THE IDENTIFICATION OF MULTIPLE OUTLIERS IN ONLINE MONITORING DATA

Ursula Gather, Marcus Bauer, Roland Fried
Department of Statistics
University of Dortmund
44221 Dortmund, Germany

ABSTRACT

We present a new procedure for routine detection of isolated and patchy outliers in univariate time series. This procedure is especially designed for online identification of outliers, but of course it can be used retrospectively, too. It is based on an embedding of the time series which allows to regard the time series as a multivariate sample with identically distributed but non independent observations. Thus, multivariate outlier identifiers can be transferred into the context of time series. This gives interesting insights in some features of outliers in time dependent data, which are not recognizable by other methods. Some applications to online monitoring data from intensive care are included.

KEY WORDS

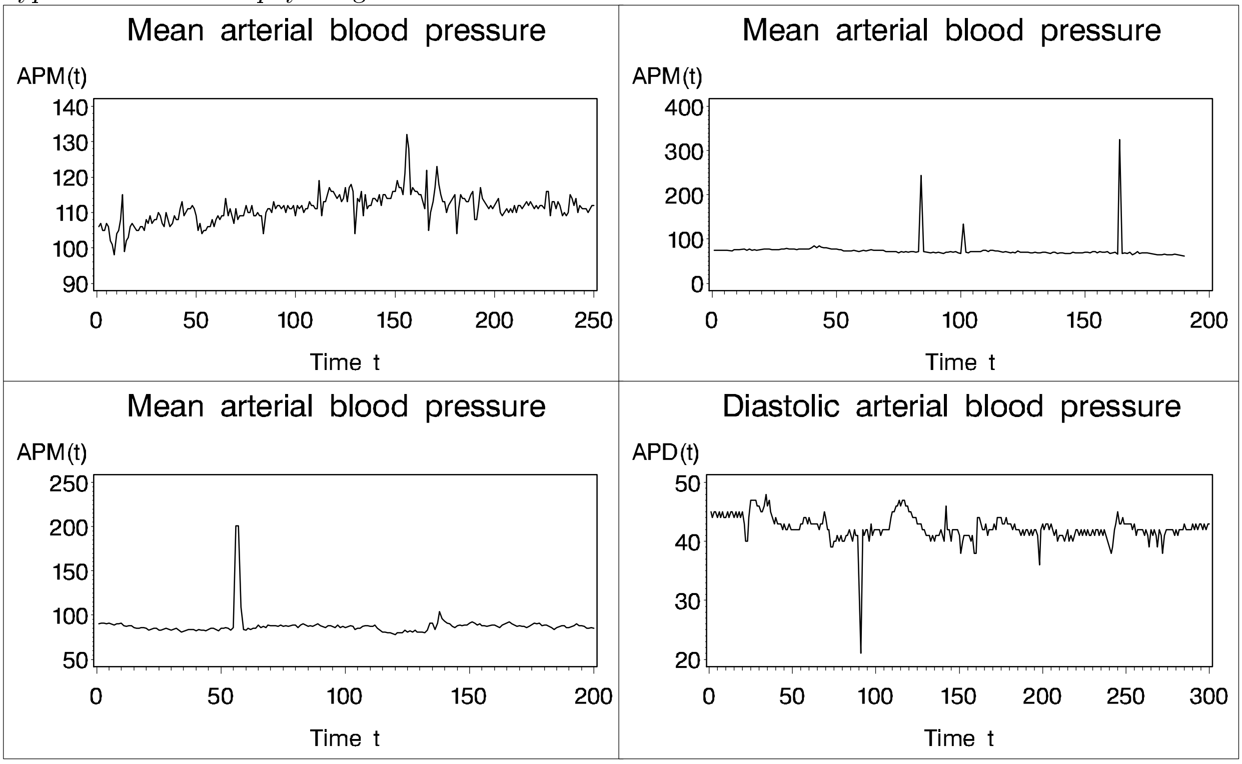
Multivariate sample, online monitoring, outlier identification, process control, time series.

1 INTRODUCTION

Increasing technical capabilities for online recording process data produce many challenges for statistical methods. In fields like intensive care medicine, industrial process control, supply chain management, or electrical energy systems more and more devices with integrated microprocessors are in use (see Imhoff (1992), Kirschen, Bruce and Wollenberg (1992), MacGregor (1997)). They allow an improved acquisition and storage of the data in real time. A common aim in all fields where data are recorded online is the monitoring of the data generating process. The automatic detection of abrupt level shifts and trends in dynamical processes is one part of this aim. Solutions have been suggested with quite different tools like time series analysis, statistical and automatic process control, neural networks, system theory, digital signal processing, artificial intelligence and fuzzy control (see Baseville and Nikiforov (1991), Oppenheim and Schafer (1995), Larsson and Hayes-Roth (1998), Linkens and Nie (1992), Navendra and Parthasarathy (1990)).
Usually, the detection of a level shift or a trend in a system leads to an alert, such that the physician, engineer or operator can check the situation and possibly take some action. A fundamental problem here is the occurrence of outliers. They can produce false alerts and if non-robust procedures are used they can mask level shifts and trends. Sometimes outliers may also carry the most important information. Hence there is a necessity for an automatic detection of outliers in online-monitoring data.

FIGURE 1

Typical time series of physiological variables



A typical example, where automatic online identification of outliers is of major interest is data from intensive care medicine. The use of clinical information systems in intensive care makes it possible to report online, simultaneously and automatically more than 2000 physiological variables, laboratory data, device parameters, etc. in the course of time. To allow for a more differentiated approach to therapy and computer aided clinical decision making, we need intelligent alarm systems allowing for a suitable bedside decision support (see Imhoff and Bauer (1996)).

The existing alarm systems are based on fixed thresholds which are chosen by the physician; they produce a great number of alarms due to measurement artifacts, patient movements or minor problems such as transient fluctuations past the set alarm limit (see O'Carrol (1986)). Most alarms, about 87% - 94% (see Lawless (1994), Mäkivirta et al. (1991)), are irrelevant in terms of patient care.

Some typical time series of physiological variables like heart rate, arterial and pulmonalarterial blood pressure recorded in one minute intervals are shown in Figure 1. They contain isolated and patchy outliers, which would produce false alarms when using nowadays alarm systems. We note that outliers in physiological time series measured in intensive care can be of very different size as there are several possible causes. In this paper a procedure for the automatic online detection of outliers is derived. This is important since constructing an alarm system based on rules for detection of level shifts and trends only neglecting outliers at all is not satisfactory. Firstly, some outliers are clinical relevant but should not immediately trigger an alarm. For instance, outliers in the heart rate of a patient point at a disturbance of the cardiac rhythm. Therefore, detection of outliers is of medical interest, here. Secondly, outliers may strongly influence any data analysis. Therefore it is important for any alarm and decision support system based on automatic statistical analysis to reduce the effects of outliers. A procedure for online outlier detection and replacement as presented here is a viable means for this task.

Imhoff et al. (1997) as well as West, Prado and Krystal (1999) indicate potential benefits of statistical time series analysis when studying such complex biomedical processes. In time series analysis ARIMA models (with batch processing of the data) and state space models (with sequential processing) are very common. Since the fundamental article of Fox (1972) there is an increasing amount of literature on outliers in time series. Apparently there are two major approaches. One possibility is to eliminate or at least reduce the effects of outliers by using robust estimators. Denby and Martin (1979), Martin (1981), Martin and Yohai (1985, 1986), and Bustos and Yohai (1986) do this for ARIMA-models, Beran (1994) for Gaussian long-memory models, and Chan and Cheung (1994), Connor (1996), Tatum and Hurvich (1993) and Kleiner and Martin (1979) treat nonlinear time series analysis, neural networks, and frequency domain analysis respectively.

On the other hand methods of outlier detection can be applied and the series can be adjusted before calculating the final estimations. In this way joint estimates of outlier effects and model parameters are obtained. This approach is best investigated for batch processing (see Chang, Tiao and Chen (1988), Chen and Liu (1993), Justel, Peña and Tsay (2001), Sánchez and Peña (1997), Tsay (1988)). A comparison of several of these procedures is given in Flak, Schmid and Sigmund (1996). Schmid (1986) uses likelihood-ratio-tests for the identification of multivariate outliers. For a review on tests for the detection of time series outliers see Hotta and Neves (1992). Sensitivity analysis for regression models is applied in the context of outlier detection in time series, for instance in Abraham and Chuang (1989), Bruce and Martin (1989), Cavanaugh and Johnson (1999), Ledolter (1990), Ljung (1993), Peña (1990). An early review on robust methods for univariate ARIMA-models is given by Stockinger and Dutter (1987).

In retrospective analysis, the method by Chen and Liu (1993) is standard nowadays. Typical problems of all retrospective ARIMA-based procedures for detecting outliers are biased estimators of initial parameter values, inappropriate model specification and masking effects, even when used iteratively (Justel, Peña and Tsay (2001), Le, Raftery and Martin (1996), Sánchez and Peña (1997), Tsay (1986)). Also, because of additional model identification steps, complicated ARIMA based procedures are not really appropriate for online monitoring.

State space models, and especially linear dynamic multiprocess models as introduced by Harrison and Stevens (1976), have also been proposed for online identification of outliers and other disturbances like effects of interventions (West and Harrison (1986)). A very well known example of a multiprocess model applied to online monitoring data after renal transplants is given by Smith and West (1983). Related work applies a linear growth model and multiprocess Kalman filtering to analyse medical online monitoring time series from anaesthesia (Daumer and Falk (1998)). Harvey and Koopman (1992) as well as Kirkendall (1992) discuss the detection of outliers in structural models using the Kalman filter, too.

Within the state space approach we find robustifications of the Kalman filter in Kitagawa (1987), Peña and Guttman (1988), Schick and Mitter (1994), and West (1981). However, all procedures are not robust against multiple or patchy outliers and some procedures fail already if two or more outliers arise within a short distance of each other (Sánchez and Peña (1997)).

All approaches in state space models can be sensitive with respect to violations of the assumption of normality and stationarity. In practical applications this leads to false classifications of clean observations as outliers which yields false alerts in online monitoring systems. A routine application of such models in intensive care units or operating rooms is thus not practised yet, also because of their very strong sensitivity against misspecification of the hyperparameters, their insensitivity against moderate level shifts, and the extreme computational effort, which algorithms in multiprocess models require. This is especially true when

several variables are controlled (Bolstad (1986)).
We choose a different, partially graphically motivated approach yielding formal, automatic procedures which can be applied online. Using similar ideas as Takens (1980, 1981) and Packard et al. (1980) we apply a mapping of the time series into a multivariate Euclidian space. Hereby we can transfer the methodology of outlier identification in multivariate data into the time series context. Therefore our paper is located in the overlap of the areas of linear and nonlinear time series, robust multivariate methods, and process control. This makes it necessary to show the connection of our proposal to these fields in order to clarify the coherences and differences.
We proceed as follows. In Section 2 we develop a simple model for the steady state of a stationary Gaussian process and we show how the construction of predictions based on this model are connected with classical AR-models. We introduce the concept of outlier identifiers for time dependent data via a multivariate embedding in Section 3 and use this concept to construct procedures for detecting outliers in time series retrospectively as well as in an online way in Section 4. In Section 5 we give a comparison with classical approaches to outlier identification in time series. The procedure is applied to online-monitoring data from intensive care in Section 6.

2 MULTIVARIATE EMBEDDING MODELS

Typically, time series models are formulated by some feedback equations. This is appropriate if prediction is the main goal, where as in (nonlinear) ARIMA models the current observation y_t depends via some function $f(\cdot)$ on former observations $\underline{y}_{t-1} = (y_{t-1}, y_{t-2}, \ldots, y_{t-p})'$ and is superposed by disturbances $\underline{\epsilon}_t = (\epsilon_t, \epsilon_{t-1}, \ldots, \epsilon_{t-q})'$ with ϵ_j i.i.d. $N(0, \sigma_\epsilon^2),\ j \in \mathbb{Z}$, such that

$$y_t = f(\underline{y}_{t-1}, \underline{\epsilon}_t),\ t \in \mathbb{Z}.$$

Often, e.g. in state space models, the observed variable is also related via a function to an unobservable state variable, which depends on its own past. For modeling outliers and effects of interventions usually an intervention term is added to such a steady state model or it is assumed that the disturbances have a contaminated distribution.
If one is only interested in describing the equilibrium or steady state of a system and in the detection of deviations from this equilibrium, it is sufficient to consider only the dependence structure of the underlying process. Therefore, in the following we formulate a simple model - the multivariate embedding model - for the underlying process which only affords a few necessary assumptions to construct a procedure for the (online-)identification of patterns in time series like outliers and effects of interventions. Because of this parsimonious model, the corresponding identification procedures afford very low computational effort.
Our approach is motivated by the observation that mapping a time series into a multivariate Euclidean space by forming vectors of time-delayed measurements may reveal structure which is difficult to discover in the original series. This idea was introduced by Takens (1980, 1981) and at the same time by Packard et al. (1980) to analyse nonlinear deterministic, especially chaotic systems. For an overview about the aspects of chaos in nonlinear time series see Tong (1995). We would like to stress here that we only apply this idea of a time series embedding often used in chaos theory. We do not treat chaotic processes but linear stochastic processes here. Nevertheless, in the following we give a short introduction into the analysis of chaotic time series and show what we can learn from it for linear stochastic processes.

Let $\{y_t\}_{t\in\{1,\ldots,N\}}$ be a time series of length N. Consider the set of m-dimensional vectors, where the components are the time delayed elements of the time series:

$$\underline{y}_t = (y_t, y_{t+T}, y_{t+2T}, \ldots, y_{t+(m-1)T})',$$

$$\underline{y}_t \in \mathbb{R}^m, \;\; T, m \in \mathbb{N}\backslash\{0\} \;\; (t = 1, \ldots, N-(m-1)T),$$

with $m, T \ll \mathbb{N}$. Here, m is called the embedding dimension and T is the time delay. It should be chosen appropriately, such that the dynamical information of the univariate time series is transformed into a spatial information within an m-dimensional space $I\!R^m$. In the theory of chaotic systems, the multivariate embedding $\{\underline{y}_t \,|\, t = 1, \ldots, N-(m-1)T\}$ is called phase space reconstruction or embedding. The claim, based on the theory of Takens (1981), is that the analysis of the phase space vectors permits certain asymptotical inferences concerning the qualitative, especially geometric behaviour of the original system. Of course there are assumptions for the approach to work, see e.g. Broomhead and King (1986), Bezerianos et al. (1995), Buzug, von Stamm and Pfister (1993).
Also, a suitable choice of T is nontrivial. For nonlinear dynamical systems it is often recommended to choose the time delay T according to the first zero crossing of the autocorrelation function. We note that different choices of T lead to different phase space reconstructions and estimated process properties.
In the present paper we consider linear stochastic systems, where the situation is less difficult; some rules for choosing m and T are given in subsection 2.1.
The methods based on multivariate embeddings, which are developed in theoretical physics, can make use of very large sample sizes and of data arising from controlled experiments. When analysing data from biological or economical systems, only relatively small data sets are available and stochastic disturbances must be taken into account. Stochastic disturbances in nonlinear dynamics are considered by Nyschka et al. (1992) as well as by Yao and Tong (1994), who investigate nonlinear autoregressive processes and their properties. In nonlinear systems the state of equilibrium (for a definition of this term in nonlinear stochastic systems see Chan and Tong (1994)) has often very complex features, which are known as 'strange attractors'. But in many medical, environmental or ecological applications one can assume that 'for all practical purposes' the underlying process can be approximated by a linear stochastic process. Hence, we use the technique of a multivariate embedding as a tool for analysing linear stochastic processes with the special aim of (online-)identification of outliers. Our approach could also be used for the detection of level shifts. This task is not treated here, but the key idea for analysing level shifts is given in the discussion. As an illustration the following examples are given:

Examples:

1.) Consider a nonlinear deterministic difference equation, which is part of the Henon attractor:

$$Y_t = 1 - 1.4Y_{t-1}^2 + 0.3Y_{t-2}, \;\; t \in \mathbb{Z}.$$

The simulated trajectory with starting value x and its 2-dimensional embedding with $T = 1$ is shown in Figure 2a and Figure 2b. The deterministic structure of the process is easy to recognize.

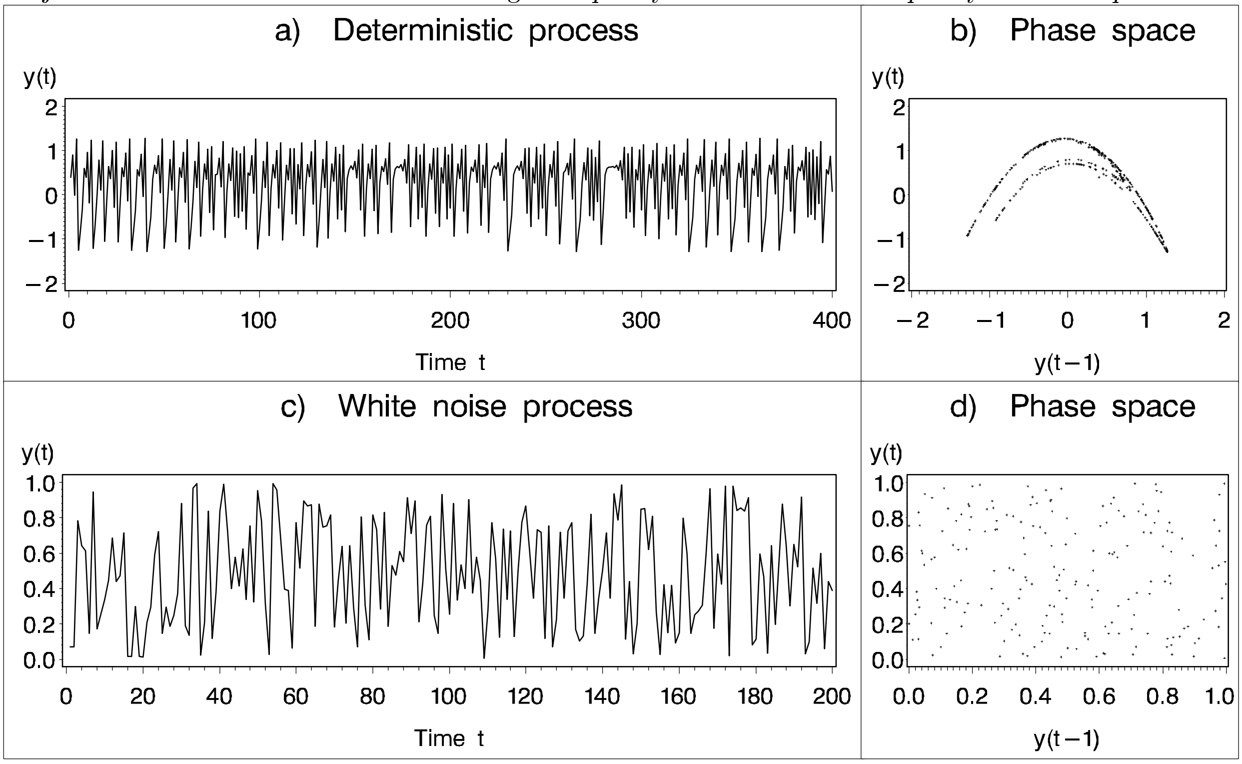
2.) In contrast to Example 1 we consider a pure White-Noise-process $\{\epsilon_t\}_{t\in\mathbb{N}}$ with $E(\epsilon_t) = 0$ and covariance

$$\gamma_k = E(\epsilon_t, \epsilon_{t+k}) = \begin{cases} 1, & k = 0 \\ 0, & k \neq 0 \end{cases}$$

as depicted in Figure 2c. The corresponding two-dimensional embedding is given in Figure 2d, where no structure is visible.

FIGURE 2

Trajectories and two-dimensional embeddings of a purely deterministic and a purely stochastic process.



3.) For a stationary AR(1)-process $Y_t = \alpha Y_{t-1} + \epsilon_t$, $\epsilon_t \sim N(\mu, \sigma_\epsilon^2)$, $\alpha \in (-1, 1)$ the time-delay vectors are concentrated in an elliptic cloud. This can be seen in Figure 3a and Figure 3b, where a trajectory of an AR(1)-process with parameters $\sigma_\epsilon^2 = 1.0$ and $\alpha = 0.5$ and its two-dimensional embedding with $T = 1$ are shown. Subsequent vectors are connected to show the movement through space. The connections are omitted in Example 1 because the visible structure would then be covered. Adding an outlier to the trajectory of an AR(1)-process at time point 337 (Figure 3c) causes a typical movement in the multivariate embedding (Figure 3d), while the outlier is barely visible in Figure 3c).

2.1 The Model

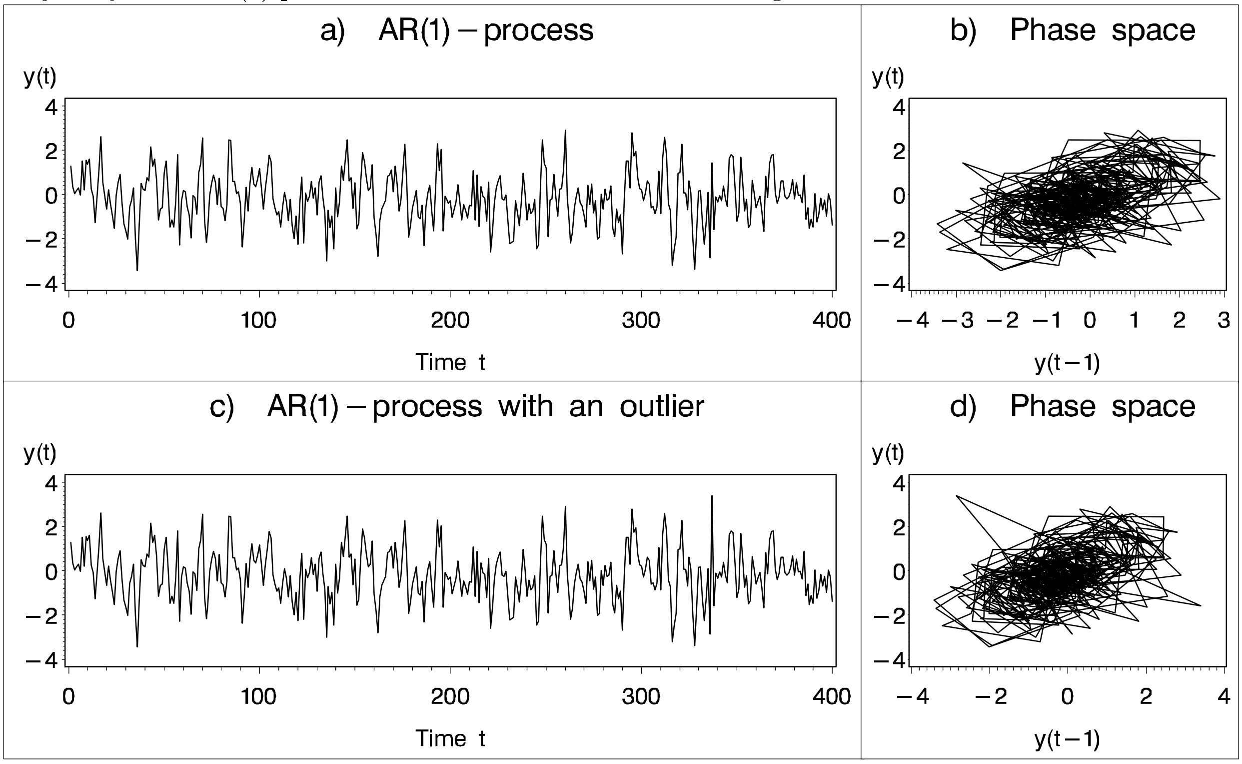
By extending Example 3 to AR(p) models and higher embedding dimensions m, we get multivariate embeddings located in m-dimensional ellipsoids, where $m = p + 1$. The smaller the dependence of consecutive process variables is, the more the shape of the ellipsoid resembles a spherical cloud. This is valid in general for stationary Gaussian processes (e.g. Brockwell and Davis, 1987, p. 98ff). Also, if we have a process with memory, but choose T such that $corr(Y_t, Y_{t-T}) = 0$, we get spherical clouds regardless of the embedding dimension m. Therefore the time delay T should be equal to one. Otherwise the information about the dependence structure, i.e. the interesting structure in linear stochastic processes is lost.

The above remarks lead to the following model. We consider a stationary Gaussian process

$$\{Y_t\}_{t\in\mathbb{Z}}, \;\; Y_t \sim N(\mu, \sigma^2)$$

FIGURE 3

Trajectory of an AR(1)-process and its two-dimensional embedding with and without outlier.



with absolutely summable autocovariance function $\gamma(h)$,

$$\sum_{h=1}^{\infty} \mid \gamma(h) \mid < \infty.$$

The (sample-) time-delay vectors $\underline{Y}_1, \ldots, \underline{Y}_{N-m+1}$ are constructed from a segment of the process with length N:

$$\underline{Y}_1 = \begin{pmatrix} Y_1 \\ \vdots \\ Y_m \end{pmatrix}, \underline{Y}_2 = \begin{pmatrix} Y_2 \\ \vdots \\ Y_{m+1} \end{pmatrix}, \ldots, \underline{Y}_{N-m+1} = \begin{pmatrix} Y_{N-m+1} \\ \vdots \\ Y_N \end{pmatrix}.$$

For linear Gaussian processes it follows immediately that the sample time-delay vectors follow a multivariate normal distribution. We recommend to choose the embedding dimension m according to

$$m = 1 + \max_{\tau}\{\tau; \mid \rho(\tau) \mid > 0\},$$

where $\rho(\tau)$ is the partial autocorrelation function (PACF).

This choice reflects the fact that it is the memory of the time series which should determine the appropriate embedding dimension. In particular, the choice $m = 2$ is not optimal for all purposes. However, one should take only those process variables into account, which have a direct influence (measured by the PACF) on the present process variable. Thus, for an $AR(2)$-process we choose $m = 3$. The absolute summability of the autocovariance function is a necessary condition for the central result given in Section 3 as well as for the construction of the online outlier-identification procedure derived in Section 4.

For a given time series the partial autocorrelation function has to be estimated from the data to choose m and therefore, we replace $\rho(\tau)$ by an estimator $\hat{\rho}(\tau)$. Typically the sample partial autocorrelation function (SPACF) based on the formula given by Durbin (1960) is used, but we note that a robust estimator would be preferable. Then we choose m according to

$$m = 1 + \max_{\tau} \left\{ \tau;\ |\ \hat{\rho}(\tau)\ | > u_{1-\alpha} \sqrt{\frac{1}{N}} \right\}, \tag{1}$$

where $u_{1-\alpha}$ is the $(1-\alpha)$-quantile of the standard normal distribution.
This model allows to interpret a time series as a multivariate sample with identically distributed but non independent observations. Thus, methods originally constructed for multivariate independent and identically distributed data can be transferred into the context of time series. One must only take care of the effects caused by the special dependence structure of the observations. Indeed, this simple model with embedding dimension m is equivalent to an AR(m-1)-model. Restricting the class of ARMA-models to AR-models is reasonable for online monitoring as this restriction allows to identify and fit a suitable model more easily using faster algorithms. We note that any ARMA-model can be approximated by an AR-model with possibly large memory $m-1$.

2.2 Predictions

The main purpose of this subsection is to reveal the relation between our approach and AR-models with respect to time series prediction. By geometric considerations a one-step-ahead prediction $\hat{Y}_{t,1}$ of the following kind is natural in the m-dimensional phase space:

$$\hat{Y}_{t,1} = arg \min_{x} \left(\sum_{i=1}^{m-1} a_i \mid\mid \underline{Y}_{t,1} - \underline{Y}_{t-i+1} \mid\mid^2 + a_0 \mid\mid \underline{Y}_{t,1} - \underline{\mu} \mid\mid^2 \right), \tag{2}$$

where $||\cdot||$ is the Euclidian norm, $x, a_i \in \mathbb{R},\ i = 0, \ldots, N-m+1$, $\underline{\mu} = (\mu, \ldots, \mu)' \in \mathbb{R}^m$ and

$$\underline{Y}_{t,1} = \begin{pmatrix} Y_{t-m+2} \\ \vdots \\ Y_t \\ x \end{pmatrix}, \quad \underline{Y}_{t-i+1} = \begin{pmatrix} Y_{t-i-m+2} \\ \vdots \\ Y_{t-i} \\ Y_{t-i+1} \end{pmatrix}, \quad (i = 1, \ldots, m-1).$$

This prediction rule demands the minimization of a weighted sum of squared distances between the $m-1$ preceding time-delay vectors and the mean vector to the new time-delay vector. This means that for predicting y_{t+1} only those observations $y_t, \ldots, y_{t-m+2}$ are taken into account which have a direct influence on y_{t+1}. This corresponds to the choice of m on the basis of the PACF in Section 2.1.
It is easy to see that minimizing (2) leads to

$$\hat{Y}_{t,1} = \left(\sum_{i=1}^{m-1} a_i Y_{t-i+1} + a_0 \mu \right) / \sum_{i=0}^{m-1} a_i, \quad \text{when} \sum_{i=0}^{m-1} a_i > 0. \tag{3}$$

It is well known that the minimum mean squared error forecast of an AR$(m-1)$-process $Y_t = \mu + \sum_{i=1}^{m-1} \alpha_i (Y_{t-i} - \mu) + \epsilon_t,\ \alpha_i \in \mathbb{R}$, is given by

$$\hat{Y}_{t,1} = \mu + \sum_{i=1}^{m-1} \alpha_i (Y_{t-i+1} - \mu). \tag{4}$$

To compare (3) and (4) let without loss of generality $\sum_{i=0}^{m-1} a_i = 1$. Then

$$\begin{aligned} \hat{Y}_{t,1} &= \textstyle\sum_{i=1}^{m-1} a_i Y_{t-i+1} + a_0 \mu \\ &= \textstyle\sum_{i=1}^{m-1} a_i Y_{t-i+1} + (1 - \sum_{i=1}^{m-1} a_i) \mu \\ &= \mu + \textstyle\sum_{i=1}^{m-1} a_i (Y_{t-i+1} - \mu). \end{aligned}$$

Thus, with $a_i = \alpha_i$ $(i = 1, \ldots, m-1)$, we can use the same prediction function as for AR$(m-1)$-processes:

$$arg\min_x \left(\sum_{i=1}^{m-1} \alpha_i \parallel \underline{Y}_{t,1} - \underline{Y}_{t-i+1} \parallel^2 + (1 - \sum_{i=1}^{m-1} \alpha_i) \parallel \underline{Y}_{t,1} - \underline{\mu} \parallel^2 \right) = \mu + \sum_{i=1}^{m-1} \alpha_i (Y_{t-i+1} - \mu).$$

The process parameters can be estimated by Yule-Walker estimators, e.g., calculated from the SPACF by solving an m-dimensional equation system. Because the SPACF has already been calculated for choosing m, the additional computational effort for prediction is not high, if the embedding dimension is low. For example, in Imhoff and Bauer (1996) and in Imhoff et al. (1997) several physiological variables from intensive care medicine are investigated and meaningful descriptions of the data are derived with low order AR models. Because of $\rho(\tau) = 0,\ \tau \geq m+1$, for an AR$(m)$ model, the embedding dimension for these applications is also small.

3 OUTLIER IDENTIFIERS FOR TIME SERIES DATA

The transformation of the dynamic information into a spatial picture leads to a multivariate sample with dependent observations. This puts us in a new situation in the framework of linear time series which to our knowledge has not been exploited so far. The transformation enables us to use multivariate methods in the embedding space of the time series.

We have just seen that outliers in time series cause a typical movement through the multivariate space (Figure 3c, 3d). Such outliers can now be detected similarly to the identification of outliers in multivariate independent and identically distributed samples, but taking into account that the vectors in the new multivariate sample are dependent.

Let us first explain what we mean by identifying outliers in a multivariate sample.

In the independent, identically distributed case with an m-variate $N(\underline{\mu}, \mathbf{\Sigma})$ target distribution, an observation $\underline{x}_i$ in some sample $\underset{\sim}{\underline{x}}_N$ of size N is identified as an outlier if it is located "too far out with respect to $N(\underline{\mu}, \mathbf{\Sigma})$", i.e. if it lies in some critical region defined by

$$OR(\underset{\sim}{\underline{x}}_N, \alpha_N) = \{\underline{z} \in \mathbb{R}^m | (\underline{z} - \underline{v})' \boldsymbol{S}^{-1} (\underline{z} - \underline{v}) \geq c\}, \tag{5}$$

where $\boldsymbol{S} = \boldsymbol{S}(\underset{\sim}{\underline{x}}_N) \in \mathbb{R}^{m \times m}$ is a symmetric and positive definite estimator of $\mathbf{\Sigma}$, $\underline{v} = \underline{v}(\underset{\sim}{\underline{x}}_N) \in \mathbb{R}^m$ is an estimate of $\underline{\mu}$ (both depending on the sample), and $c = c(m, N, \alpha_N) \geq 0$ is a critical value with $\alpha_N = 1 - (1-\alpha)^{\frac{1}{N}}$ to take the sample size into account; c can be chosen according to

$$P(\underline{X}_i \notin OR(\underset{\sim}{\underline{x}}_N, \alpha_N), \text{ for all } i = 1, \ldots, N) = 1 - \alpha \tag{6}$$

such that in a sample of size N, coming indeed independently from $N(\underline{\mu}, \mathbf{\Sigma})$, we will not identify any outlier with probability $1 - \alpha$. We call $OR(\underset{\sim}{\underline{x}}_N, \alpha_N)$ an α_N-outlier region or just outlier region. Thus the steps to choose c are the following: First one has to determine the probability distribution, the sample size N and the dimension m. This determines the shape of the critical region and choosing the level α leads to α_N which means the volume of the critical region is determined and thus c is fixed. Taking $\underline{v} = \bar{\underline{x}} = (\bar{x}_N, \ldots, \bar{x}_N)'$ and $\boldsymbol{S} = \boldsymbol{S}_N$ in (5), where $\boldsymbol{S}_N$ is the sample covariance matrix, we have described the classical outlier identifier based on the Mahalanobis-distance MD_i. Since for a multivariate independent and identically normally distributed sample we get then that $(\underline{x}_i - \underline{v})' \boldsymbol{S}^{-1} (\underline{x}_i - \underline{v})$ is equal to

$$MD_i^2 = (\underline{x}_i - \bar{\underline{x}}_N)' \boldsymbol{S}_N^{-1} (\underline{x}_i - \bar{\underline{x}}_N) \overset{N \to \infty}{\sim} \chi_m^2, \ \ i = 1, \ldots, N,$$

we set $c = \chi^2_{m,1-\alpha_N}$ (for the asymptotic χ^2_m-distribution of MD_i^2 in this case see e.g. Fahrmeir, Hamerle and Tutz (1996, p. 31)). An observation $\underline{x}_i$ can then be regarded as outlying if MD_i^2 is larger than $\chi^2_{m,1-\alpha_N}$. If an outlier is detected at time point t we replace it by its optimal linear prediction when estimating the autocovariances and the mean of the past observations at the next time point to calculate the empirical Mahalanobis distance. The resulting estimates can be expected to perform remarkably well in the presence of additive outliers (see e.g. Sejling et al. (1994)). However, to avoid masking and swamping effects in outlier identification, the use of robust estimates $\underline{v}$ of the location and $\boldsymbol{S}$ of the covariance matrix is highly recommended (see Rousseeuw (1985), Becker and Gather (1999)). The method of Ma and Genton (2000) seems very promising for the estimation of the autocovariances. Using this method the maximal temporal breakdown point possible against additive outliers, which is 25% contaminated observations, can be achieved keeping the asymptotic Gaussian efficiency at 82.27%. The method is based on a scale approach and is location-free. Therefore, it also provides some robustness against undetected trends in the data. For detailed information and formal definitions of outlier identification in multivariate data we refer to Schwager and Margolin (1982), Caroni and Prescott (1992), Davies and Gather (1993), Barnett and Lewis (1994), Gather and Becker (1998)), and for robust methods to Rousseeuw and van Zomeren (1990) as well as Becker and Gather (1999).

Now, following the model in Section 2.1 we consider time series data, which are transformed into a multivariate sample. The definition of an outlier identifier with respect to $N(\underline{\mu}, \boldsymbol{\Sigma})$ given in equation (5) can be adopted as it stands, but the dependence of the sample vectors has some consequences here.

If we consider a simple analogue of the above Mahalanobis distance for time series, i.e.

$$MDTS_t = \sqrt{(\underline{y}_t - \underline{\bar{y}}_{N-m+1})' \boldsymbol{S}^{-1}_{Y,N-m+1} (\underline{y}_t - \underline{\bar{y}}_{N-m+1})} \ \ (t = 1, \ldots, N-m+1), \tag{7}$$

we can now define an analogue to the classical multivariate outlier identifier. Here $\underline{\bar{y}}_{N-m+1} = \frac{1}{N-m+1} \sum_{t=1}^{N-m+1} \underline{y}_t$ is the arithmetic mean of the time-delay vectors and $\boldsymbol{S}_{Y,N-m+1}$ is the sample covariance matrix

$$\boldsymbol{S}_{Y,N-m+1} = \begin{pmatrix} \hat{\gamma}_N(0) & \hat{\gamma}_N(1) & \cdots & \hat{\gamma}_N(m-1) \\ \hat{\gamma}_N(1) & \hat{\gamma}_N(0) & & \vdots \\ \vdots & \vdots & & \vdots \\ \hat{\gamma}_N(m-1) & \cdots & \cdots & \hat{\gamma}_N(0) \end{pmatrix},$$

with $\hat{\gamma}_N(h) = \frac{1}{N} \sum_{t=1}^{N-h} (y_t - \mu)(y_{t+h} - \mu)$ $(h = 0, \ldots, m-1)$ where $\mu = \bar{y}_N = \frac{1}{N} \sum_{t=1}^{N} y_t$.

Definition 1 *Let* $\underset{\sim}{\mathbf{Y}}_{N-m+1} = (\underline{Y}_1, \ldots, \underline{Y}_{N-m+1})$ *be an m-dimensional random sample constructed from a set of process variables* $\{Y_t\}_{t \in 1,\ldots,N}$ *arising from a normal process with embedding dimension m. Then*

$$OR_{MDTS}(\underset{\sim}{\mathbf{Y}}_{N-m+1}, \alpha_N) =$$
$$\{\underline{z} \in \mathbb{R}^m \mid (\underline{z} - \underline{\bar{Y}}_{N-m+1})' \boldsymbol{S}^{-1}_{Y,N-m+1} (\underline{z} - \underline{\bar{Y}}_{N-m+1}) \geq \chi^2_{m,1-\alpha}\}$$

is called Mahalanobis-type outlier identifier for univariate time series. The normalization for choosing $c(m, N-m+1, \alpha_{N-m+1})$ *is according to (6) (see below).*

As in the multivariate independent and identically distributed case one could also use robust estimates for $\underline{\mu}$ and $\boldsymbol{\Sigma}$ to avoid masking and swamping. In section 5.3 we briefly compare outlier identification based on classical and robust estimates, where we apply the MCD (Rousseeuw and Leroy (1987)) for robust estimation.

It only remains to investigate the asymptotic properties of $MDTS_t$ under the new conditions. The following theorem states that $MDTS_t^2$ is still asymptotically χ^2-distributed in the dependent situation of a multivariate time series embedding (the proof is given in the appendix).

Theorem 1 *Let $\{Y_t\}_{t\in\mathbb{Z}}$ be a stationary normal process with mean μ and absolutely summable autocovariance function $\gamma(h)$, $h \in \mathbb{N}$. If a set of process variables $\{Y_t\}_{t=1,\ldots,N}$ is given and $\underline{Y}_1,\ldots,\underline{Y}_{N-m+1}$ denote the multivariate normally distributed time-delay vectors with mean $\underline{\mu}$ and covariance matrix $\mathbf{\Sigma}$, then*

$$MDTS_t^2 \overset{N\to\infty}{\sim} \chi_m^2 \quad \textit{for} \quad t = 1,\ldots,N-m+1.$$

If we choose now $\alpha_{N-m+1} = \frac{\alpha}{N-m+1}$, this guarantees asymptotically the inequality

$$P\left(\bigcup_{t=1}^{N-m+1} \{\underline{Y}_t \in OR(\underset{\sim}{\mathbf{Y}}_{N-m+1}, \alpha_{N-m+1})\}\right) \\ \leq (N-m+1)P(\underline{Y}_t \in OR(\underset{\sim}{\mathbf{Y}}_{N-m+1}, \alpha_{N-m+1})) \leq \alpha.$$

and hence the normalization fulfills (6).

4 IDENTIFICATION PROCEDURES

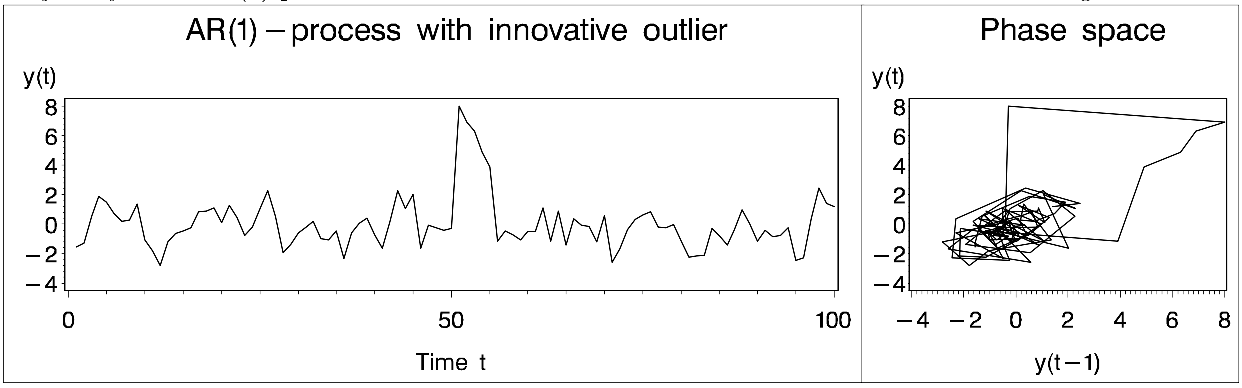
In the fundamental paper of Fox (1972) two types of outliers are defined: additive and innovative outliers. An additive outlier is an observation with an extreme distance from the rest of the data, whereas an innovative outlier is a short-term deviation from the steady state of the system. It should be kept in mind, that the perception of what is an outlier depends on the adopted model.

Following Fox, it is possible to model short-term deviations with ARIMA-models by adding an impulse function to the noise at a fixed time point. In the context of state space models we cannot speak of innovative outliers like in ARIMA-models, because there is no autoregressive structure in the observation equation. However, an innovative outlier is nothing else than a decaying level shift, which can be modeled by a contaminated noise in the state equation. In fact, a short-term deviation can also be interpreted as a level shift or an outlier patch with decreasing effect (depending on the dynamics of the process) and not as a new type of outlier. If a short term deviation is present, the time-delay vectors used in our approach leave the ellipsoid of the steady state in the m-dimensional embedding; they move then to the main diagonal and fall back along the main diagonal into the ellipsoid (Figure 4). We thus would label the first, more extreme observations after the occurrence of an innovative outlier to be outlying. We do not want to consider this as a new type of outlier, because especially in online monitoring situations one should avoid overmodeling. We rather want to identify isolated additive outliers and patchy additive outliers. A further task is the discrimination of level shifts from isolated and patchy outliers. However, we restrict here to the automatic detection of additive outliers in time series. If a time-delay vector falls into an outlier region $OR_{MDTS}(\underset{\sim}{\mathbf{Y}}_{N-m+1}, \alpha_{N-m+1})$ we have to decide, which component of the vector is the outlier in the time series.

Usually one additive outlier in a time series generates m outliers in the corresponding m-dimensional embedding. Then the additive outlier is the observation, which is exactly that component which arises in all outlying time-delay vectors. But there are two special situations, where the decision is more difficult. Firstly,

FIGURE 4

Trajectory of an AR(1)-process with an innovative outlier and its two-dimensional embedding.



in the situation where $\underline{y}_t \in OR_{MDTS}(\underset{\sim}{\mathbf{Y}}_{N-m+1}, \alpha_{N-m+1})$ but $\underline{y}_{t-1}, \underline{y}_{t+1} \notin OR_{MDTS}(\underset{\sim}{\mathbf{Y}}_{N-m+1}, \alpha_{N-m+1})$ it is not clear, which observation is an outlier.

Secondly, consider a process with embedding dimension two. Let us assume that multiple outliers occur, for instance at time points $t, t+1$, and $t+3$:

$$\begin{array}{cccccc} & AO & AO & & AO & \\ & \downarrow & \downarrow & & \downarrow & \\ y_{t-1} & y_t & y_{t+1} & y_{t+2} & y_{t+3} & y_{t+4} \end{array}$$

Then the time-delay vectors $\underline{y}_t, \underline{y}_{t+1}, \underline{y}_{t+2}, \underline{y}_{t+3}$ and $\underline{y}_{t+4}$ might be located in an outlier region. The same could be true if additive outliers in the time series occur at time points $t, t+2$ and $t+3$:

$$\begin{array}{cccccc} & AO & & AO & AO & \\ & \downarrow & & \downarrow & \downarrow & \\ y_{t-1} & y_t & y_{t+1} & y_{t+2} & y_{t+3} & y_{t+4} \end{array}$$

Clearly, the movements through space in these two cases differ and one could distinguish between these cases by a closer look at these movements. But there is a shorter recursive way to identify the right observations as additive outliers in the retrospective as well as in the online case.

First we consider the retrospective case, where the entire time series $\{y_t\}_{t\in\{1,\ldots,N\}}$ is observed. Then the PACF is estimated and according to (1) the embedding dimension is chosen. If m is determined, the dimension of $\underline{\mu}$ and $\mathbf{\Sigma}$ is known, too, and according to Section 3 outlier identifiers $OR_{MDTS}(\underset{\sim}{\mathbf{Y}}_{N-m+1}, \alpha_{N-m+1})$ can be constructed. To determine the outlier component of elements of $OR_{MDTS}(\underset{\sim}{\mathbf{Y}}_{N-m+1}, \alpha_{N-m+1})$ exactly, one should check the time-delay vectors in chronological order. If $\underline{y}_t$ falls into an outlier region, then y_t is regarded as additive outlier and it is replaced by its one-step ahead forecast $\hat{y}_{t-1,1}$, see section 2.2. Then the next time-delay vector is examined and so on.

Since there is little information at the beginning of the time series ($1 \leq t \leq m$), outlier identification is more difficult there. One may consider the vectors $\underline{y}_t = (\hat{\mu}\mathbf{1}'_{m-t}, y_1, \ldots, y_t)$ $(t = 1, \ldots, m)$. If such a vector falls into an outlier region, then y_t should be replaced by $\hat{\mu}$.

To identify outliers online we use a running time window of length N, $W_{t_{current}} = \{y_{t_{current}}, \ldots, y_{t_{current}-N+1}\}$, which is moved through the data. The estimated process parameters $\hat{\mu}_t, \hat{\gamma}_t(h)$ $(h = 0, 1, \ldots, m-1)$, and eventually also m and the parameters a_i have to be updated continually. At the beginning it is necessary to observe and analyse a starting sequence retrospectively. Then outliers can be identified online recursively and analogously to the retrospective case. The estimates of the

process parameters are updated based on the observations within the current time window $W_{t_{current}}$. In our simulations, most of the outliers could be identified correctly in this way.
Parameter estimation using a moving time window may help to overcome non-stationarity problems which can be described by slowly varying model parameters. However, the procedure described above cannot be used directly for time series which are not mean-stationary. In many applications the differenced series $\{d_t\}_{t\in\{2,\ldots,N\}}$, where $d_t = y_t - y_{t-1}$, can be regarded as stationary. The identification procedure can then be applied to the differenced series, where d_{t+1} should be replaced by $y_{t+1} - \hat{y}_{t-1,1}$, if d_t is identified as an additive outlier.
Similar to control charts from quality control we investigate two possibilities for choosing the level α. Either one can use probability limits or some kind of control limits. Choosing probability limits means to choose the level α in (6) itself. If the assumptions of normality and stationarity are fulfilled then a condition in terms of the expected number of falsely identified observations as outliers is simply given by

$$E(\text{number of falsely identified outliers}) = N_G \alpha_{N-m+1},$$

where N_G is the size of the entire data set.
In many practical situations only a deviation from the process level of more than $100k\%$, $k \in [0,1]$, is of interest and may be fixed by the physician, engineer, or operator. In monitoring situations it is often difficult to keep the number of false alarms low when using probability limits. This is due to the fact that the level α is fixed, whereas possibly the variability of the process fluctuates. If the variability of the process is small then the probability limits are very sensitive and too many outliers are detected. If the variability is large the procedure is very insensitive. Hence, an alternative approach in such situations is to choose the level α depending on the process variance. We propose a kind of control limit for choosing the level α adaptively to avoid too many false alarms. To fix ideas, let $m = 2$. Further we assume that from a practical point of view one expects, depending on the mean μ, that each measurement of a process should be in the range $[(1-k)\mu, (1+k)\mu]$ corresponding to 100k% control limits. For $m = 2$ this would lead to a control region $S = \{y_1, y_2) : (1-k)\mu \le y_1, y_2 \le (1+k)\mu\}$, where the process is said to be 'in control' at time point $t-1$ and t, when $(y_{t-1}, y_t)' \in S$. If the process variables Y_{t-1} and Y_t are independent we inscribe a circle into the square region, and generally if there are dependencies in the process we inscribe an ellipse into the square S, or equivalently for the shifted situation:

$$\begin{pmatrix} Y_{t-1} - \mu \\ Y_t - \mu \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & \gamma \\ \gamma & \sigma^2 \end{pmatrix} \right) = N\left(\mathbf{0}, \mathbf{\Sigma}_2\right).$$

Then the set

$$\{(y_1, y_2) |\ f_{\theta,\eta}(y_1, y_2) \le \chi^2_{2,1-\alpha}\}$$

$$\text{with} \quad f_{\theta,\eta}(y_1, y_2) = \begin{pmatrix} y_1 & y_2 \end{pmatrix} \begin{pmatrix} \theta & \eta \\ \eta & \theta \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}, \qquad \begin{pmatrix} \theta & \eta \\ \eta & \theta \end{pmatrix} = \mathbf{\Sigma}_2^{-1}$$

is an ellipse. The equation $f_{\theta,\eta}(y_1, y_2) = \chi^2_{2,1-\alpha}$ is equivalent to

$$\theta y_1^2 + 2\eta y_1 y_2 + \theta y_2^2 = \chi^2_{2,1-\alpha}.$$

If we want to inscribe an ellipse into the square Q we have to consider one of the functions

$$\begin{aligned} y_1 &= \tfrac{1}{2\theta}\left(-2\eta y_2 \overset{+}{-} \sqrt{4\eta^2 y_2^2 - 4\theta(\theta y_2^2 - \chi^2_{2,1-\alpha})}\right) \\ &= -\tfrac{\eta y_2}{\theta} \overset{+}{-} \tfrac{1}{\theta}\sqrt{(\eta^2 - \theta^2) y_2^2 + \theta \chi^2_{2,1-\alpha}}, \qquad -h < y_2 < h, \quad h = \sqrt{\tfrac{\theta \chi^2_{2,1-\alpha}}{\theta^2 - \eta^2}}, \end{aligned}$$

which describe the upper and lower part of the ellipse.
With

$$\begin{aligned} h &= \sqrt{\tfrac{\theta\chi^2_{2,1-\alpha}}{\theta^2-\eta^2}} \\ \iff \quad \chi^2_{2,1-\alpha} &= \tfrac{h^2(\theta^2-\eta^2)}{\theta} \end{aligned}$$

and $h = k\mu$ we get

$$\chi^2_{2,1-\alpha} = \frac{k^2\mu^2(\theta^2-\eta^2)}{\theta},$$

and hence

$$\alpha = 1 - F_{\chi^2_2}\left(\frac{k^2\mu^2(\theta^2-\eta^2)}{\theta}\right).$$

For a time window of length N it follows (to guarantee (6))

$$\alpha_{N-1} = \frac{1 - F_{\chi^2_2}\left(\frac{k^2\mu^2(\theta^2-\eta^2)}{\theta}\right)}{N-1}.$$

The parameters μ, θ and η are unknown and should be replaced by their estimators in practice. Using a moving time window in online monitoring applications, the current level α_{N-m+1} is calculated from the data of the current time window.
For a higher embedding dimension m we chose α_{N-m+1} analogously. E.g., for $m = 3$ and $(Y_{t-2} - \mu, Y_{t-1} - \mu, Y_t - \mu)' \sim N(\mathbf{0}, \mathbf{\Sigma}_3)$ one gets

$$\alpha_{N-2} = \frac{1}{N-2}\left[1 - F_{\chi^2_3}\left\{k^2\mu^2\left(\frac{\theta^2-\xi^2}{\theta} + \frac{(2\eta\xi - \theta(\eta+\xi))^2}{4\theta(\eta^2-\theta^2)}\right)\right\}\right]$$

where

$$\begin{pmatrix} \theta & \eta & \xi \\ \eta & \theta & \eta \\ \xi & \eta & \theta \end{pmatrix} = \mathbf{\Sigma}_3^{-1}.$$

The differences between using a fixed and an adaptive level are investigated in Section 6 with time series from intensive care medicine.

5 COMPARISONS

5.1 Retrospective Outlier Identification Based on Likelihood Ratio Statistics

Although the proposed procedure particularly aims at the online detection of outliers, in some situations it also has some advantages in comparison to classical retrospective procedures, where the whole sample is used for retrospective outlier detection. Fox (1972) derived an off line procedure for detecting additive and innovative outliers in AR(p)-processes, which is based on the maximum-likelihood-ratio test statistic. Chang, Tiao and Chen (1988) extended Fox results to general ARIMA(p, d, q)-models

$$\begin{aligned} &\Phi(B)\alpha(B)Y_t = \Theta(B)\epsilon_t, && \Phi(B) = 1 - \phi_1 B - \ldots - \phi_p B^p \\ &\Theta(B) = 1 - \theta_1 B - \ldots - \theta_q B^q, && \alpha(B) = (1-B)^{d_1}(1-B^s)^{d_2}, \end{aligned}$$

with $d = d_1 + sd_2, d_1, d_2, s \in \mathbb{N}$. They recommended to compute the likelihood-ratio statistic $\lambda_{A,t}$ for additive outliers, given by

$$\lambda_{A,t} = \frac{\hat{\omega}_{A,t}}{\sigma_\epsilon\left(\sum_{i=0}^{\infty}\pi_i^2\right)^{-\frac{1}{2}}},$$

where $\omega_{A,t} = \rho^2\Pi(F)\Pi(B)y_t$, $\rho^2 = (1+\pi_1^2+\ldots+\pi_{N-t}^2)^{-1}$, F (B) is the forward (backward) shift operator and $\Pi(B) = \Phi(B)\alpha(B)/\Theta(B) = (1-\pi_1 B - \ldots)$. For an AR(1)-process $\lambda_{A,t}$ is given by

$$\lambda_{A,t} = \frac{y_t - \frac{\phi_1}{1+\phi_1^2}(y_{t-1}+y_{t+1})}{\sqrt{\frac{1}{1+\phi_1^2}}\sigma_\epsilon}, \tag{8}$$

because for an AR(1)-process $\Pi(B) = \Phi(B) = 1-\phi_1 B$. On the other hand, in the case of an AR(1)-process the Mahalanobis-distance for time series reads

$$MDTS_t = \frac{y_t^2 - 2\phi_1 y_{t-1}y_t + y_{t-1}^2}{(1-\phi_1^2)\sigma_\epsilon^2}. \tag{9}$$

We can now see that in general the equations (8) and (9) do not lead to the same results. In the classical case the test statistic is based on the conditional expectation

$$\lambda_{A,t} = \frac{y_t - E(Y_t \mid y_{t-1}, y_{t+1})}{\sqrt{\frac{1}{1+\phi_1^2}}\sigma_\epsilon},$$

whereas $MDTS_t$ uses the marginal distribution of subsequent observations. Using the conditional expectation is especially crucial when multiple outliers are present, because then the estimation of outlier effects $\hat{\omega}_{A,t}$ can be mislead and thus clean observations are identified as outliers, whereas genuine outliers are not detected. If one single outlier is present, Chen and Liu (1993) realized: "when the critical value is too low, there is a higher frequency to misidentify the location of an outlier by one time period. This is due to the higher correlation between neighboring test statistics." This fact is illustrated in the example in Figure 5. In Figure 5a) a simulated AR(1)-process of length 2000 with process parameters $\phi_1 = 0.7$ and $\sigma_\epsilon = 1.0$ is simulated and its 2-dimensional embedding is shown in Figure 5c). In Figure 5b) the observations marked with '+' are genuine outliers, namely generated by

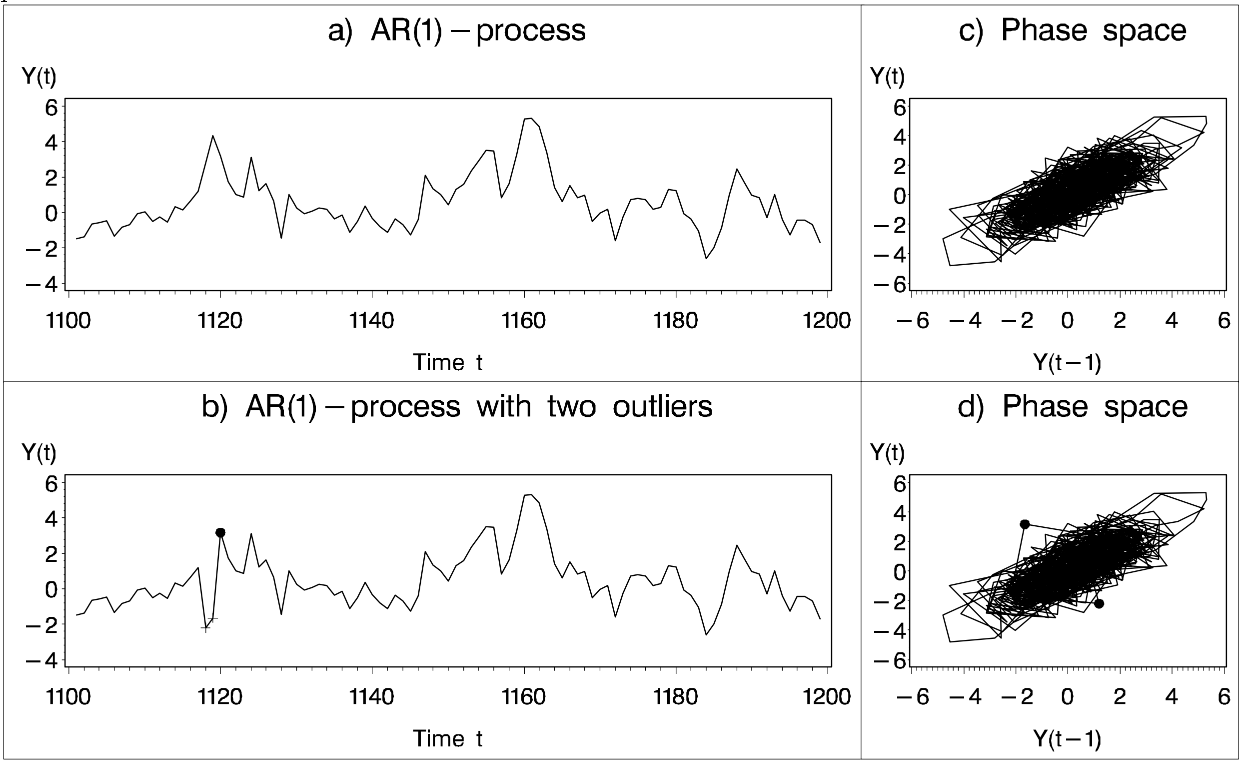
$$y_t^* \quad = \quad y_t - 5.0 I_{1118} - 6.0 I_{1119}.$$

where y_t^* is the contaminated process and I_t is an impulse function. In fact using the classical test statistic $\lambda_{A,t}$, the observation y_{1120} (marked by '•') is identified as an outlier using a level $\alpha = 0.01$, whereas the true outliers are not detected although the contamination is relatively large ($5.0\sigma_\epsilon$, $6.0\sigma_\epsilon$). Applying the outlier identifier $OR_{MDTS}(\underset{\sim}{\mathbf{Y}}_{N+1}, \alpha_{N+1})$ leads to an identification of the observations 1118 and 1120 as outliers with $\alpha = 0,01$. They are marked by '•' in Figure 5d. This example shows that OR_{MDTS} can detect genuine outliers which are difficult to identify otherwise. Obviously an outlier leads also to a larger estimate of the outlier effect $y_t - E(Y_t \mid y_{t-1}, y_{t+1}, y_{t-2}, y_{t+2}\ldots)$ at adjacent time points t and thus the test statistic at time t is larger than a critical value. $MDTS_{t-1}$ can also be smaller than the critical value so that $MDTS_t$ passes the limit. But we prefer another interpretation: something is wrong in the dynamic at the time points $t-1$ and t, but it is hard to say which observation is the outlier. This is illustrated in Figure 5d showing the two-dimensional embedding where it is obvious that changing y_{t-1} or y_t can move the process back into the steady state. Using the classical test statistic $\lambda_{A,t}$ to determine the exact time point can definitely lead to wrong decisions. Thus, it is better to investigate short time intervals to look for deviations.

When usual test statistics designed for a single outlier are applied in the 'patch of outlier situation' these outliers may mask each other, while possibly the test statistic for adjacent, genuine observations increases significantly. This effect can even arise when procedures which are based on test statistics for single outliers are used to iteratively identify multiple outliers. Chen and Liu (1993) for instance suggest an improved iterative procedure in which a single AO at any time point is tested in each step using a statistic constructed

FIGURE 5

Example of a wrong identification of the time point where an outlier occurs with a classical outlier detection procedure



from regression residuals deleting detected outliers. This already reduces masking effects. However, this procedure may fail if genuine observations are regarded as outliers at an early stage of the analysis. Then these observations are replaced by their optimal interpolation, the conditional expectation given the rest of the series, which is influenced strongly by adjacent outliers. As a consequence, the effects of these true outliers will be underestimated in the subsequent iterations and possibly we do not detect them at all using an automated procedure. In a certain sense, any known retrospective procedure can be understood as a greedy search for the smallest number of observations which do not fit to the rest of the data, particularly to their adjacent observations. It then regards these observations as outliers.

On the other hand, using Mahalanobis-distances for time series one compares the vector of time delayed observations to a mean estimate, which is much less influenced by adjacent outliers than the conditional mean. Therefore the danger of identifying a genuine observation as an outlier is reduced significantly. Moreover, if one is able to get a clean sequence for initialization of the procedure, possibly after a careful visual examination and retrospective application of outlier diagnostics, one gets the chance to detect outliers in their temporal ordering. Masking effects are overcome as long as one detects all outliers up to the current time point and replaces them by optimal predictions. Otherwise, using robust estimates of the mean and the autocovariances can easily keep the masking effect at a minimum, while this is more difficult for test statistics based on local characteristics of the time series.

5.2 Outlier Identification Based on Online Prediction

The procedure suggested in the present paper is designed particularly for the online detection of outliers. A natural competitor for this task is based on the one-step ahead prediction, which is compared to the incoming observation. More precisely, the squared standardized prediction error

$$SSPE_t = \frac{(y_t - E(Y_t|y_{t-1}, y_{t-2}, \ldots))^2}{\sigma_\epsilon^2} \tag{10}$$

could be compared to quantiles of the χ_1^2-distribution. Here, the conditional expectation of Y_t given the past observations, which is easy to calculate for an ARMA(p, q)-model, is used instead of the conditional expectation given both the past and the future. For the retrospective procedure suggested by Chen and Liu (1993), the test statistic for an AO at the end of the time series reduces to this statistic if we assume the model parameters to be known. For the practical online application the model orders have to be identified from an initial interval and the model parameters have to be estimated from past data.

Again we consider AR(1)-processes as a particular example. In this case the squared standardized prediction error reads

$$SSPE_t = \frac{(y_t - \phi_1 y_{t-1})^2}{\sigma_\epsilon^2}. \tag{11}$$

In the following we perform a Monte Carlo study to compare this approach based on ARMA(p, q)-predictions to our approach. We simulate 1000 time series of length 120 each from various ARMA(p, q)-models with errors following a standard normal distribution, and insert several outlying patterns. We consider AR(1)-processes with positive autocorrelations $\phi_1 \in \{0.0, 0.4, 0.8\}$, MA(1)-processes with positive parameters $\theta_1 \in \{0.4, 0.8\}$ and ARMA(1,1)-processes with $\phi_1 = \theta_1 \in \{0.4, 0.8\}$. In our comparison we use a (possibly misspecified) AR(1)-model for prediction, and choose the corresponding embedding dimension $m = 2$ for our approach. The determination of the model orders is difficult in an online application, hence it is important to check the robustness of a procedure with respect to deviations from the model assumptions.

A time window of length $N = 60$ is moved through the series and the model parameters are estimated from the Yule-Walker equations in each step. For $t = 61, \ldots, 120$ we calculate the squared standardized prediction error $SSPE_t$ and the Mahalanobis-distance $MDTS_t$ and compare them to the χ_1^2- and the χ_2^2-distribution respectively. The choice of the window width $N = 60$ and the total length 120 of the series is motivated by our application, where we have measurements in time lags of one minute.

If the correct parameter values were used, $SSPE_{61}, \ldots, SSPE_{120}$ would form a series of independent random variables, while $MDTS_t$ and $MDTS_{t+1}$ still would be dependent, the critical value α_N proposed in Section 3 is a Bonferroni bound. Therefore we compare $SSPE_t$ to the $\chi^2_{1,0.999}$-quantile resulting in an overall significance level of approximately 6%, and choose from our simulations the critical value $c = 11.6$ for $MDTS_t$ to achieve approximately the same significance level. When an outlier is detected using either method, it is replaced by its optimal one-step ahead prediction, which is the same in both approaches, compare Section 2.2.

After these fixings we insert several outlying patterns starting at $t = 100$. We consider a single additive outlier of magnitude $5\sigma_\epsilon^2$, two subsequent additive outliers having the magnitudes $3\sigma_\epsilon^2$ and $5\sigma_\epsilon^2$ as well as three subsequent additive outliers with magnitudes $3\sigma_\epsilon^2$, $5\sigma_\epsilon^2$ and $3\sigma_\epsilon^2$. The latter patterns describe a step-wise reaction of the system to a local disturbance. Table 5.2 provides the number of correctly identified outlying patterns by using $SSPE_t$ (upper line) and $MDTS_t$ (lower line). In this comparison, we consider a pattern to be detected if the procedure indicates an outlier at any distorted time point.

As could be expected, using the squared standardized prediction error is better than using the Mahalanobis-distance in the multivariate embedding in case of a single outlier. This is due to the fact that in the former approach we compare the incoming observation to the best prediction available if the model is correctly

TABLE 1

Number of correctly identified outlying patterns in 1000 time series generated from ARMA(1,1)-models. Upper line: Detection based on $SSPE_t$. Lower line: Detection based on $MDTS_t$.

	True model parameters (ϕ_1, θ_1)						
Outliers at	$(0.0, 0.0)$	$(0.4, 0.0)$	$(0.8, 0.0)$	$(0.0, 0.4)$	$(0.0, 0.8)$	$(0.4, 0.4)$	$(0.8, 0.8)$
$t = 100$	971	968	968	973	968	968	964
	960	959	962	958	952	955	947
$t = 100, 101$	977	967	973	980	978	982	978
	994	987	992	993	997	993	997
$t = 100, 101, 102$	981	986	981	985	983	971	977
	997	998	999	998	999	994	994

specified. However, in case of subsequent outliers $SSPE_t$ is more prone to masking effects. While it detects the first outlier more often than the Mahalanobis-distance, it very often fails to detect the second outlier also if it has already ignored the first one, compare formula (11). This also happens in case of three subsequent outliers. The Mahalanobis-distance in the multivariate embedding, however, almost always signals a disturbance at any of the inflicted observations. This is due to the simultaneous comparison of subsequent observations to the steady state, which helps to overcome masking effects. These findings are very important for the online detection of level shifts, too. All the results obtained are remarkably robust with respect to the deviations from the AR(1)-model assumptions considered here.

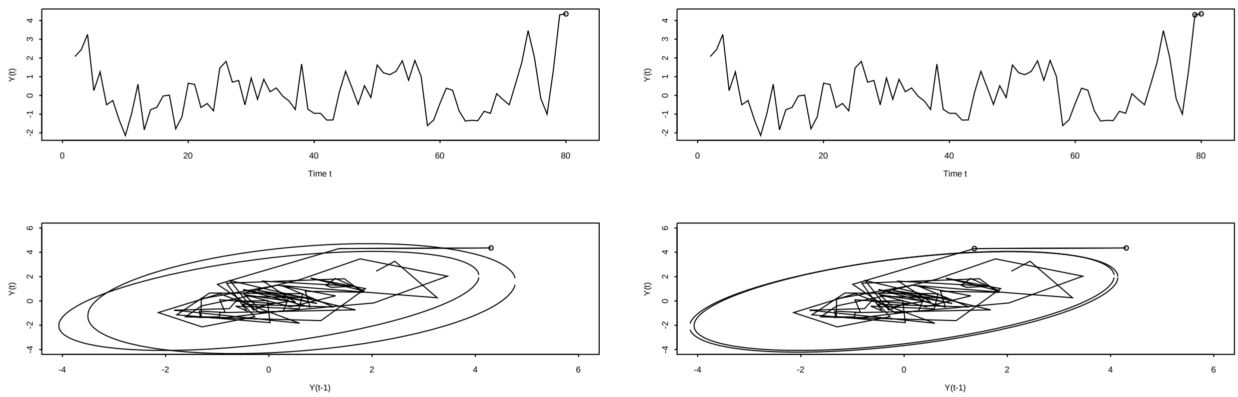
5.3 Online Outlier Identification Based on Robust Estimates

As mentioned above using the Mahalanobis-distance for outlier identification reduces the masking effect within a patch of outliers. Nevertheless, masking effects may occur if there are non-detected outliers within the current time window, which is used for the calculation of the Mahalanobis-distance. Note that replacing outliers using predictions keeps the masking effect low as long as we detect the outliers. To further explore the situation we simulate data from several AR(1)-models and replace some of the observations by outliers. We then apply the procedure for online outlier identification using the classical estimators as proposed above as well as an analogous online procedure based on the MCD (Rousseeuw and Leroy (1987)). For the latter, we do not need to replace detected outliers as the MCD is highly robust already. We choose a window width of $N = 50$ and set $\alpha = 0.1$. Normalizing constants for the MCD in a finite sample can be found in Becker and Gather (2001).

We do not want to go into too much details with respect to this comparison here since in our practical experience we have not found extreme masking and swamping effects between subsequent patches of outliers when using the non-robust estimators. The reason for this is that the procedure detects and replaces large outliers in most of the cases. Thus masking occurs only when there are small outliers, say of size 2σ or perhaps 3σ, that are followed by other outliers. In these cases, the results provided by the procedure based on non-robust estimates can be further improved using robust estimates. Figure 6 provides a time series simulated from an AR(1)-model with $\alpha_1 = 0.5$ as an illustrative example. We have replaced the observations at the time points 73, 74, 75, 78, 79, 80 by additive 2σ, 3σ, 2σ, 2σ, 3σ, 4σ-outliers respectively. The non-detected small outliers at $t = 73, 74, 75$ displace slightly the 'classical' control ellipse at the following time points, while the control ellipse based on the MCD is almost unaffected. Therefore, the robustified procedure not only detects the 4σ-outlier at $t = 80$, but also the 3σ-outlier at $t = 79$.

FIGURE 6

Outlier identification based on classical (left) and based on robust estimates (right). Time series with detected outliers marked by a circle (top), and a two-dimensional embedding with 'true' (dotted) and estimated control ellipse (solid) at time point $t = 80$. *The non-detected small outliers at* $t = 73, 74, 75$ *displace the 'classical' control ellipse.*



6 EXAMPLES

We apply the proposed procedure to online-monitoring data from intensive care medicine. We concentrate on the time series shown in Figure 1. The data are stored with a Clinical Information System (Emtek Continuum 2000, Version 4.1M3, Decision support system (DSS), Sybase SQL server 4.9.2), which has been in use successfully for eight years at the surgical intensive care unit of the Community Hospital at Dortmund (Imhoff (1995)). We consider the mean arterial blood pressure (map) and the diastolic arterial blood pressure (dpap) as examples of time series containing isolated (series 1 and 2) and/or patchy outliers (series 3 and 4).

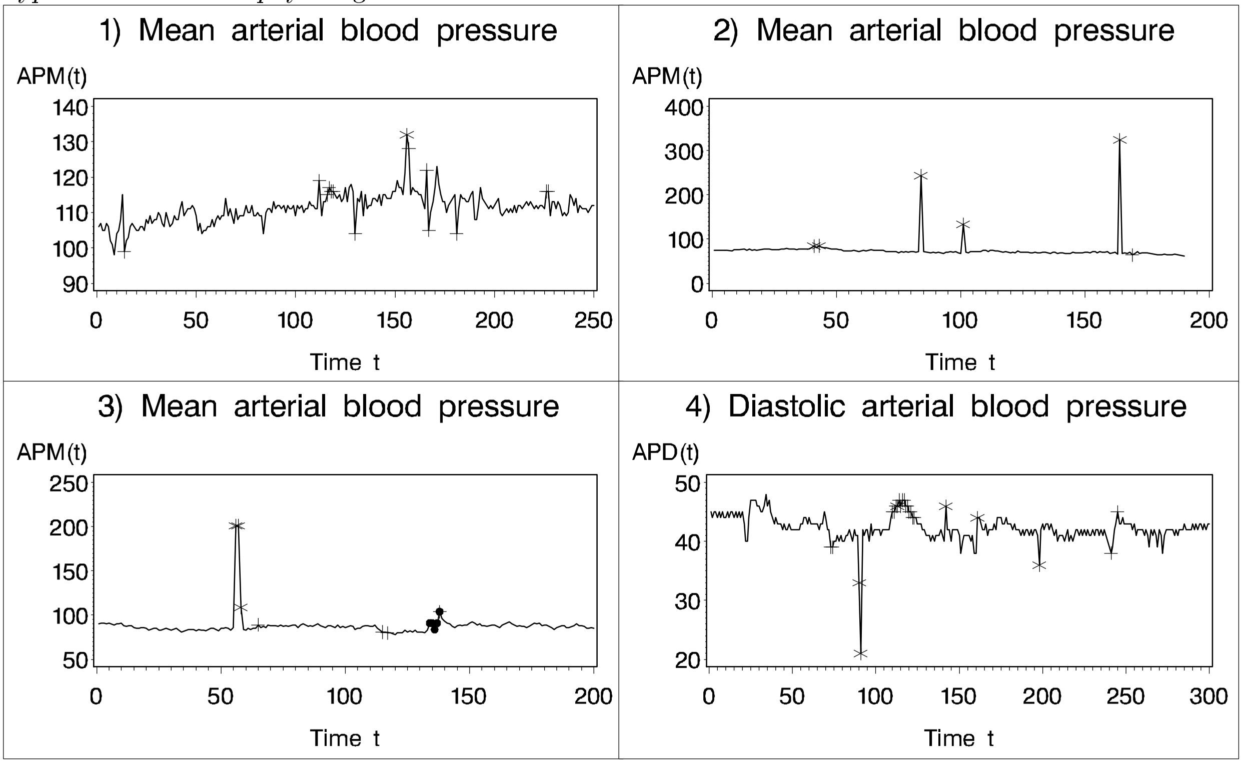
For the four time series depicted in Figure 7 we estimate the PACF on the basis of the first 50 observations. According to (1) we get for the series 1, 3 and 4 the embedding dimension $m = 2$ and for series 2 we get $m = 3$.

Using the procedure presented above we analyse each time series with a fixed level as well as an adaptive one. The fixed level is $\alpha = 0.01$. For the adaptive significance level, the physician claimed that every observation, which deviates more than 10% from the past level of the series, is interpreted as an artifact, thus we choose $k = 0.1$. Finally the length of the time window is set to $N = 30$.

The results of the analysis are summarized in Table 2 and the observations identified as outliers are marked in Figure 7. Observations which are identified with the fixed as well as the adaptive level are marked with a star $*$. Those observations, which are only identified using the fixed (adaptive) level are marked with $+$ ($\bullet$). Crosschecking these identifications with a senior intensivist yielded that all artifacts labeled by him were reliably identified with the new procedure. With the fixed level more observations not labeled by the

FIGURE 7

Typical time series of physiological variables



intensivist were identified as additive outliers than with the adaptive level.

7 CONCLUDING REMARKS

The proposed procedure for detecting outliers in time series consists of a combination of a multivariate embedding of linear stochastic processes and techniques for the identification of outliers in multivariate data. To our knowledge the advantages of such an embedding of time series have not been investigated up to now. Though Brockwell and Davis (1987) as well as Box, Jenkins and Reinsel (1994) plot such 2-dimensional embeddings in the context of regression, there is no theoretical extension and further use of this tool so far. The main properties of our method are:

- it is able to automatically detect outliers which are hard to detect otherwise and it is not prone to the masking effect,
- it detects critical sequences of observations and patchy outliers,
- it works online,
- it is easy to implement and has low computational effort,
- it gives new insights into the behaviour of outliers in time series.

These properties are fundamental for automatic online outlier detection to work in practice. We applied the proposed procedure to both simulated and real physiological time series and found the results to agree

TABLE 2

Identified outliers in the investigated time series.

Series 1		fixed level $\alpha_N = 0.000345$	adaptive level	
AO at time point	'real' artifact	$1-p_t$	α_N	$1-p_t$
14	no	0.999874		
112	no	0.999997		
116	no	0.999901		
117	no	1.000000		
118	no	0.999980		
119	no	0.999979		
130	yes	1.000000		
156	yes	1.000000	0.000032	1.000000
157	yes	1.000000		
166	no	0.999995		
167	no	0.999996		
181	no	0.999964		
226	no	0.999956		
227	no	0.999942		

Series 2		fixed level $\alpha_N = 0.000357$	adaptive level	
AO at time point	'real' artifact	$1-p_t$	α_N	$1-p_t$
41	no	1.000000	<0.000001	1.000000
43	no	1.000000	<0.000001	1.000000
84	yes	1.000000	0.030591	1.000000
101	yes	1.000000	0.000013	1.000000
164	yes	1.000000	<0.000001	1.000000
169	no	0.999887		

Series 3		fixed level $\alpha_N = 0.000345$	adaptive level	
AO at time point	'real' artifact	$1-p_t$	α_N	$1-p_t$
56	yes	1.000000	<0.000001	1.000000
57	yes	1.000000	<0.000001	1.000000
58	yes	1.000000	<0.000001	1.000000
65	no	0.999963		
115	no	0.999867		
117	no	0.999838		
134-137	no		<0.000735	>0.999924
138	no	1.000000		
139-141	no		<0.000735	>0.999924

Series 4		fixed level $\alpha_N = 0.000345$	adaptive level	
AO at time point	'real' artifact	$1-p_t$	α_N	$1-p_t$
73	no	0.999995		
74	no	0.999998		
90	yes	1.000000	0.000768	1.000000
91	yes	1.000000	0.000559	1.000000
110-123	LS	>0.999807		
142	yes	1.000000	0.006098	1.000000
161	no	0.999978	0.000048	1.000000
198	yes	1.000000	0.000001	0.999996
241	no	0.999985		
245	no	0.999934		

with the judgement of experienced physicians.

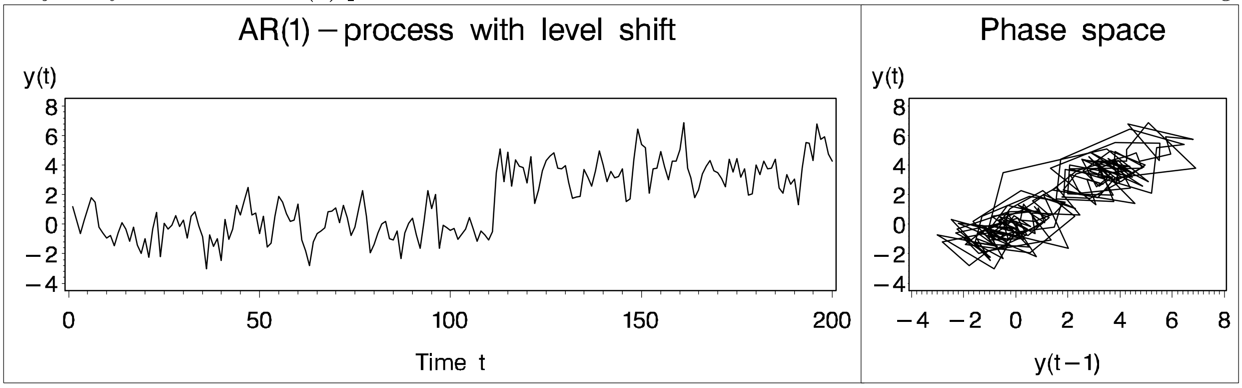
There is a connection to statistical process control because the presented procedure can be seen as a general Shewhart chart for autocorrelated data. In fact, if we have independent and identically distributed data the embedding dimension is one and thus the outlier region is $OR(\underset{\sim}{x}_N, \alpha_N) = \{x : |x - \bar{x}| > z_{1-\frac{\alpha_N}{2}} s\}$. The limits $\bar{x} - z_{1-\frac{\alpha_N}{2}} s$ and $\bar{x} + z_{1+\frac{\alpha_N}{2}} s$ are exactly the same as for a Shewhart chart for one-at-a time data. Thus for independent and identically distributed data, i.e. in the non time series situation, our procedure is identical to a Shewhart chart, if no moving window is used.

The multivariate control charts of Alt and Smith (1988), Wierda (1994), Liu (1995) as well as Nomikos and MacGregor (1995) are also only constructed for independent and identically distributed data. A multivariate extension of our procedure will give a new generalization of these multivariate control charts for dependent data.

Robust versions of the new method are straightforward, because instead of $OR_{MDTS}(\underset{\sim}{\mathbf{Y}}_{N-m+1}, \alpha_{N-m+1})$ we can use outlier identifiers based on robust estimators as well. First experiences with an implementation using the MCD estimator (Rousseeuw and Leroy (1987)) are promising. Furthermore, the procedure may be extended to identify other patterns like level shifts and trends. Figure 8 shows a 2-dimensional embedding illustrating the effect of a level shift on a simulated AR(1) process. The vectors which are affected by the level shift form a second ellipse. The different movements of the time-delay vectors through the multivariate space for different patterns may be used to identify and discriminate such patterns.

FIGURE 8

Trajectory of an AR(1)-process with a level shift and its two-dimensional embedding.



APPENDIX: PROOFS

Proof of Theorem 1: Because $\{Y_t\}_{t\in\mathbb{Z}}$ is stationary, normal and has an absolutely summable autocovariance function

$$\sum_{i=1}^{+\infty} \gamma(h) < +\infty,$$

the process is mean and covariance ergodic. That means that $\underline{\hat{\mu}}_N = \underline{\bar{y}}_{N-m+1}$ and $\hat{\gamma}_N(h) = \frac{1}{N}\sum_{t=1}^{N-h}(y_t - \mu)(y_{t+h} - \mu)$ are consistent estimators for $\underline{\mu}$ and $\gamma(h)$:

$$\begin{aligned} Var(\underline{\hat{\mu}}_N) &= O(N^{-1}), \\ Var(\hat{\gamma}_N(h)) &= O(N^{-1}),\ h = 0, 1, 2, \ldots \end{aligned}$$

It follows that $(\underline{\bar{Y}}_N, \boldsymbol{S}_{Y,N-m+1})$ are $\sqrt{N}$-consistent estimators for $(\underline{\mu}, \boldsymbol{\Sigma})$. We consider the function $g(\underline{v}) = (\underline{x} - \underline{v})'\boldsymbol{S}^{-1}(\underline{x} - \underline{v})$. A Taylor expansion with rest term r yields

$$\begin{aligned} g(\underline{v}) &= g(\underline{\mu}) + \left[\left.\frac{\partial g(\underline{v})}{\partial \underline{v}}\right|_{\underline{v}=\underline{\mu}} \right]' (\underline{v} - \underline{\mu}) \\ &\quad + (\underline{v} - \underline{\mu})' \left.\left[\frac{\partial}{\partial \underline{v}} \left(\frac{\partial g(\underline{v})}{\partial \underline{v}} \right)' \right]\right|_{\underline{v}=\underline{\mu}} (\underline{v} - \underline{\mu}) + r \\ &= (\underline{x} - \underline{\mu})'\boldsymbol{S}^{-1}(\underline{x} - \underline{\mu}) + 2(\underline{\mu} - \underline{x})'\boldsymbol{S}^{-1}(\underline{v} - \underline{\mu}) \\ &\quad + 2(\underline{v} - \underline{\mu})'\boldsymbol{S}^{-1}(\underline{v} - \underline{\mu}) + r. \end{aligned}$$

Because of $\sqrt{N}$-consistency we may neglect the rest term. Let $\underline{v} = \underline{\bar{Y}}_N$ and $\boldsymbol{S} = \boldsymbol{S}_{Y,N-m+1}$. Then, using the consistency of v and S and the statement that $(\underline{Y}_t - \underline{\mu})'\boldsymbol{\Sigma}^{-1}(\underline{Y}_t - \underline{\mu}) \sim \chi^2_m$, we get that $(\underline{Y}_t - \underline{\bar{Y}}_N)'\boldsymbol{S}^{-1}_{Y,N-m+1}(\underline{\bar{Y}}_t - \underline{\bar{Y}}_N)$ is asymptotically χ^2_m-distributed.

ACKNOWLEDGMENTS.

This work was supported by the Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 475. We thank Dr. Michael Imhoff for many helpful suggestions and for the expert classification of patterns.

REFERENCES

ABRAHAM, B. and CHUANG, A. (1989). Outlier Detection and Time Series Modeling. *Technometrics* 31, 241-248.

ALT, F.B. and SMITH, N.D. (1988). Multivariate Process Control. *Handbook of Statistics* 7, eds. P.R. Krishnaiah, C.R. Rao, Amsterdam: North Holland, pp. 333-352.

BARNETT, V. and LEWIS, T. (1994). *Outliers in statistical data* (3rd edn). New York: Wiley.

BASSEVILLE, M. and NIKIFOROV, I.V. (1991). *Detection of abrupt changes: Theory and applications*. Englewood Cliffs: Prentice-Hall.

BEZERIANOS, A., BOUNTIS, T., PAPAIOANNOU, G. and POLDOROPOULOS, P. (1995). Nonlinear time series analysis of electrocardiograms. *Chaos* 5, 95-101.

BECKER, C. and GATHER, U. (1999). The masking breakdown point of multivariate outlier identification rules. *J. Am. Statist. Assoc.* 94, 947-955.

BECKER, C. and GATHER, U. (2001). The largest nonidentifiable outlier: a comparison of multivariate simultaneous outlier identification rules. *Computational Statistics & Data Analysis* 36, 119-127.

BERAN, J, (1994). On a class of M-estimators for Gaussian long-memory models. *Biometrika* 81, 755-66.

BOLSTAD, W.M. (1986). An efficient algorithm for Harrison–Stevens forecasting using the multiprocess multivariate dynamic linear model. *Communications in Statistics B* 15, 819-828.

BOX, G.E.P., JENKINS, G.M. and REINSEL, G.C. (1994). *Time Series Analysis. Forecasting and Control.* (3rd edn). Englewood Cliffs: Prentice-Hall.

BROCKWELL, P.J. and DAVIS, R.A. (1987). *Time Series: Theory and Methods.* New York: Springer.

BROOMHEAD, D.S. and KING, G.P. (1986). Extracting qualitative dynamics from experimential data. *Physica* 20D, 217-236.

BRUCE, A.G. and MARTIN, R.D. (1989). Leave-k-out diagnostics for time series. *JRSS B* 51, 363-401.

BUSTOS, O.H. and YOHAI, V.J. (1986). Robust estimates for ARMA models. *J. Am. Statist. Assoc.* 81, 155-168.

BUZUG, T., VON STAMM, J. and PFISTER, G. (1993). Characterization of period-doubling scenarios in Taylor-Couette flow. *Physical Review E* 47, 1054-1065.

CARONI, C. and PRESCOTT, P. (1992). Sequential application of Wilk's multivariate outlier test. *Applied Statistics* 41, 355-364.

O'CARROL, T.M. (1986). Survey of alarms in an intensive therapy unit. *Anaesthesia* 41, 742-744.

CAVANAUGH, J.E. and JOHNSON, W.O. (1999). Assesing the predictive influence of cases in a state space process. *Biometrika* 86, 183-190.

CHAN, W.-S. and CHEUNG, S.-H. (1994). On robust estimation of threshold autoregressions. *Journal of Forecasting* 13, 37-49.

CHAN, K.S. and TONG, H. (1994). A note on noisy chaos. *JRSS Series B* 56, 301-311.

CHANG, I., TIAO, G.C. and CHEN, C. (1988). Estimation of time series parameters in the presence of outliers. *Technometrics* **30**, 193-204.

CHEN, C., and LIU, L.-M. (1993). Joint estimation of model parameters and outlier effects in time series. *J. Am. Statist. Assoc.* 88, 284-297.

CONNOR, J.T. (1996). A robust neural network filter for electricity demand prediction. *Journal of Forecasting* 15, 437-458.

DAUMER, M. and FALK, M. (1998). On-line change-point detection for state space models using multi-process Kalman filters. *Proceedings of the International Linear Algebra Society symposium on fast algorithms for control, signals and image processing*, ed. D. O'Leary, Amsterdam: Elsevier, pp. 125-135.

DAVIES, P.L. and GATHER, U. (1993). The identification of multiple outliers. Invited paper with discussion and reply. *J. Am. Statist. Assoc.* 88, 782-792.

DENBY, L. and MARTIN, D. (1979). Robust estimation of the first-order autoregressive parameter. *J. Am. Statist. Assoc.* 74, 140-146.

DURBIN, J. (1960). The fitting of time series models. *Revue Inst. Int. De Statist.* 28, 233-243.

FAHRMEIR, L., HAMERLE, A. and TUTZ, G. (1996). *Multivariate Statistische Verfahren.* Second Edition. Berlind: De Gruyter.

FLAK, T., SCHMID, W. and SIGMUND, R. (1996). A comparison of several procedures for identifying outliers in contaminated ARMA processes. *Statistics* 2, 175-195.

FOX, A.J. (1972). Outliers in time series. *JRSS B* 3, 350-363.

GATHER, U. and BECKER, C. (1998). Convergence rates in multivariate robust outlier identification. *Results in Mathematics* 34, 101-107.

HARRISON, P.J. and STEVENS, C.F. (1976). Bayesian forecasting. *JRSS B* 38, 205-247.

HARVEY, A.C. and KOOPMAN, S.J. (1992). Diagnostic checking of unobserved components time series models. *Jounal of Economics and Business Statistics* 10, 377-389.

HOTTA, L.K. and NEVES, M.M.C. (1992). A brief review on tests for detection of time series outliers. *Estadistica* 44, 103-148.

IMHOFF, M. (1992). Acquisition of ICU data: concepts and demands. *International Journal of Clinical Monitoring and Computing* **9**, 229-237.

IMHOFF, M. (1995). A clinical information system on the intensive care unit: dream or night mare? *Medicina Intensiva 1995*, ed. J.A.G. Rubi, XXX. Congreso SEMIUC, Murcia.

IMHOFF, M. and BAUER, M. (1996). Time series analysis in critical care monitoring. *New Horizons* 4, 519-531.

IMHOFF, M., BAUER, M., GATHER, U. and LÖHLEIN, D. (1997). Time series analysis in intensive care medicine. *Applied Cardiopulmonary Pathophysiology* 6, 263-281.

JUSTEL, A., PEÑA, D. and TSAY, R.S. (2001). Detection of outlier patches in autoregressive time series. *Stat. Sinica.* 11, 651-673.

KIRKENDALL, N.J. (1992). Monitoring for outliers and level shifts in Kalman filter implementations of exponential smoothing. *Journal of Forecasting* 11, 543-560.

KITAGAWA, G. (1987). Non-Gaussian state-space modeling of nonstationary time series. *J. Am. Statist. Assoc.* 82, 1032-1063.

KIRSCHEN, D.S., BRUCE, F. and WOLLENBERG, B.F. (1992). Intelligent alarm processing in power systems. *Proceedings of IEEE* 80, 663-672.

KLEINER, B. and MARTIN, D. (1979). Robust estimation of power spectra. *JRSS B* 41, 313-351.

LARSSON, J.E. and HAYES-ROTH, B. (1998). Guardian: An intelligent autonomous agent for medical monitoring and diagnosis. *IEEE Intelligent Systems and their applications.* New York: IEEE.

LAWLESS, S.T. (1994). Crying wolf: False alarms in a pediatric intensive care unit. *Critical Care Medicine* 22, 981-985.

LE, M.D., RAFTERY, A.E. and MARTIN, R.D. (1996). Robust bayesian model selection for autoregressive processes with additive outliers. *J. Am. Statist. Assoc.* 91, 123-131.

LEDOLTER, J. (1990). Outlier diagnostics in time series analysis. *Journal of Time Series Analysis* 11, 317-324.

LINKENS, D.A. and NIE, J. (1992). A unified real-time approximate reasoning approach for use in intelligent control: Part 1. Theoretical development. *Int. J. Control* 56, 347-363.

LIU, R.Y. (1995). Control charts for multivariate processes. *J. Am. Statist. Assoc.* 90, 1380-1387.

LJUNG, G.M. (1993). On outlier detection in time series. *JRSS B* 55, 559-567.

MA, Y. and GENTON, M.G. (2000). Highly robust estimation of the autocovariance function. *J. Time Series Analysis* **21**, 663-684.

MAC GREGOR, J.F. (1997). Using on-line process data to improve quality: challenges for statisticians. *International Statistical Review* 65, 309-323.

MÄKIVIRTA, A., KOSKI, E., KARI, A. and SUKUVAARA, T. (1991). The median filter as a preprocessor for a patient monitor limit alarm system in intensive care. *Computer Methods and Programs in Biomedicine* 34, 139-144.

MARTIN, R.D. (1981). Robust methods for time series. *Applied Time Series Analysis II*, ed. D.F. Findley, New York: Academic Press, pp. 683-759.

MARTIN, R.D. and YOHAI, V.J. (1985). Robustness in time series and estimating ARMA models. *Handbook of Statistics* 5, ed. P.R. Krishnaiah, C.R. Rao, Amsterdam: North Holland, pp. 119-155.

—— (1986) Influence fuctionals for time series. *The Annals of Statistics* 14, 781-818.

NAVENDRA, K.S. and PARTHASARATHY, K. (1990). Identification and control of dynamical systems using neural networks *IEEE Transactions on Neural Networks* 1, 4-27.

NOMIKOS, P. and MAC GREGOR, J.F. (1995). Multivariate SPC charts for monitoring batch processes. *Technometrics* 37, 41-59.

NYSCHKA, D., ELLNER, S., MCCAFFREY, D. and GALLANT, A.R. (1992). Finding chaos in noisy systems. *JRSS Series B* 54, 399-426.

OPPENHEIM, A.V. and SCHAFER, R.W. (1995). Discrete-time signal processing. Englewood Cliffs: Prentice-Hall.

PACKARD, N.H., CRUTCHFIELD, J.P., FARMER, J.D. and SHAW, R.S. (1980). Geometry from a time series. *Physical Review Letters* 45, 712-716.

PEÑA, D. (1990). Influential observations in time series. *Journal of Business & Economic Statistics* 8, 235-241.

PEÑA, D. and GUTTMAN, I. (1988). Bayesian approach to robustifying the Kalman filter. *Bayesian Analysis of Time Series and Dynamic Models*, ed. J.C. Spall, New York: Decker, pp. 227-258.

ROUSSEEUW, P.J. (1985). Multivariate estimation with high breakdown point. *Mathematical Statistics and Applications* 8, eds. W. Grossmann, G. Pflug, I. Vincze, W. Wertz, Dordrecht: Reidel,pp. 283-297.

ROUSSEEUW, P.J. and LEROY, A.M. (1987). Robust regression and outlier detection. New York: Wiley.

ROUSSEEUW, P.J. and VAN ZOMEREN, B.C. (1990). Unmasking multivariate outliers and leverage points. *J. Am. Statist. Assoc.* 85, 633-639.

SÁNCHEZ, M.J. and PEÑA, D. (1997). The identification of multiple outliers in ARIMA models. *Statistics and Econometrics Series 27.* Universidad Carlos III de Madrid, Working Paper 97-76.

SCHICK, I.C. and MITTER, S.K. (1994). Robust recursive estimation in the presence of heavy-tailed observation noise. *Annals of Statistics* 22, 1045-1080.

SCHMID, W. (1986). The multiple outlier problem in time series analysis. *Australian Journal of Statistics* 28, 400-413.

SCHWAGER, S.J. and MARGOLIN, B.H. (1982). Detection of multivariate normal outliers. *Annals of Statistics* 10, 943-954.

SEJLING, K., MADSEN, H., HOLST, J., HOLST, U. and ENGLUND, J.-E. (1994). Methods for recursive robust estimaton of AR parameters. *Computational Statistics and Data Analysis* 17, 509-536.

SMITH, A.F.M. and WEST, M. (1983). Monitoring renal transplants: an application of the multiprocess Kalman filter. *Biometrics* 39, 867-878.

STOCKINGER, N. and DUTTER, R. (1987). Robust time series analysis: a robust survey. *Kybernetica*, Supplement 23, 1-90.

TAKENS, F. (1980). Dynamical systems and turbulence. *Vol. 898 of Lecture Notes in Mathematics.* Berlin: Springer.

TAKENS, F. (1981). Detecting strange attractors in turbulence. *Dynamical Systems and Turbulence*, eds. D.A. Rand and L.S. Young. New York: Springer.

TATUM, L.G. and HURVICH, C.M. (1993). A frequency domain approach to robust time series analysis. *New Directions in Statistical Data Analysis and Robustness*, eds. S. Morgenthaler, B. Ronchetti, W.A. Stahel. Basel: Birkhäuser, pp. 195-204.

TONG, H. (1995). A personal overview of nonlinear time series analysis from a Chaos perspective. *Scandinavian Journal of Statistics* 22, 399-421.

TSAY, R.S. (1986). Time series model specification in the presence of outliers. *J. Am. Statist. Assoc.* 81, 132-141.

TSAY, R.S. (1988). Outliers, level-shifts and variance changes in time series. *Journal of Forecasting* 7, 1-20.

WEST, M. (1981). Robust sequential approximate Bayesian estimation. *JRSS B* 43, 157-166.

WEST, M. and HARRISON, P.J. (1986). Monitoring and adaption in Bayesian forecasting models. *J. Am. Statist. Assoc.* 81, 741-750.

WEST, M., PRADO, R. and KRYSTAL, A.D. (1999). Evaluation and Comparison of EEG Traces: Latent Structure in Nonstationary Time Series. *J. Americ. Statist. Assoc.* 94, 1083-1095.

WIERDA, S.J. (1994). Multivariate statistical process control – recent results and directions for future research. *Statistica Neerlandica* 48, 147-168.

YAO, Q. and TONG, H. (1994). On prediction and chaos in stochastic systems. *Phil. Trans. R. Soc. Lond. A* 348, 357-369.