]^2 \tag{3}$$

(Nickerson and Basawa, 1992). We denote the number of regressors by subscripts here. In the denominator of $\tilde{\varphi}_k$ we sum up to $t = n-1$ instead of $t = n$ for bias reduction. Since $\tilde{\varphi}_k$ may turn out to be larger than one we restrict it to be at most 0.99. In order to improve these estimates we can use a two-step approach. First we estimate $\boldsymbol{\beta}$ by ordinary least squares to get an estimate $\tilde{\boldsymbol{\Sigma}}$ of $\boldsymbol{\Sigma}$ inserting $\tilde{\varphi}_k$ and $\tilde{\sigma}_k$. Then we calculate a feasible generalized least squares estimate $\hat{\boldsymbol{\beta}}_k = \hat{\boldsymbol{\beta}}_{\tilde{\boldsymbol{\Sigma}}}$. Inserting $\hat{\boldsymbol{\beta}}_k$ into (2) and (3) instead of $\tilde{\boldsymbol{\beta}}_k$ we get two-step weighted least squares estimates $\hat{\varphi}_k$ and $\hat{\sigma}_k$.

A higher dimensional parameterization $\boldsymbol{\mu}_t = \boldsymbol{f}(t)' \boldsymbol{\beta}$, $\boldsymbol{\beta} = (\beta_0, \dots, \beta_k)'$, provides flexibility for time windows with non-linear trend. In a steady state, however, it may result in instability of the mean estimates as we fit an overparameterized model then. A possibility to overcome this problem is data-driven shrinkage of the mean estimates in the full model towards the estimates in a reduced model corresponding to a steady state. Particularly, we consider a convex combination $\hat{\boldsymbol{\beta}}_S$ of $\hat{\boldsymbol{\beta}}_k$ and $\hat{\boldsymbol{\beta}}_0 = (\hat{\mu}, 0, \dots, 0)'$ corresponding to fitting a $(k+1)$-dimensional trend function and a constant mean respectively to the current time window,

$$\hat{\boldsymbol{\beta}}_S = \hat{\boldsymbol{\beta}}_k - \frac{c\hat{\sigma}_k^2}{(\hat{\boldsymbol{\beta}}_k - \hat{\boldsymbol{\beta}}_0)' \left(\boldsymbol{X}'\tilde{\Sigma}^{-1}\boldsymbol{X}\right)(\hat{\boldsymbol{\beta}}_k - \hat{\boldsymbol{\beta}}_0)} \,(\hat{\boldsymbol{\beta}}_k - \hat{\boldsymbol{\beta}}_0)\,. \tag{4}$$

The amount of shrinkage is controlled by a factor which is just c times the inverse of a χ^2-statistic which compares the fit of the reduced model to the fit of the full model. We restrict this factor not to exceed one. Shrinkage estimators $\hat{\varphi}_S$ and $\hat{\sigma}_S$ can be obtained by inserting $\hat{\boldsymbol{\beta}}_S$ into formulas (2) and (3). Shrinkage estimation of the parameters of a regression model with correlated errors has been treated by Nickerson and Basawa (1992) and by Chaturvedi and Wan (2000). The former authors show that the shrinkage estimator of the mean has smaller expected weighted mean square error than the weighted least squares estimator if the covariance matrix is known up to a scale parameter, $n \geq k$ and $0 \leq c \leq 2(k-1)(n-k)/(n-k+3)$. For unknown covariances, they prove that shrinkage estimation of the mean has higher asymptotic efficiency than generalized least squares if the true mean is approximately constant. In the latter paper it is shown that the shrinkage estimator of the mean dominates feasible generalized least squares in case of known variances and unknown correlations. Thus, shrinkage may well improve the estimation of the mean; see Sargan (2001) for a discussion of this topic. We set $c = 4$, that is at the upper end of the interval suggested by Nickerson and Basawa since we have $k = 3$.

We are not interested in the mean estimates themselves but we use them to detrend the data and to estimate the AR parameters. Therefore we check in the next section whether these improved regression estimates also result in better estimates of the autocovariances. We compare the performance of $(\tilde{\varphi}_1, \tilde{\sigma}_1^2)$, $(\tilde{\varphi}_3, \tilde{\sigma}_3^2)$, $(\hat{\varphi}_3, \hat{\sigma}_3^2)$ and $(\hat{\varphi}_S, \hat{\sigma}_S^2)$, where the index denotes the number of regressors in the trend function, while S denotes shrinkage. The first estimator is computationally cheap as we fit a simple linear trend, while the second needs fitting a trend with four parameters and the third is a two-step estimator, that needs iteration. Shrinkage estimation only affords additional evaluation of equation (4) once in each step.

There are further possible variations of the approach suggested above. The weighted sum test statistic could be applied to the residuals derived from fitting an AR model to the data instead of the untransformed observations. Then we could use the standardization for independent observations and would not need to adjust the control limits for the autocorrelations. However, experience shows that it is often better to use the original observations than residuals from a possibly misspecified time series model (see Lu and Reynolds, 1999, and the references cited therein). Another approach would be to use isotonic regression (Wu, Woodroofe and Mentz, 2001) for detrending if it was known in advance that all changes of the mean are monotonic. However, this restriction may be unduly severe if the mean shows non-monotonic behavior in the steady state. Alternatively, a referee suggested using local polynomials instead of imposing a parametric form for detrending. Indeed, Einbeck and Kauermann (2003) apply local polynomials for online trend detection and they shrink a local linear to a local constant fit. However, their approach is distinct from ours in several aspects: Firstly, their procedure is based on the comparison of two level estimates, while ours estimates the mean only for estimation of the autocorrelations and standardizing the test statistic, which uses optimal weights. Secondly, they do not consider time-varying autocorrelations. Thirdly, they control the amount of shrinkage by a local slope estimate and not by comparing the residuals, which we regard as preferable. Finally, they use longer time windows in their nonparametric procedure than we do imposing parametric assumptions.

3 A Simulation Study

In the following we perform Monte Carlo experiments to check the reliability of the parameter estimators, to derive critical values and to check the power of the proposed procedure.

3.1 The simulation design

We simulate time series $Y_1, \ldots, Y_{300}$ of length $N = 300$ corresponding to five hours of measurement from a process

$$Y_t = \mu_t + E_t\,,$$

$$E_t = \varphi E_{t-1} + U_t$$

sampled every minute, where U_t denotes $N(0,1)$-distributed disturbances. We consider $\varphi \in \{0.0, 0.1, \ldots, 0.9\}$ and several mean functions μ_t. The latter either include no trend, a linear trend $\mu_t^{(1)} = a(t-100)\, 1_{100<t<200} + 100a 1_{199<t}$ starting at time point $t = 101$ with a duration of 100 time

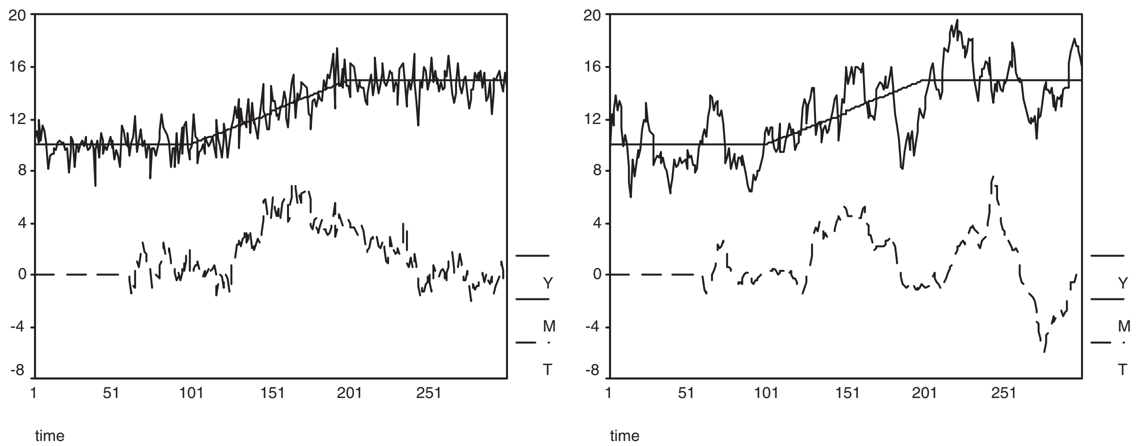

Fig. 1 Simulated time series (solid), underlying mean M (bold solid) and test statistic T (dashed) for zero ($\varphi = 0.0$, left) and large autocorrelations ($\varphi = 0.9$, right).

points and slope $a \in \{0, 0.05, 0.1\}$, or a non-linear trend $\mu_t^{(2)}$ having the shape of the ascending part of the sinus-function, i.e. zero derivatives at the endings, and causing the same total change in mean as a linear trend with slope $a \in \{0.05, 0.1\}$. For monitoring, we move a time window with $n = 60$ observations through the series.

First we illustrate the problems resulting from positive autocorrelations and the behavior of shrinkage estimates as proposed above. Figure 1 shows simulated time series with a linear trend, slope $a = 0.05$ and $\varphi = 0.3$ (small autocorrelations) or $\varphi = 0.9$ (large autocorrelations). The test statistic T with shrinkage-based standardization as described above is also depicted. For $\varphi = 0.9$ the trend is barely visible as there are monotonic increasing as well as monotonic decreasing patterns. Nevertheless, T increases during the trend period even in case of such high autocorrelations, but it takes more time until it becomes large. If φ is small or moderate (not shown here) T increases strongly briefly after the start of the trend.

Figure 2 shows the shrinkage factors for the previous time series as a function of time. Additional simulations for $\varphi = 0.0$ and $\varphi = 0.6$ reported in our discussion paper (Fried and Imhoff, 2002) show that for small to moderate autocorrelations this factor is close to one in a steady state, while it is close

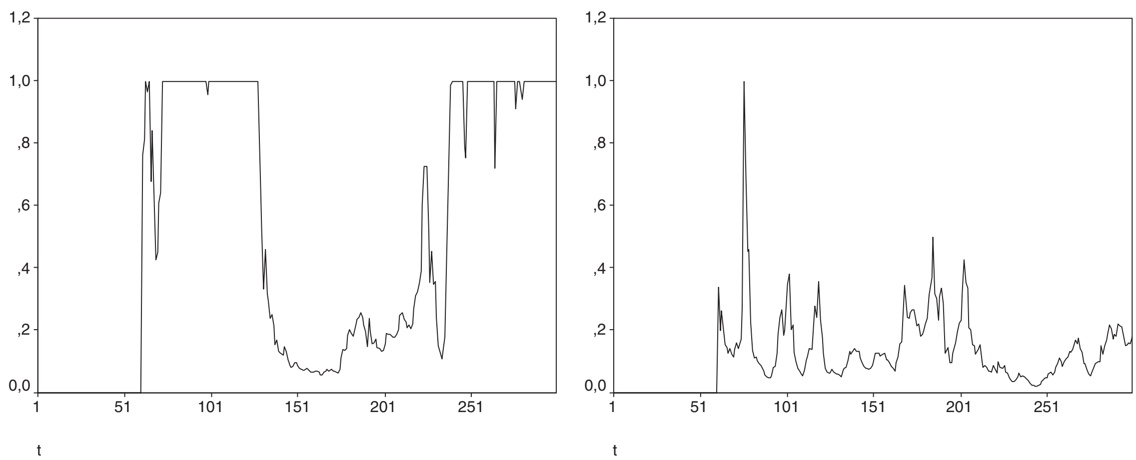

Fig. 2 Shrinkage factors for simulated time series with inserted trend and small ($\varphi = 0.3$, left) or large autocorrelations ($\varphi = 0.9$, right).

to zero in a trend period. In case of large autocorrelations ($\varphi = 0.9$), monotonic patterns cause the shrinkage factor to be small most of the time.

A comparison of the mean estimates reveals that a four dimensional regression function fits the data most of the time well, but during the steady state the data are overfitted as could be expected. Shrinkage reduces this overfitting, while fitting a simple linear trend means a crude approximation particularly at the beginning and the end of a trend period.

3.2 Comparison of the parameter estimators

For standardizing the weighted sum we need a reliable estimator of its standard deviation τ. In order to explore the properties of the distinct methods we simulate 500 time series for each of $\varphi = 0.0, 0.3, 0.6, 0.9$ and a linear trend with slope 0.05 between time points 101 and 200. We average the parameter estimates across the 500 time series for each model and each time point to assess the bias of the distinct methods.

Figure 3 depicts the averaged estimates of τ. The estimator based on the sample autocovariances without detrending is not depicted here since it shows a very large positive bias in a trend period, i.e.

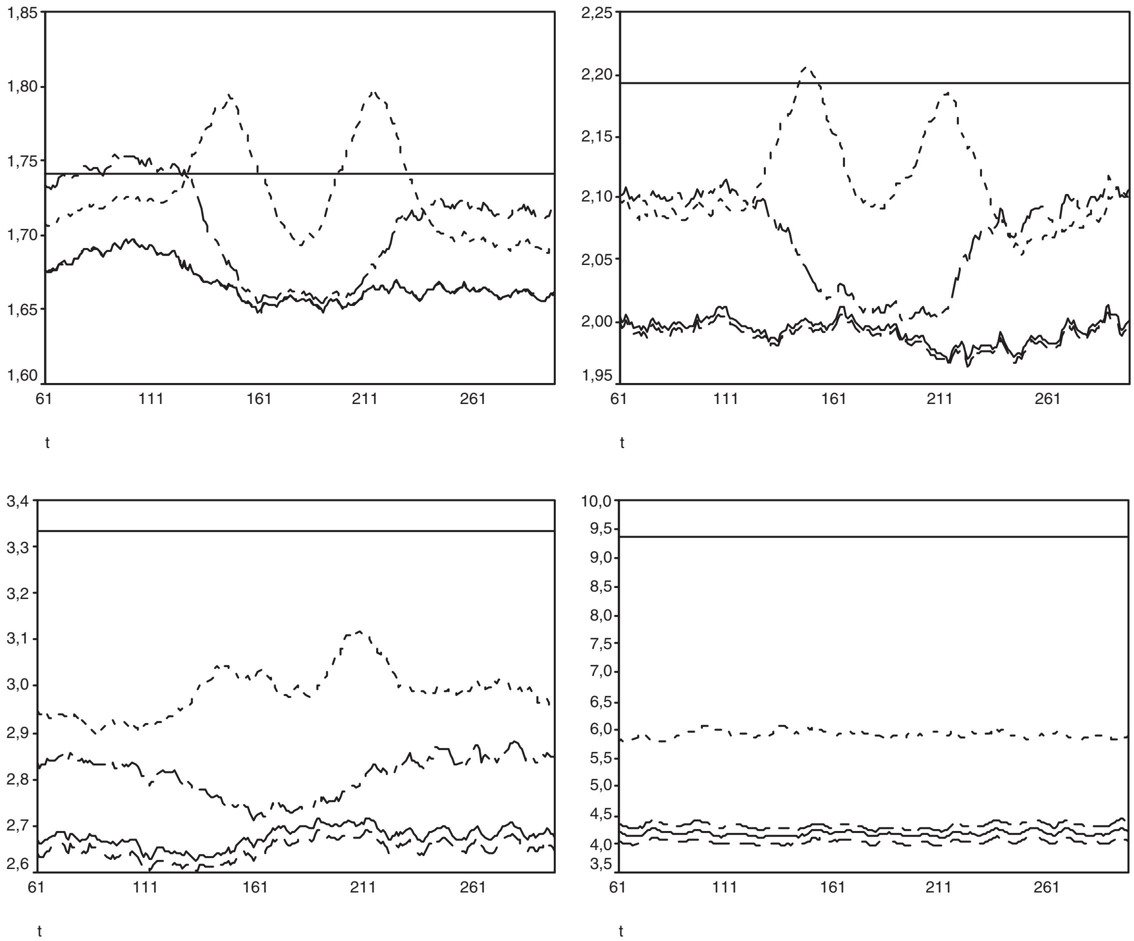

Fig. 3 Mean estimates of τ in dependence on φ in case of a linear time trend between $t = 100$ and $t = 200$: Zero ($\varphi = 0.0$, top left), small ($\varphi = 0.3$, top right), moderate ($\varphi = 0.6$, bottom left) and large ($\varphi = 0.9$, bottom right) autocorrelations. True parameter (bold solid), simple linear detrending (dotted), one-step/two-step third order detrending (dashed/solid) and shrinkage estimates (dashed-dotted).

including it would prohibit discerning differences between the other methods. All other methods show a negative bias, that increases with increasing autocorrelations. When all observations within the current time window arise from a steady state, simple linear detrending has the smallest (negative) bias. Two-step estimation provides only minor improvement as compared to one-step third order detrending, that is almost negligible in case of small autocorrelations. This is in line with the results of Bloomfield and Nychka (1992), who find the ordinary least squares and the optimal (in the sense of mean square error) unbiased estimate of a simple linear trend in case of AR(1) disturbances to be very close to each other if the autocorrelations are moderate to small. Grenander (1954) shows that ordinary least squares is asymptotically fully efficient in a broad range of regression models with correlated errors. Shrinkage estimation is usually in between simple linear and third-order detrending in a steady state.

In a trend period, simple linear detrending results in estimates that are typically larger than in a steady state. Its bias may even become positive if the autocorrelations are small. Hence linear detrending reduces the effect of a trend on the test statistic. The negative bias of shrinkage estimation, however, increases even slightly in a trend period if the autocorrelations are not very large. This is due to the fact that the shrinkage estimators are close to the ordinary, not trend corrected estimators in a steady state and close to third order regression during a trend. This is advantageous since it increases the differences between the values of the test statistic in a trend period and in a steady state even more, while we can cope with a bias in a steady state adjusting the estimator or the critical value.

Further considerations for the AR parameters φ and σ lead to very similar results (Fried and Imhoff, 2002). For σ, shrinkage results even in a smaller bias than simple linear detrending. An additional analysis for non-linear, sinusoidal trends leads to essentially the same results, and an analysis of the variances of the estimators does not reveal large differences.

In view of these results we prefer shrinkage detrending if computation time is not extremely critical. The increasingly negative bias of these estimators in trend periods increases the power of the procedure. Two-step estimation without shrinkage improves the results only in case of very large autocorrelations. The estimates obtained from fitting a straight line increase in a trend period, but the estimates without detrending are much worse.

Although bias corrections for AR parameters exist (Fuller, 1996, chapter 6.2), we are not aware of them for shrinkage estimation. Therefore, we estimate the bias of $\hat{\varphi}_S$ by simulations. For each of $\varphi = 0.0, \ldots, 0.9$, we simulate 200 time series of length 300. Then we calculate the sample means of the resulting $200 \cdot 241 = 48200$ shrinkage estimates, cf. Table 1. The standard error is about 0.0006 for each value of φ. Plotting the simulated bias against φ reveals that a quadratic function might be appropriate. We find the linear term not to be significantly distinct from zero and the adjusted R^2 to increase from 0.949 to 0.955 when neglecting it. Fitting a pure quadratic function we get the approximately bias-corrected estimates $\hat{\varphi}_{S,bc} = \hat{\varphi}_S(1 + 0.305\hat{\varphi}_S) + 0.0424$.

3.3 Critical values

As stated above, Brillinger's (1989) rule for retrospective trend detection is too sensitive even for long time series. This problem may become even more serious in online monitoring since we perform multiple testing at subsequent time points. To overcome this problem, we derive approximate critical values for the test statistic via simulations. We simulate 5001 time series for each of $\varphi = 0.0, 0.1, \ldots, 0.9$ and a constant mean. Shrinkage estimation with the bias correction derived

Table 1 Simulated and fitted bias (multiplied by -1) of the shrinkage estimator $\hat{\varphi}_S$ in case of a steady state for several φ.

φ	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
sim.	0.0365	0.0496	0.0511	0.0668	0.0763	0.0977	0.1051	0.1254	0.1619	0.2084
fit.	0.0432	0.0436	0.0496	0.0594	0.0748	0.0922	0.1175	0.1435	0.1670	0.1887

Table 2 Percentiles of Max T_i in case of a constant mean for $\varphi = 0.0, \ldots, 0.9$.

φ	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
90%	3.476	3.535	3.644	3.707	3.789	3.961	4.163	4.496	5.269	7.379
95%	3.724	3.840	3.925	4.027	4.143	4.338	4.590	5.051	6.058	8.593
97.5%	3.961	4.032	4.184	4.347	4.521	4.640	5.102	5.571	6.725	9.856
99%	4.235	4.317	4.604	4.771	4.991	5.170	5.630	6.344	7.748	11.415
99.5%	4.537	4.652	4.832	5.008	5.285	5.573	6.006	6.984	8.391	12.360

above is used to standardize the weighted sum and for each time series the maximal absolute value of the test statistic is calculated.

Table 2 provides empirical percentiles of these maxima for each φ. The $(1-\alpha)$-percentile means an approximate 2α-significance limit for a test whether a monotonic trend occurs during 300 observations corresponding to five hours of measurement in our application. The increase of the percentiles for large φ is not completely satisfactory as we hence should choose critical values in dependence on the estimate of φ. However, the percentiles are rather stable for small to moderately large autocorrelations allowing us to regard $c = 5$ as a conservative 5% significance bound for wrong detection of a trend within five hours of measurements if we estimate φ to be less than 0.5. If the estimate is larger we should use a larger critical value, that can be chosen from Table 2 by interpolation.

3.4 Statistical power

Now we inspect the power of the proposed procedure. We simulate 200 time series of length 300 for each of several models. Either a linear or a sinusoidal trend is inserted between $t = 101$ and $t = 200$ causing a total change of 5σ or 10σ. We use the weighted sum test statistic T with $n = 60$ for monitoring and apply the bias-corrected shrinkage estimators for standardization. We choose $\alpha = 5\%$ and select the critical value corresponding to the shrinkage estimate for the current time window from Table 2 by interpolation. Then we count in how many time series a trend is detected between $t = 101$ and $t = 260$ for the first time as a signal outside this period means a false alarm. To check the validity of the critical values we also analyze time series without a trend. Here, we calculate the number of time series in which a trend is detected at any time point since all alarms are false then.

Table 3 Number of identified trends (first line) and average delay of trend detection (second line) for several trend sizes (measured in multiples of the standard deviation σ): No trend (top), linear trend (center) and sinusoidal trends (bottom).

Size	φ									
	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0σ	7	7	10	5	16	8	12	20	24	63
5σ	200	199	198	198	198	187	164	127	88	92
	41.3	43.7	44.5	49.0	56.2	58.4	65.6	68.5	70.7	85.4
10σ	199	199	199	196	198	200	194	185	153	132
	31.3	32.4	33.5	36.6	38.4	42.3	49.4	54.4	62.1	68.4
5σ	197	197	200	200	191	180	151	115	97	84
	34.4	36.9	40.4	45.9	50.8	57.2	65.1	70.4	77.2	76.2
10σ	200	199	198	198	198	197	194	186	176	137
	20.8	22.7	25.3	27.4	31.7	35.2	44.4	51.5	57.0	67.7

Table 3 provides the numbers of time series in which a trend is detected. All trends considered here can be detected reliably if φ is small to moderate. Sinusoidal trends pose larger difficulties than linear trends, which might be caused by their smooth beginning. However, the power increases with increasing steepness also for such non-linear trends.

For large autocorrelations, say $\varphi \geq 0.8$, it is difficult to distinguish whether monotonic sequences are due to autocorrelations or due to a trend. For $\varphi = 0.9$ the number of false alarms largely exceeds the percentage of false alarms regulated by the significance level even though we derived critical values from simulations. Woodward and Gray (1993) note that deterministic trends are very difficult to detect in short time series if the autocorrelations are high. The large percentage of false alarms found here may be caused by the need to estimate φ for choosing the critical value, while these values were derived assuming φ to be fixed and known. The results might improve if we demand that T exceeds the critical value at some subsequent time points in order to reduce the impact of minor fluctuations.

In Table 3 the average delay of trend detection is also provided. We only consider those cases where a correct alarm is given. The delay increases slightly for small to moderate autocorrelations, and substantially for large autocorrelations. For small autocorrelations on average about 45 (34) observations are sufficient to detect a linear trend with slope 0.05 (0.1), while more than 60 (50) observations are needed in case of large autocorrelations. When the autocorrelations are very large, reliable and fast discrimination of a trend from a steady state affords a trend to be stronger than those considered here. The average delay for a sinusoidal trend is typically smaller, but we have to keep in mind the lower detection rate. The time delay could be reduced by using a shorter time window but the expense would be a larger number of false alarms, particularly for large autocorrelations.

The trend for $\varphi = 0.3$ shown in Figure 1 is detected with a delay of 44 observations, while in the time series with $\varphi = 0.9$ no trend is detected at all as the large positive autocorrelations cause monotonic patterns with duration less than 30 observations. Here, the systematic changes are small in comparison to the random variability found in the data.

4 Application to Real Time Series

In order to judge the performance of the proposed procedure for real data we analyzed time series representing physiological variables like the heart rate or blood pressure and found the procedure to perform well. Analyzing a large number of such time series, Imhoff et al. (2002) find that the autocor-

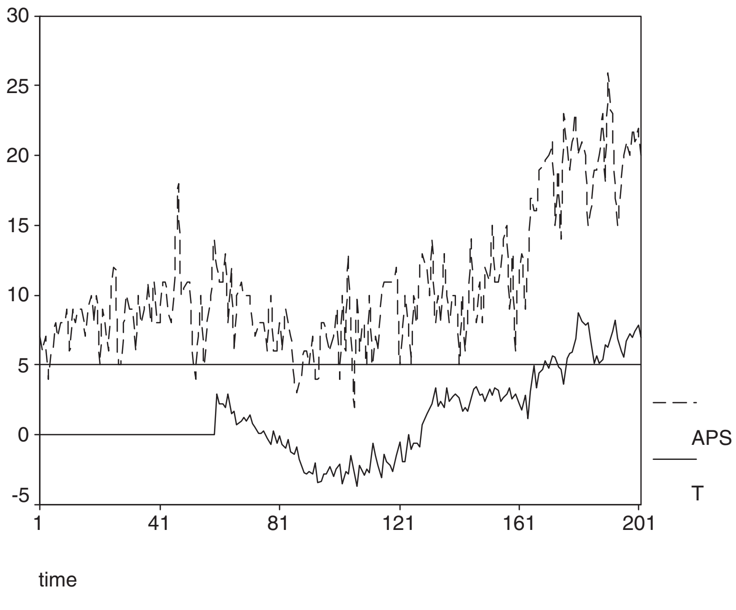

Fig. 4 Time series representing arterial pressure and standardized weighted sum statistic T. We subtracted 100 from the pressure for the reason of illustration. At $t = 165$ an upward trend is detected.

relations can be described well by AR models of low order during a steady state. We present a single example in the following. Figure 4 shows two hundred measurements of the systolic arterial pressure of a critically ill patient taken every minute. The time series first drifts slightly around a steady state and then starts increasing slowly at about $t = 100$. From about $t = 164$ on it increases more strongly. Analyzing the autocorrelations we find an AR(1) model with φ between 0.1 and 0.4 to be adequate. Figure 4 also provides the test statistic T for a moving time window of length $n = 60$. The test statistic remains well within the non-critical limits up to time point 163, then it increases and crosses the critical value $c = 5.0$ at $t = 165$.

5 Discussion

We have proposed a procedure for online detection of monotonic trends in time series with time-varying autocorrelations, which are modelled by a low order AR model for the noise. We have also investigated some variations of parametric detrending for estimation of the autocorrelations and have found that shrinkage estimation improves the discriminatory power of the test statistic. Although further work remains to be done, we consider the results to be encouraging so far if the autocorrelations are not very large. Shrinkage estimators of the AR parameters are only mildly influenced by trends, and the proposed test statistic usually results in much larger absolute values in a trend period than in a steady state. Both linear and nonlinear trends were identified correctly rather soon in most of the cases. We applied the procedure to a couple of long physiologic time series observed in intensive care and found the results to agree well with the opinion of an experienced physician. Possible problems w.r.t. the clinical relevance of detected trends can be facilitated replacing σ by a constant percentage of the current mean in the standardization. Further improvements are possible by smoothing the parameter estimates for subsequent time windows.

Trend detection is a hard problem in case of very large positive autocorrelations since the behavior of an undisturbed process is close to non-stationarity with frequent monotonic sequences. Comparing several tests for retrospective trend detection, Woodward and Gray (1993) found the rate of false alarms to be often larger than 50% then. Our procedure may be modified for very large autocorrelations by fitting a simple linear trend to the data and shrinking it towards a constant mean as detrending may be too flexible if trend functions of higher order are fitted. Some experiments showed that another possibility might be a local linear fit with a large bandwidth. In case of small to moderate autocorrelations, however, this nonparametric method did not give better results than the approach taken here. Parametric shrinkage estimation has the advantage of resulting in smaller estimates of the standard deviation in a trend period than in a steady state. Combining local polynomials with data driven shrinkage seems worthwhile, but further experience is needed e.g. with respect to the suitable automatic choice of the bandwidth for autocorrelated data (Altman, 1993). Anyway, a time window of length $n = 60$ is perhaps not sufficient to distinguish slow trends and very large positive autocorrelations as both result in similar patterns.

The proper specification of the window width n is an important issue, similarly as a suitable forgetting factor λ is needed for an EWMA chart. We consider the choice of a time period to be easier for an operator than the choice of a weighting factor. In our medical application, an experienced physician may well consider a monotonic change of a vital sign over one hour to be clinically relevant. Moreover, the power of EWMA and CUSUM charts is best for sudden shifts of the mean, whereas this is the worst case for the weighted sum statistic. This statistic mirrors the start and the end of the time window as the corresponding weights have different signs. This is similar to comparing time delayed means, that has also been suggested for trend detection (Daumer, 1997, 1998; Daumer and Neiss, 2001). Comparison of time delayed means corresponds to weighted sums with weights $-1/m_1, \ldots, -1/m_1, 0, \ldots, 0, 1/m_2, \ldots 1/m_2$. Therefore, such test statistics have lower worst-case power than that advocated here as the latter uses optimal weights.

Like EWMA and CUSUM charts the procedure as presented here is not robust against outliers. A simple possibility to overcome this deficiency is to apply additionally a procedure for online outlier detection, or alternatively the test statistic could be robustified using M-estimators for instance. Gather, Bauer and Fried (2003) propose a control chart for online outlier detection applying multivariate outlier detection rules to a multivariate embedding of the time series in order to incorporate autocorrelations. Detected outliers and missing values can be replaced by a convex combination of the current mean estimate and previous observations regulated by the estimated AR parameters. This takes better account of the dynamics of the time series than simply inserting the mean or the previous observation as these bias the lag one sample autocovariance towards zero and one, respectively. The performance of the proposed procedure will not be affected a lot as long as there are a few outliers or missings only within a single time window. Proper resolutions of further practical difficulties, which need to be found before a more extensive validation of the procedure with real data and its real-world usage certainly depend on the individual application.

Acknowledgements We thank two anonymous referees for their comments and suggestions. The financial support of the Deutsche Forschungsgemeinschaft (SFB 475, "Reduction of complexity in multivariate data structures") is gratefully acknowledged.

References

Abelson, R. P. and Tukey J. W. (1963). Efficient utilization of non-numerical information in quantitative analysis: general theory and the case of simple order. *Annals of Mathematical Statistics* **34**, 1347–1369.

Altman, N. S. (1993). Estimating error correlation in nonparametric regression. *Statistics Probability Letters* **18**, 213–218.

Bloomfield, P. and Nychka, D. W. (1992). Climate spectra and detecting climate change. *Climatic Change* **21**, 275–287.

Brillinger, D. R. (1989). Consistent detection of a monotonic trend superposed by a stationary time series. *Biometrika* **76**, 23–30.

Cembrowski, G. S., Westgard, J. O., Eggerl, A. A., and Toren, E. C. (1975). Trend detection in control data: Optimization in Trigg's technique for trend analysis. *Clinical Chemistry* **21**, 1395–1405.

Chang, J. T. and Fricker, R. D. (1999). Detecting when a monotonically increasing mean has crossed a threshold. *Journal of Quality Technology* **31**, 217–234.

Chaturvedi, A. and Wan, A. T. K. (2000). Exact results on the inadmissibility of the feasible generalized least squares estimator in regression models with non-spherical disturbances. *Biometrical Journal* **42**, 481–487.

Cochrane, D. and Orcutt, G. H. (1949). Application of least squares regression to relationships containing autocorrelated error terms. *Journal of the American Statistical Association* **44**, 32–61.

Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *The Annals of Statistics* **25**, 1–37.

Daumer, M. (1997). Adaptive Drifterkennung und Intelligente Alarmsysteme. *Biomedizinische Technik* **43**, Ergänzungsband 1, 214–215.

Daumer, M., 1998: Online Monitoring von Changepoints: Erweiterungen und Alternativen zum Schwellenwertalarm. *Journal für Anästhesie und Intensivbehandlung*, 1. Quartal, 38–40.

Daumer, M. and Neiss, A. (2001). A new adaptive algorithm to detect shifts, drifts and outliers in biomedical time series. In: J. Kunert, G. Trenkler (eds.): *Mathematical Statistics with Applications in Biometry.* Josef EUL Verlag, Lohmar-Köln, 265–275.

Einbeck, J. and Kauermann, G. (2003). Online monitoring with local smoothing methods and adaptive ridging. *Journal of Statistical Computation and Simulation*, to appear.

Endresen, J. and Hill, D. W. (1977). The present state of trend detection and prediction in patient monitoring. *Intensive Care Medicine* **3**, 15–26.

Fried, R. and Imhoff, M. (2002). On the online detection of monotonic trends in time series. Discussion Paper 68/02, SFB 475, University of Dortmund, Germany. Available from: http://www.sfb475.uni-dortmund.de/dienst/en/content/veroeff-e/tr02-e.html

Fuller, W. A. (1996). *Introduction to Statistical Time Series.* 2nd Edition. Wiley, New York.

Gather, U., Bauer, M., and Fried, R. (2003). The identification of multiple outliers in online monitoring data. *Estadistica*, to appear.

Gather, U., Fried, R., Lanius, V., and Imhoff, M. (2001). Online monitoring of high-dimensional physiological time series – a case-study. *Estadistica* **53**, 259–298.

Gather, U., Imhoff, M., and Fried, R. (2002). Graphical models for multivariate time series from intensive care monitoring. *Statistics in Medicine* **21**, 2685–2701.

Gerodette, T. (1987). A power analysis for detecting trends. *Ecology* **68**, 1364–1372.
Grenander, U. (1954). On the estimation of regression coefficients in the case of an autocorrelated disturbance. *Annals of Mathematical Statistics* **25**, 252–272.
Högel, J. (2000). Applications of statistical process control techniques in medical fields. *Allgemeines Statistisches Archiv* **84**, 337–359.
Imhoff, M., Bauer, M., Gather, U., and Fried, R. (2002). Pattern detection in intensive care monitoring time series with autoregressive models: Influence of the model order. *Biometrical Journal* **44**, 746–761.
Lu, C.-W. and Reynolds, M. R. (1999). EWMA control charts for monitoring the mean of autocorrelated processes. *Journal of Quality Technology* **31**, 166–187.
Montgomery, D. C. and Mastrangelo, C. M. (1991). Some statistical process control methods for autocorrelated data. *Journal of Quality Technology* **23**, 179–193.
Nickerson, D. M. and Basawa, I. V. (1992). Shrinkage estimation for linear models with correlated errors. *Shankhya* **54**, 411–424.
Sargan, J. D. (2001). Model building and data mining. *Econometric Reviews* **20**, 159–170.
Schack, B. and Grieszbach, G. (1994). Adaptive methods for trend detection and their applications in analysing biosignals. *Biometrical Journal* **36**, 429–452.
Schmid, W. and Steland, A. (2000). Sequential control of non-stationary processes by nonparametric kernel control charts. *Allgemeines Statistisches Archiv* **84**, 315–336.
Sun, H. and Pantula, S. G. (1999). Testing for trends in correlated data. *Statistics Probability Letters* **41**, 87–95.
Trigg, D. W. (1964). Monitoring a forecasting system. *Operational Research Quarterly* **15**, 271–274.
Woodward, W. A. and Gray, H. L. (1993). Global warming and the problem of testing for trend in time series data. *Journal of Climate* **6**, 953–962.
Woodward, W. A., Bottone, S., and Gray, H. L. (1997). Improved tests for trends in time series data. *Journal of Agricultural, Biological and Environmental Statistics* **2**, 403–416.
Wu, W. B., Woodroofe, M., and Mentz, G. (2001). Isotonic regression: Another look at the changepoint problem. *Biometrika* **88**, 793–804.