The Identification of Multiple Outliers in Online Monitoring Data

Marcus Bauer*, Ursula Gather*, Michael Imhoff**

* Department of Statistics, University of Dortmund
** Surgical Department, Community Hospital Dortmund

Abstract

We present a robust graphical procedure for routine detection of isolated and patchy outliers in univariate time series. This procedure is suitable for retrospective as well as for online identification of outliers. It is based on a phase space reconstruction 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 transfered into the context of time series which is done here. Some applications to online monitoring data from intensive care are given.

Key words: Multivariate sample, online monitoring, outlier identification, phase space reconstruction, process control, time series.

1 Introduction

Increasing technical possibilities for online recording process data produce manifold challenges for statistical methods. In many 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 (Imhoff, 1992; Mc Gregor, 1997; Kirschen *et al.*, 1992). They allow an improved acquisition and storage of the data in real time. A common aim in all fields where online data are recorded 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 are 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 (Basseville and Nikiforov, 1991; Oppenheim and Schafer, 1995; Larsson and Hayes-Roth, 1998; Linkens and Nie, 1992; Navendra and Parthasarathly, 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 occurence of outliers. They can produce false alerts and if nonrobust procedures are used they can mask level shifts and trends. Also outliers may carry the most important information. Hence there is a necessity for an automatic detection of outliers in online monitoring data.

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, simultanously 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, Imhoff and Bauer (1996).

The existing alarm systems based on fixed threshholds which are chosen by the physician; they produce a great number of alarms due to measurement artefacts, patient movements or minor problems such as transient fluctuations past the set alarm limit (O'Carrol, 1986). Most alarms, about 87% - 94% (Mäkivirta, 1991; Lawless, 1994) are irrelevant in terms of patient care. This poor reliability may lead to critical or even life-threatening situations.

A few typical time series of physiological variables like heart rate, arterial and pulmonal arterial blood pressure recorded in one minute intervals are shown in Figure 1. They contain

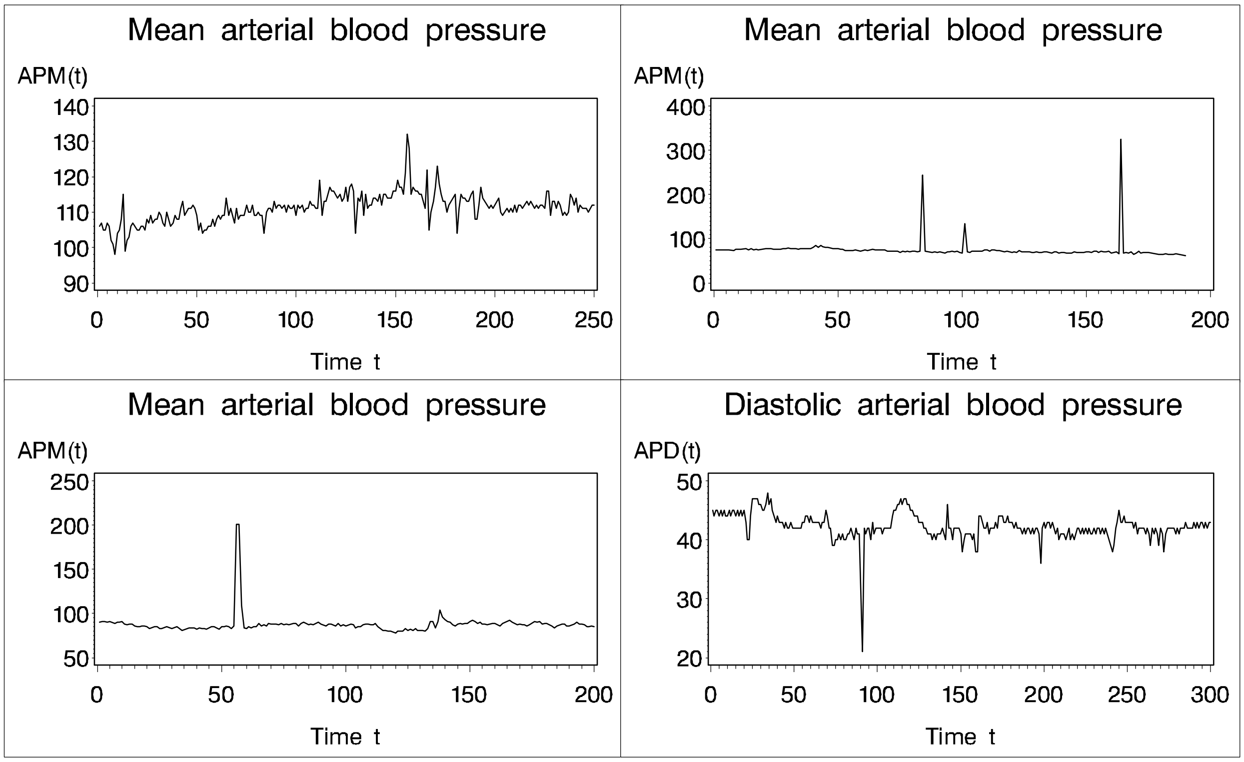

Figure 1: Typical time series of physiological variables

isolated and patchy outliers, which would produce false alarms when using nowadays alarm systems. The aim of this paper is to construct a procedure for the automatic detection and adjustment of outliers which can reduce the false alarm rate in online monitoring systems.
We find an increasing amount of literature in this field since the fundamental article of Fox (1972) and there seem to be two major approaches. Outliers in time series can be represented within the framework of ARIMA models (with batch processing of the data) or within state space models (with sequential processing). It is well known that ARIMA and state space models are related, but outliers have been treated mainly by using ARIMA representations. Furthermore, other areas like nonlinear time series analysis, neural networks, and frequency domain analysis deal with outliers too (see, e.g., Chan and Cheung, 1994; Conner, 1996; Tatum and Hurvich, 1993; Kleiner and Martin, 1979).
Within batch processing the best investigated approach comprises iterative outlier detection and adjustment procedures to obtain joint estimates of outlier effects and model parameters (see, e.g., Chang *et al.*, 1988; Tsay, 1988; Chen and Liu, 1993 as well as Sánchez and Peña, 1997; Justel *et al.*, 1998). A comparison of several of these procedures is given by Flak *et al.* (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 detec-

tion in time series, for instance by Bruce and Martin (1989), Abraham and Chuang (1989), Ljung (1993), Peña (1990), and Ledolter (1990). By robust approaches - as proposed in articles of Denby and Martin (1979), Martin (1981), Martin and Yohai (1985, 1886), Bustos and Yohai (1986) - parameter estimation methods in time series models are modified such that the influence of outliers is reduced or eliminated. An early review on robust methods for univariate ARIMA-models is given by Stockinger and Dutter (1987).

Typical problems of ARIMA-based procedures for detecting outliers are biased estimators of initial parameter values, inappropriate model specification and masking effects, especially if multiple outliers are present (Sánchez and Peña, 1997; Justel *et al.*, 1998; Le *et al.*, 1996; Tsay, 1986; Chen and Liu, 1993). Also, because of lengthy model identification steps, ARIMA based procedures are not really appropriate when analysing online monitoring data. Within the state space approach we find robustifications of the Kalman filter in West (1981), Peña and Guttman (1989) as well as Kitagawa (1987). A review is given by Schick and Mitter (1994). However all these procedures are not robust against multiple or patchy outliers and some procedures fail already if two or more outliers arise within less than w time intervals distance of each other (Schick and Mitter, 1994). Harvey and Koopman (1992) as well as Kirkendall (1992) discuss the detection of outliers in structural models using the Kalman filter, too. All approaches in state space models can be sensitive w.r.t. 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.

State space models, and especially linear dynamic multiprocess models as introduced by Harrison and Stevens (1976) are often used 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 in a linear growth model with multiprocess Kalman filtering is due to Daumer *et al.* (1998) who analyses medical online monitoring time series from anaesthesia. A routine application of such models in intensive care units or operating rooms is not practised yet because of a very strong sensitivity against misspecifications of the hyperparameters, the 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).

This paper choses a different, partially graphical approach which leads to procedures which are able to work online. We proceed as follows. In Section 2 we develop a simple model (called phase space 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 in the context of time dependent data in Section 3 and give a comparison with the classical way of outlier identification in time series in Section 4. The new outlier identifiers are used in Section 5 to construct procedures for detecting outliers in time series retrospectively as well as in an online way. The procedure is applied to online monitoring data from intensive care in Section 6.

2 Phase Space Models

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

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

Often like 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 modelling 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 phase space model - for the underlying process which only affords a few assumptions necessary 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 correspondig identification procedures afford very low computational effort.

The so-called phase space reconstruction is a simple and fundamental tool introduced by Takens (1980, 1981) and at the same time by Packard *et al.* (1980) to analyse nonlinear deterministic, especially chaotic systems.

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:

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

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

with $m, T \ll \mathbb{N}$. Here, m is called the embedding dimension and T is the time delay. The dynamical information of the univariate time series is thus transformed into a spatial information within an m-dimensional space $I\!R^m$, the so-called phase space. The set $\{\vec{\mathbf{y}}_t \mid t = 1, \ldots, N-(m-1)T\}$ 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, e.g. $m > 2d_H + 1$, where d_H is the Hausdorff dimension of the attractor of the original system. For special systems a lower embedding dimension can be chosen (Broomhead, King, 1986); this is confirmed by many applications (e.g. Bezerianos *et al.*, 1995; Buzug et al., 1993).

Also, a suitable choice of T is nontrivial. For nonlinear dynamical systems it is often recommended to choose the time delay T such that $Corr(Y_t, Y_{t-\tau}) = 0 \; \forall \tau \geq T \; \forall t \in \mathbb{Z}$. Some authors prefer a slight correlation. 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 data analytical methods based on phase space reconstructions, 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 with phase space techniques. 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 phase space reconstruction as a tool for analysing linear stochastic processes with the

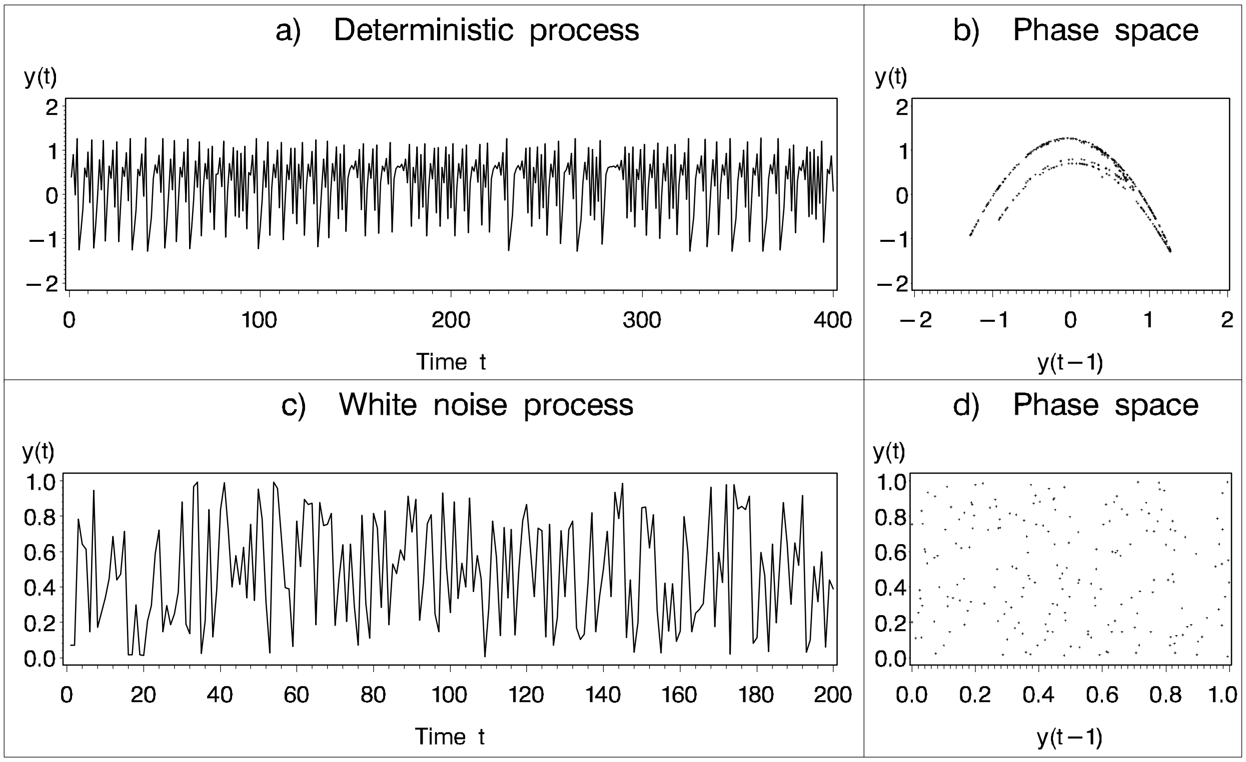

Figure 2: Trajectories and their phase space of a pure deterministic and a pure stochastic process.

special aim of (online-)identification of outliers. The phase space reconstruction could also be used for the detection of level shifts in the phase space. This is not within the scope of the paper, 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:

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

The simulated trajectory with starting value x and its 2-dimensional phase space reconstruction 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 phase space reconstruction is given in Figure 2d, where no structure is visible.

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 phase space 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 phase space reconstruction 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 2c) causes a typical movement through the phase space (Figure 2d).

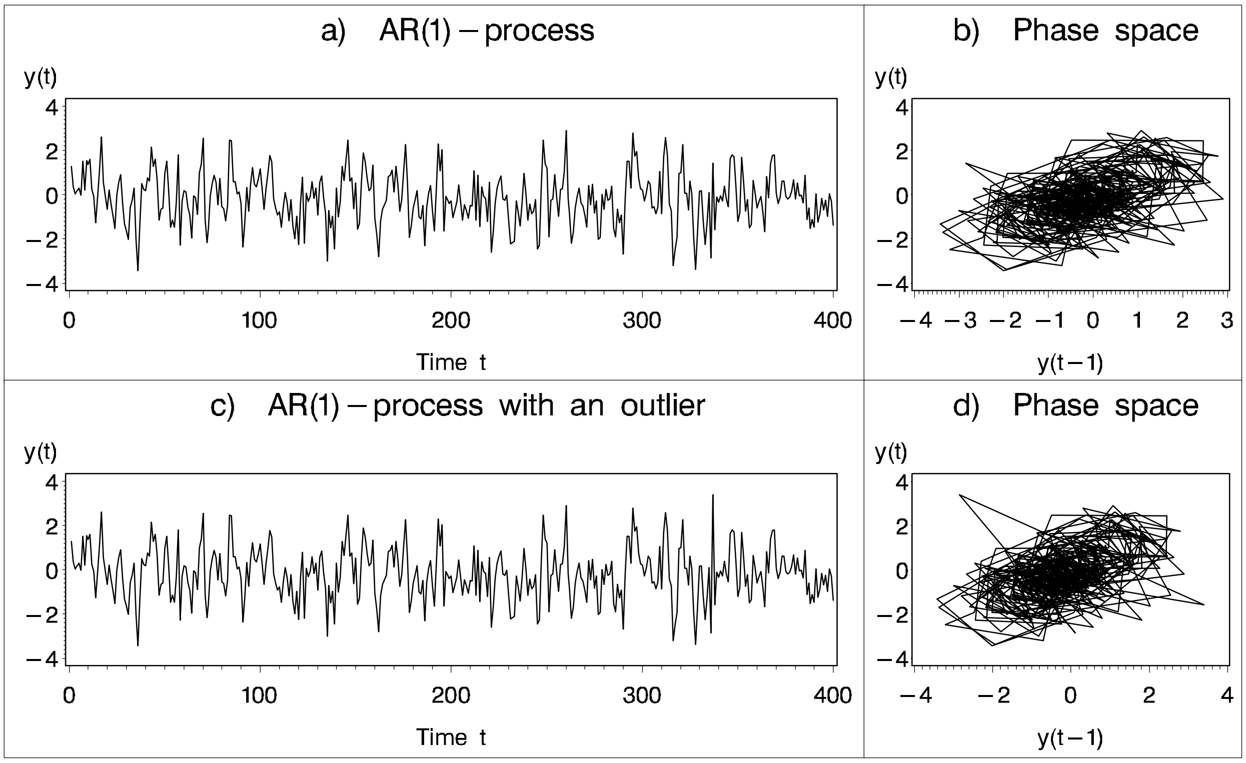

Figure 3: Trajectory of an AR(1)-process and its phase space with and without outlier.

2.1 The model

Extending Example 3 to AR(p) models and higher embedding dimensions m, we get phase space reconstructions located in m-dimensional ellipsoids. The smaller the dependence of consecutive process variables is, the more the shape of the ellipsoid resembles a spherical cloud. Clearly, this is valid in general for stationary Gaussian processes. 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 strucure in linear stochastic processes is lost.

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

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

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

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

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

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

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).

By the assumption of normality for the process variables it follows immediately that the sample phase space vectors follow a multivariate normal distribution. For the choice of m we suggest to take only those process variables into account, which have a direct influence (measured by the PACF) on the present process variable. 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 sample partial autocorrelation function (SPACF) has to be calculated from the data to choose m and therefore, we replace $\rho(\tau)$ by an estimator $\hat{\rho}(\tau)$, typically based on the formula given by Durbin (1960). Then we choose m according to

$$m = 1 + \max_{\tau}\{\tau;\ \mid \hat{\rho}(\tau) \mid > u_{1-\alpha}\sqrt{\frac{1}{N}}\}, \tag{2.1}$$

where $u_{1-\alpha}$ is the $(1-\alpha)$-quantile of a standard normal distribution and $Var(\hat{\rho}(\tau)) = \frac{1}{N}$.

This model allows to interpret a time series as a multivariate sample with identically distributed but non independent observations. Thus, there is a chance for several methods originally constructed for multivariate i.i.d. data to be transfered to the context of a univariate time series. One must only take care of the effects caused by the special dependency structure of the observations.

2.2 Predictions

It is possible to derive predictions on the basis of phase space models. Let $\hat{y}_{t,1}$ be the one-step-ahead prediction of a time series. By geometric considerations the following estimation is natural in an m-dimensional phase space:

$$\hat{Y}_{t,1} = arg \min_x \left(\sum_{i=1}^{m-1} a_i \mid\mid \vec{\mathbf{Y}}_{t,1} - \vec{\mathbf{Y}}_{t-i+1} \mid\mid^2 + a_0 \mid\mid \vec{\mathbf{Y}}_{t,1} - \mu \mid\mid^2 \right), \tag{2.2}$$

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

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

It is easy to see that minimizing (2.2) leads to

$$\hat{Y}_{t,1} = \left(\sum_{i=1}^{m-1} a_i Y_{t-i+1} + a_0 \mu \right) / \sum_{i=0}^{m-1} a_i, \quad \text{when} \sum_{i=0}^{m-1} a_i > 0. \tag{2.3}$$

The predictions are constructed such that the weighted sum of the quadratic distances of the $m-1$ preceding phase space vectors and the mean vector to the new one is minimal. This means that for predicting y_{t+1} only those observations $y_t, \ldots, y_{t-m+2}$ are relevant, 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 well known that the minimum mean square 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

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

The prediction in (2.3) can be compared with the minimum mean squared error forecast for AR-processes. For this purpose let $\sum_{i=0}^{m-1} a_i = 1$, this guarantees also that $\hat{Y}_{t,1}$ minimizes (2.2). 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:

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

The process parameters can be estimated, e.g. by Yule-Walker estimators, calculated from the SPACF by solving an m-dimensional equation system. Because the SPACF is already calculated for choosing m, the additional computational effort to get predictions is not high, if the embedding dimension is low. For example, in Imhoff *et al.* (1997) as well as in Imhoff and Bauer (1996) 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 this application is small, when the model order is low.

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

Let us first explain what we mean by identifying outliers in a multivariate sample. We follow Davies and Gather (1993) as well as Becker and Gather (1997) and formalize the task of identifying outliers by using the concept of so-called α-outliers. Such α-outliers are only characterized by their extreme position in the sample w.r.t. the anticipated distribution. For a multivariate sample an α-outlier w.r.t $N(\mu,\Sigma)$ is just defined as an element of

$$out(\alpha,\mu,\Sigma) := \{x \in \mathbb{R}^m;\ (x-\mu)'\Sigma^{-1}(x-\mu) > \chi^2_{m,1-\alpha}\}, \tag{3.1}$$

the so called α-outlier region w.r.t. $N(\mu,\Sigma)$. For a sample of size N, one can also speak of an α_N-outlier region $out(\alpha_N,\mu,\Sigma)$ according to the condition

$$P(\bigcap_{i=1}^{N}\{X_i \in \mathbb{R}^m \backslash out(\alpha_N,\mu,\Sigma)\}) = 1-\alpha, \tag{3.2}$$

for $X_i \sim N(\mu,\Sigma)$ $(i = 1,\ldots,N)$ and some given $\alpha \in (0,1)$. For an i.i.d. sample this leads to $\alpha_N = 1-(1-\alpha)^{\frac{1}{N}}$. Note, that the α_N-outlier region depends on the unknown parameters μ and Σ. Therefore, the outlier region is typically unknown and has to be estimated from the data and estimating an α-outlier region is equivalent to identifying all α-outliers in a given data set. For this reason α_N outlier identifiers are now defined as follows. Let $\mathbf{x}_N = \{x_1,\ldots,x_N\}$ be a sample of size N, $x_i \in \mathbb{R}^m$ $(i = 1,\ldots,N)$, such that more than half, say $\frac{N}{2} < n \le N$, of the observations come i.i.d (!) from an m-dimensional normal distribution. Then, for $\alpha_N \in (0,1)$, an α_N outlier identifier is defined as a region (depending on the sample)

$$OR(\mathbf{x}_N, \alpha_N) := \{z \in \mathbb{R}^m \mid (z-v)'S^{-1}(z-v) \ge c\}, \tag{3.3}$$

where $S = S(\mathbf{x}_N) \in \mathbb{R}^{m\times m}$ is symmetric and positive definite, $v = v(\mathbf{x}_N) \in \mathbb{R}^m$, and $c = c(m, N, \alpha_N),\ c \in \mathbb{R},\ c \ge 0$. Every point $x \in OR(\mathbf{x}_N, \alpha_N)$ is then "identified" as an α_N outlier with respect to $N(\mu, \Sigma)$. The normalization constant c can be chosen analogously to (3.2)

$$P(X_i \in \mathbb{R}^m \backslash OR(\mathbf{x}_N, \alpha_N),\ i = 1, \ldots, N) = 1 - \alpha.$$

Taking $v = \bar{x}_N$ and $S = S_N$ in (3.3), where S_N is the sample covariance matrix, leads to the classical Mahalanobis-distance as outlier identifier (Barnett and Lewis, 1994, p. 271; Schwager and Margolin, 1982; Caroni and Prescott, 1992; Gather and Becker, 1997). Since in the i.i.d. case

$$MD_i = \sqrt{(x_i - \bar{x}_N)'S_N^{-1}(x_i - \bar{x}_N)},\ \ i = 1, \ldots, N, \quad MD_i^2 \sim \chi^2_m,$$

we set $c = \chi^2_{m,1-\alpha}$ by using (3.2). This is a weighted Euclidian distance, which considers the distance as well as the direction of an observation to the centre of the data. An observation x_i is regarded as outlying if MD_i is larger than $\chi^2_{m,1-\alpha}$.

However, the use of robust estimates for v and S like Minimum Volume Ellipsoid (MVE) estimators (Rousseeuw, 1985) or S-estimators for location and scale is recommended to avoid masking and swamping (Gather and Becker, 1997; Becker and Gather, 1999).

Up to now we have concentrated on the i.i.d. case. Now, following the model in Section 2.1 we consider time series data, which are transformed into a multivariate sample by a phase space reconstruction. Definition (3.1) of the α outlier region w.r.t. $N(\mu, \Sigma)$ can be adopted as it stands, but the dependence of the sample vectors has some consequences here.

For example, if we consider the simple analogue of the classical outlier identifier, based on the above Mahalanobis distance for time series (MDTS)

$$MDTS_t = \sqrt{(\vec{\mathbf{y}}_t - \bar{\vec{\mathbf{y}}}_{N-m+1})'S^{-1}_{Y,N-m+1}(\vec{\mathbf{y}}_t - \bar{\vec{\mathbf{y}}}_{N-m+1})}\ \ (t = 1, \ldots, N-m+1), \tag{3.4}$$

we must investigate its asymptotic properties under the new conditions. In (3.4) $\bar{\vec{\mathbf{y}}}_{N-m+1} = \frac{1}{N-m+1}\sum_{t=1}^{N-m+1} \vec{\mathbf{y}}_t$ is the arithmetic mean of the phase space vectors and $S_{Y,N-m+1}$ is the sample covariance matrix

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

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 $\mu = \bar{y}_N = \frac{1}{N}\sum_{t=1}^{N} y_t$. The following theorem shows that $MDTS_t^2$ is still asypmtotically χ^2-distributed in the dependent phase space situation.

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 $\vec{\mathbf{Y}}_1,\ldots,\vec{\mathbf{Y}}_{N-m+1}$ denote the multivariate normally distributed phase space vectors with mean μ and covariance matrix Σ, then*

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

Proof:

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 $\hat{\mu}_N = \bar{\vec{\mathbf{y}}}_{N-m+1}$ and $\hat{\gamma}_N(h) = \frac{1}{N}\sum_{t=1}^{N-h}(y_t-\mu)(y_{t+h}-\mu)$ are consistent estimators for μ and $\gamma(h)$:

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

It follows that $(\bar{\vec{\mathbf{Y}}}_N, S_{Y,N-m+1})$ are $\sqrt{N}$-consistent estimators for (μ,Σ). Similar to Becker (1997) we consider the function $g(v) := (x-v)'S^{-1}(x-v)$. A Taylor expansion with rest term r yields

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

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

To adjust the outlier region correctly to the case of dependent samples, a choice of $\alpha_N = \frac{\alpha}{N-m+1}$ seems appropriate. This guarantees the inequality

$$P\left(\bigcup_{t=1}^{N-m+1}\left\{\vec{\mathbf{Y}}_t \in out(\alpha_{N-m+1}, \mu, \Sigma)\right\}\right) \\ \leq (N-m+1)P(\vec{\mathbf{Y}}_t \in out(\alpha_{N-m+1}, \mu, \Sigma)) \leq \alpha.$$

On the basis of Theorem 3.1 we can now define the analogue to the classical multivariate outlier identifier for time series.

Definition 1 *Let $\alpha_N \in (0,1)$ and $\mathbf{Y}_{N-m+1} = (\vec{\mathbf{Y}}_1, \ldots, \vec{\mathbf{Y}}_{N-m+1})$ 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*

$$OR_{MDTS}(\mathbf{Y}_{N-m+1}, \alpha_{N-m+1}) := \\ \{z \in \mathbb{R}^m \mid (z - \vec{\bar{\mathbf{Y}}}_{N-m+1})' S^{-1}_{Y,N-m+1}(z - \vec{\bar{\mathbf{Y}}}_{N-m+1}) \geq \chi^2_{m,1-\alpha}\}$$

is called Mahalanobis-type outlier identifier for univariate time series. Further, each vector $\vec{y}_t$ with

$$\vec{y}_t \in OR_{MDTS}(\mathbf{Y}_{N-m+1}, \alpha_{N-m+1})$$

is called α_N- phase space outlier (α_N-outlier-PSV).

The normalization by choosing $c(m, N, \alpha_N)$ is according to (3.2) and α_{N-m+1} is given by $\alpha_{N-m+1} = \frac{\alpha}{N-m+1}$.

An outlier identifier based on the MVE can analogously be defined for time series data, because the dependences in the sample do not influence the MVE. In the following definition $\vec{\mathbf{Y}}_{MVE,N-m+1}$ denotes the center of the MVE and $S^{MVE}_{Y,N-m+1}$ denotes the estimated empirical covariance matrix calculated from the observations in the MVE.

Definition 2 *Under the assumptions of Definition 1*

$$OR_{MVETS}(\mathbf{Y}_{N-m+1}, \alpha_{N-1}) := \\ \{z \in \mathbb{R}^m \mid (z - \vec{\mathbf{Y}}_{MVE,N-m+1})'(S^{MVE}_{Y,N-m+1})^{-1}(z - \vec{\mathbf{Y}}_{MVE,N-m+1}) \geq c(m, N, \alpha_N)\}$$

is called MVE outlier identifier for univariate time series.

The MVE can therefore be used to calculate robust estimators of the process mean and the (partial) autocorrelation function. If these robust estimators are used in the Yule Walker equation, this leads also to robust estimators of the process parameters and therefore to robust predictions.

4 Comparison to classical outlier identification in time series

In the fundamental paper of Fox (1972) two types of outliers are defined: additive and innovative outliers. An additive outlier is considered as 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 mentioned that the interpretation of extraordinary observations depends on the adopted model. Following Fox it is possible to model short-term deviations with ARIMA-models by adding an impuls 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. When the term innovative outlier is used in the state space approach it means nothing else than a level shift, modeled by a contaminated noise in the state equation. In fact, a short-term deviation can also be interpreted as a level shift with a fast decreasing effect and not as a new type of outlier.

If a short term deviation is present, the phase space vectors leave the ellipsoid of the steady state, then move to the main diagonal and fall back on the main diagonal into the ellipsoid (Figure 4). We do not want to consider this as a new type of outlier, because especially in online monitoring situations one should avoid overmodelling. We rather want to identify isolated additive outliers and patchy additive outliers. In some sense a short-term deviation is a special case of the pattern of patchy additive outliers. A further task would be the discrimination of level shifts from isolated and patchy outliers. We restrict here to the automatic detection of additive outliers.

For AR(p)-processes Fox (1972) derived the maximum-likelihood-ratio test for additive and innovative outliers. Chang *et al.* (1988) extended Fox results to general ARIMA(p, d, q)-models

$$\Phi(B)\alpha(B)Y_t = \theta(B)\epsilon_t, \qquad \Phi(B) = (1 - \Phi_1 B - \ldots - \Phi_p B^p)$$
$$\theta(B) = (1 - \theta_1 B - \ldots - \theta_q B^q), \quad \alpha(B) = (1 - B)^{d_1}(1 - B^s)^{d_2},$$

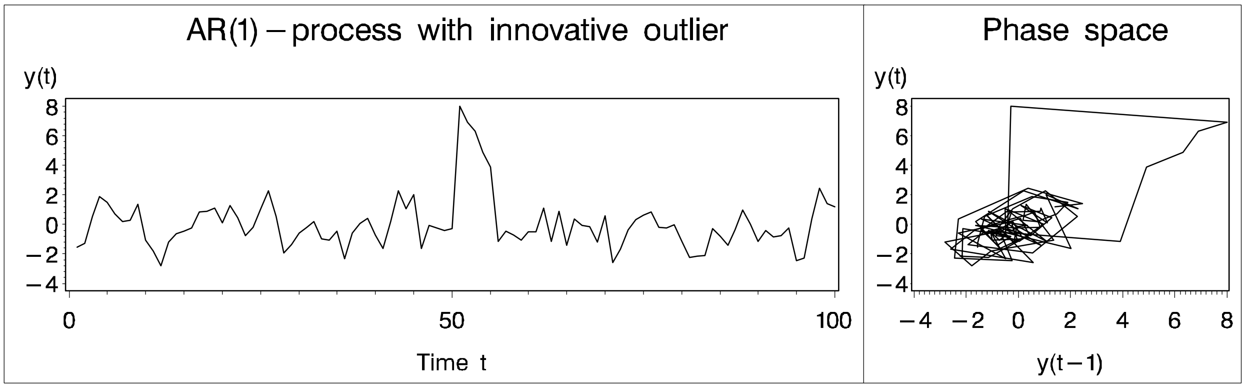

Figure 4: Trajectory of an AR(1)-process and its phase space with an innovative outlier.

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

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

where $\hat{\omega}_{A,t} = \rho^2 \pi(F)\pi(B)y_t$, $\rho^2 = (1 + \pi_1^2 + \ldots + \pi_{N-t})^{-1}$ and F (B) is the forward (backward) shift operator, $\hat{\sigma}_\epsilon$ is an estimator of the residual standard deviation 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

$$\lambda_{A,t} = \frac{y_t - \frac{\Phi_1}{1+\Phi_1^2}(y_{t-1} + y_{t+1})}{\sqrt{\frac{1}{1+\Phi_1^2}}\sigma_\epsilon}, \tag{4.5}$$

because for an AR(1)-process $\pi(B) = \phi(B) = 1 - \Phi_1 B$. In the case of an AR(1)-process the Mahalanobis-distance for time series is

$$MDTS_t = \frac{y_t^2 - 2\Phi_1 y_{t-1} y_t + y_{t-1}^2}{\sigma_\epsilon^2}. \tag{4.6}$$

We can now see, that in general Equations 4.5 and 4.6 do not lead to the same results. In the classical case the test statistic is based on the conditional expectation

$$\lambda_{A,t} = \frac{y_t - E(y_t \mid y_{t-1}, y_{t+1})}{\sqrt{\frac{1}{1+\Phi_1^2}}\sigma_\epsilon}.$$

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, Chan and Liu (1993) realized already: "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

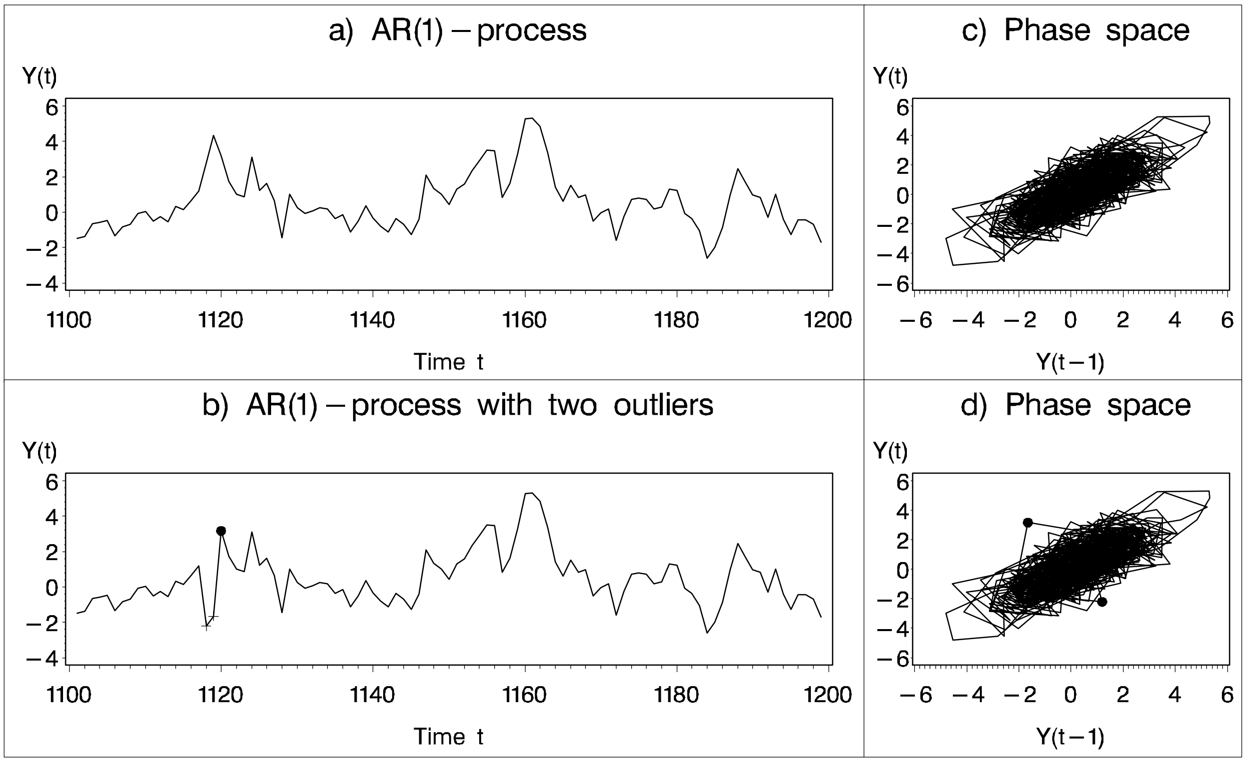

Figure 5: Example of a wrong identification of the time point where an outlier occurs with a classical outlier detection procedure

the example in Figure 5. In Figure 5a a simulated AR(1)-process of time length 2000 with process parameters $\Phi_1 = 0.7$ and $\sigma_\epsilon = 1.0$ is simulated and its 2-dimensional phase space is shown (Figure 5b). In Figure 5c the observations marked with '+' are outliers and are 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 impuls function. In fact using the classical test statistic (4.5) 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 Mahalanobis-distance for time series leads to an identification of observations 1118 and 1120 as α-outlier-PSV. They are marked by '•' in Figure 5d.

Obviously a slight change in y_{t-1} may lead to a larger estimation of the outlier effect $y_t - E(y_t \mid y_{t-1}, y_{t+1}, y_{t-2}, y_{t+2} \ldots)$ and thus the test statistic at time t is larger than a critical value. Using Mahalanobis-distances for time series, $MDTS_{t-1}$ can also be smaller than the critical value and only $MDTS_t$ can pass the limit. But here we have another interpretation: something is wrong in the dynamic at the time points $t-1$ and t, but it is impossible to say which observation is the outlier. This is illustrated in Figure 5d, where it is obvious that changing y_{t-1} or y_t can move the process back into the steady state. Using the test statistic (4.5) to determine the exact time point can lead to definitely wrong decisions. Thus it is better to investigate short time intervals to look for deviations.

5 Identification Procedures

In this chapter we use the above concept to construct procedures for detecting additive outliers in time series. If a phase space vector falls into an empirical outlier region $OR_{MDTS}(\mathbf{Y}_{N-m+1}, \alpha_{N-m+1})$ we have still to decide, which component of the vector is the additive outlier in the time series. As remarked in Section 4 in some cases this problem cannot be solved. In practical situations however a decision is needed.

Usually one additive outlier in a time series generates m outliers in the corresponding phase space sample. Then the additive outlier is the observation, which is exactly that component which arises in all outlying phase space vectors. But there are two special situations, where the decision is more difficult. First, in the situation where $\vec{y}_t \in OR(\mathbf{Y}_{N-m+1}, \alpha_{N-m+1})$ but $\vec{y}_{t-1}, \vec{y}_{t+1} \notin OR(\mathbf{Y}_{N-m+1}, \alpha_{N-m+1})$ it is not clear, which observation is an additive outlier. Second, 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 phase space vectors $\vec{y}_t, \vec{y}_{t+1}, \vec{y}_{t+2}, \vec{y}_{t+3}$ and $\vec{y}_{t+4}$ are located in an outlier region. The same phase space vectors would be α_N-outlier-PSV's, 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 on 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 (2.1) the embedding dimension is chosen. If a priori knowledge is available, the embedding dimension can be regarded as known. If m is thus determined, the dimension of μ and Σ is known too and according to Section 3 outlier identifiers $OR_{MDTS}(\mathbf{Y}_{N-m+1}, \alpha_{N-m+1})$ can be constructed. To determine the outlier component of an α_N-outlier-PSV exactly, one should check then the phase space vectors in chronological order. If a phase space vector $\vec{y}_t$ falls into an outlier region, then y_t is regarded as additive outlier and it is replaced by its forecast $\hat{y}_{t-1,1}$. The one step ahead forecasts $\hat{y}_{t-1,1}$ can be calculated according to section 2.2. Then the next phase space vector is examined and so on.

Since there is little information at the beginning of the time series ($1 \le t \le m$) the outlier identification is complicated. One may consider here vectors $\vec{y}_t = (\hat{\mu}1'_{m-t}, y_1, \ldots, y_t)$ ($t = 1, \ldots, m$). If such a vector falls in an outlier region, then y_t is 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 recursivly and analogously to the retrospective case, only the estimators of the process parameters are updated based on the observations within the current time window $W_{t_{current}}$.

Note that the procedures described above cannot be used for nonstationary time series. In many applications the differenced series $\{d_t\}_{t\in\{2,\ldots,N\}}$, where $d_t = y_t - y_{t-1}$, can be regarded however as stationary. The identification procedure can then be applied to the differenced series. One should only replace d_{t+1} by $y_{t+1} - \hat{y}_{t-1,1}$, if d_t is identified as an additive outlier.

We should now discuss choosing the level α. Similar to control charts from quality control we investigate two possibilities to choose the level α. Either one can use probability limits or some kind of control limits. Choosing probability limits means to choose the level α itself. If the assumptions of normality and stationarity are fulfilled then a condition in terms of the expected number of falsly identified observations as outliers is simply given by

$$E(\text{number of falsly identified AO}) = N_G \alpha_N,$$

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 engineer, physician 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 the variability of the process possibly 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.

Let us assume that $m = 2$. If the process variables Y_t, Y_{t-1} were independent, then using 100k% control limits would lead to a square control region $S = \{(y_1, y_2)|\,(1-k)\mu \leq y_1, y_2 \leq (1+k)\mu\}$, where the process is said to be "in control" at time points $t-1$ and t, when $(y_{t-1}, y_t)' \in S$. But if there are dependences in the process, the correct control region is elliptical, and the process is said to be "out of control" only if $(y_{t-1}, y_t)'$ is outside the ellipse. If $m = 2$ we can inscribe an ellipse into the square S or equivalently for the shifted situation:

$$(Y_{t-1} - \mu, Y_t - \mu) \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & \gamma \\ \gamma & \sigma^2 \end{pmatrix}\right) = N\left(0, \Sigma_2\right).$$

Then the set

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

with

$$f_{\theta,\eta}(y_1, y_2) = \begin{pmatrix} y_1 & y_2 \end{pmatrix} \begin{pmatrix} \theta & \eta \\ \eta & \theta \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}, \qquad \begin{pmatrix} \theta & \eta \\ \eta & \theta \end{pmatrix} = \Sigma_2^{-1}$$

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 in the square S 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

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

The parameters μ, θ and η are unknown and should be replaced by their estimators in praxis. Using a moving time window in online monitoring applications, the current level α_N is calculated with the data of the current time window.

For higher embedding dimension m the choice of α_N can be done analogously. For example if $m = 3$ and $(Y_{t-2} - \mu, Y_{t-1} - \mu, Y_t - \mu) \sim N(0, \Sigma_3)$ one gets

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

where

$$\begin{pmatrix} \theta & \eta & \xi \\ \eta & \theta & \eta \\ \xi & \eta & \theta \end{pmatrix} = \Sigma_3^{-1}.$$

The differences in using a fixed or an adaptive level are investigated in the next section with time series from intensive care medicine.

6 Examples

We apply the proposed procedure to online monitoring data from intensive care medicine. We concentrate on the time series shown in the introduction 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 is succesfully in use for six 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). A short description of the data is given in table 1. The examples presented in this paper are a representative part of a more extensive study with 34 data sets, which contain isolated as well as patchy outliers.

Series	1	2	3	4
Variable	apm	apm	apm	apd
Size	250	190	200	300
Outliers	isolated	isolated	patchy	patchy

Table 1: Investigated time series.

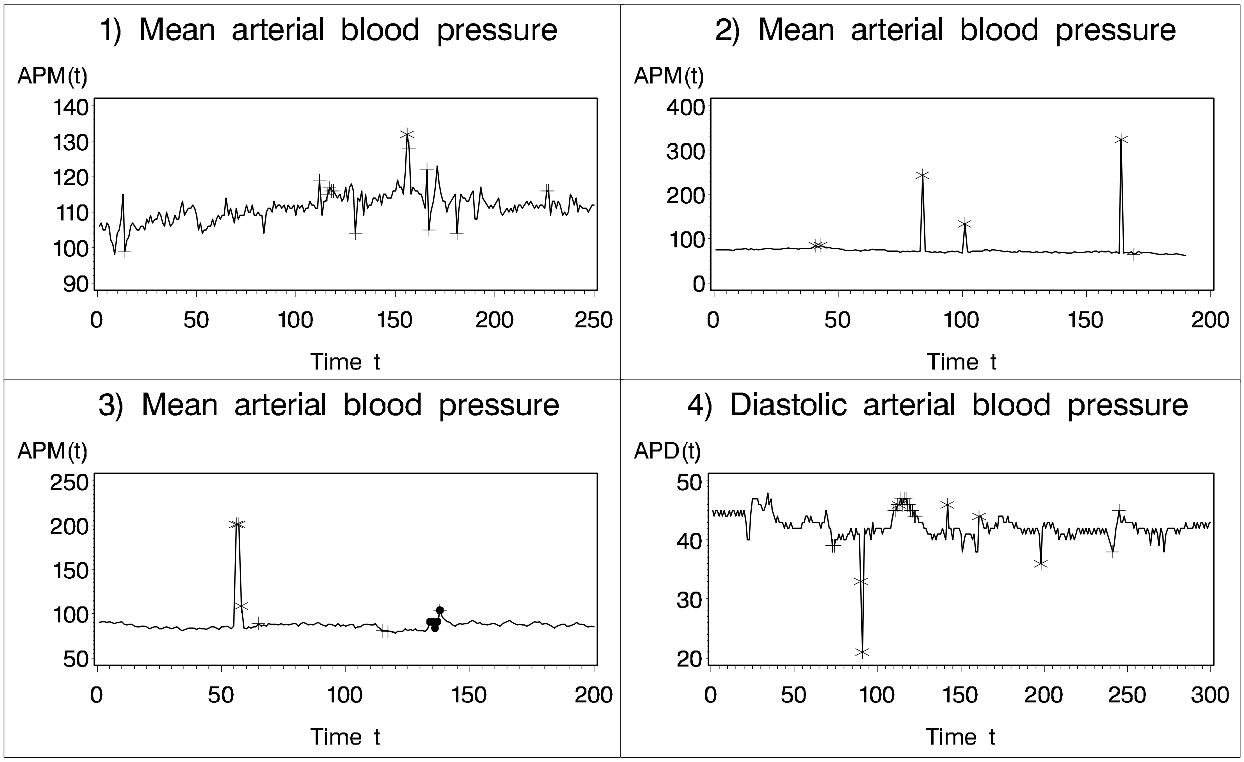

Figure 6: Typical time series of physiological variables

For the four time series depicted in Figure 6 we estimate the PACF on the basis of the first 50 observations. According to (2.1) we get for Series 1, 3 and 4 (Series 2) an embedding dimension $m = 2$ ($m = 3$).

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 level, the physician determines that every observation, which deviates more than 10% from the past level is an artefact, 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 6. Observations which are identified with the fixed as well as the adaptive level are marked with a star $*$. These obsevations, which are only identified by using the fixed (adaptive) level are marked with $+$ ($\bullet$). Crosschecking these identifications with a senior intensivist yielded that all artefacts labeled by him were reliably identified with the procedure. Also the fixed level more observations not labeled by the intensivist were identified as additive outliers than with the adaptive level.

Series 1		fixed level $\alpha_N = 0.000345$	adaptive level	
AO at time point	'real' artefact	$1 - p_t$	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' artefact	$1 - p_t$	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' artefact	$1 - p_t$	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' artefact	$1 - p_t$	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		

Table 2: Identified outliers in the investigated time series.

7 Concluding Remarks

The proposed procedure consists of a combination of the phase space embedding for linear stochastic processes and techniques for the identification of outliers in multivariate data. There is a connection to statistical process control because the presented procedure can be seen as a general Sheward chart for autocorrelated data. In fact, if we have i.i.d. data the embedding dimension is one and thus the outlier region is $out(\alpha, \mu, \sigma^2) = \{x : |x - \mu| > z_{1-\frac{\alpha}{2}}\sigma\}$. The limits $\mu - z_{1-\frac{\alpha}{2}}\sigma$ and $\mu + z_{1+\frac{\alpha}{2}}\sigma$ are exactly the same as for a Sheward chart for one-at-a time data. And thus our procedure is then identical to such a Sheward chart, if no moving window is used.

The procedure seems to be a useful tool for the automatic detection of outliers in online monitoring data. Because of a low computational effort it can be applied to very large sampling rates. Further the robust version is straightforward, because instead of $OR_{MDTS}(\mathbf{Y}_{N-m+1}, \alpha_{n-m+1})$ one can use the $OR_{MVE}(\mathbf{Y}_{N-m+1}, \alpha_{n-m+1})$ or outlier identifiers based on other robust estimators. The procedure may be extended to identify other patterns like level shifts and trends. The effect of such a level shift to 2-dimensional phase space reconstructions of a simulated AR(1) process with level shift is shown in Figure 7. The vectors affected by the level shift form a second ellipse. The different movements of the phase space vectors through the phase space for different patterns might be used to identify and discriminate such patterns.

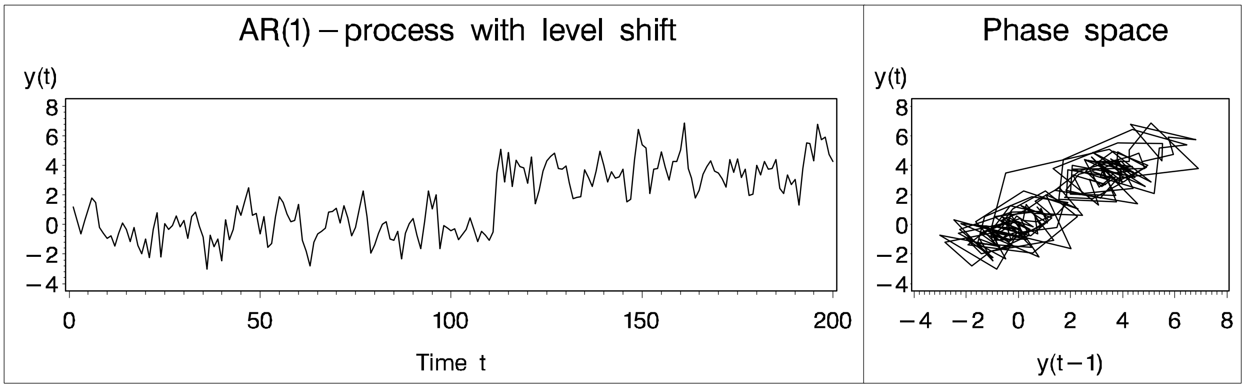

Figure 7: Trajectory of an AR(1)-process and its phase space with an innovative outlier.

Acknowledgements This work was supported by the Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 475.

References

[1] Abraham, B. and Chuang, A. (1989) Outlier detection and time series modeling. *Technometrics* 31, 241-248.

[2] Barnett, V. and Lewis, T. (1994) *Outliers in statistical data* (3rd edn). New York: Wiley.

[3] Basseville, M. and Nikiforov, I.V. (1991) *Detection of abrupt changes: Theory and applications*. Englewood Cliffs: Prentice-Hall.

[4] Bezerianos, A., Bountis, T., Papaioannou, G. and Poldoropoulos, P. (1995) Nonlinear time series analysis of electrocardiograms. *Chaos* 5 (1), 95-101.

[5] Becker, C. and Gather, U. (1997) *The masking breakdown point of multivariate outlier identification rules*. Technical Report, SFB 475. University of Dortmund.

[6] Becker, C. and Gather, U. (1999) *The masking breakdown point of multivariate outlier identification rules*. JASA, in press.

[7] 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.

[8] Broomhead, D.S. and King, G.P. (1986) Extracting qualitative dynamics from experimential data. *Physica* 20D, 217-236.

[9] Bruce, A.G. and Martin, R.D. (1989) Leave-k-out diagnostics for time series. *JRSS B* 51, 363-401.

[10] Bustos, O.H. and Yohai, V.J. (1986) Robust estimates for ARMA models. *JASA* 81, 155-168.

[11] Buzug, T., von Stamm, J. and Pfister, G. (1993) Characterization of period-doubling scenarios in Taylor-Couette flow. *Physical Review E* 47, 1054-1065.

[12] Caroni, C. and Prescott, P. (1992) Sequential application of Wilk's multivariate outlier test. *Applied Statistics* 41, 355-364.

[13] O'Carrol, T.M. (1986) Survey of alarms in an intensive therapy unit. *Anaesthesia* 41(7), 742-744.

[14] Chan, W.-S. and Cheung, S.-H. (1994) On robust estimation of threshold autoregressions. *Journal of Forecasting* 13, 37-49.

[15] Chan, K.S. and Tong, H. (1994) A note on noisy chaos. *JRSS Series B* 56 (2), 301-311.

[16] Chang, I., Tiao, G.C. and Chen, C. (1988) Estimation of time series parameters in the presence of outliers. *Technometrics* 30, 193-204.

[17] Chen, C. and Liu, L.-M. (1993) Joint estimation of model parameters and outlier effects in time series. *JASA* 88, 284-297.

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

[19] Daumer, M. and Falk, M. (1998) On-line change-point detection for state space models using multi-process Kalman filters. In: *Proceedings of the International Linear Algebra Society symposium on fast algorithms for control, signals and image processing*. Ed. D. O'Leary, pp. 125-135. Amsterdam: Elsvier.

[20] Davies, P.L. and Gather, U. (1993) The identification of multiple outliers. Invited paper with discussion and reply. *JASA* 88, 782-792.

[21] Denby, L. and Martin, D. (1979) Robust estimation of the first-order autoregressive parameter. *JASA* 74, 140-146.

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

[23] Flak, T., Schmid, W. and Sigmund, R. (1996) A comparison of several procedures for identifying outliers in contaminated ARMA processes. *Statistics* 2, 175-195.

[24] Fox, A.J. (1972) Outliers in time series. *JRSS B* 3, 350-363.

[25] Gather, U. and Becker, C. (1998) Convergence rates in multivariate robust outlier identification. *Results in Mathematics* 34, 101-107.

[26] Harrison, P.J. and Stevens, C.F. (1976) Bayesian forecasting. *JRSS B* 38, 205-247.

[27] 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.

[28] Hotta, L.K. and Neves, M.M.C. (1992) A brief review on tests for detection of time series outliers. *Estadistica* 44, 103-148.

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

[30] −− (1995) A clinical information system on the intensive care unit: dream or night mare? In: *Medicina Intensiva 1995*. Ed. J.A.G. Rubi. XXX. Congreso SEMIUC, Murcia.

[31] Imhoff, M. and Bauer, M. (1996) Time series analysis in critical care monitoring. *New Horizons* 4 (4), 519-531.

[32] Imhoff, M., Bauer, M., Gather, U. and Löhlein, D. (1997) Time series analysis in intensive care medicine. *Applied Cardiopulmonary Pathophysiology* 6, 263-281.

[33] Justel, A., Peña, D. and Tsay, R.S. (1998) Detection of outlier patches in autoregressive time series. Preprint, Universidad Carlos III de Madrid.

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

[35] Kitagawa, G. (1987) Non-Gaussian state-space modeling of nonstationary time series. *JASA* 82, 1032-1063.

[36] Kirschen, D.S., Bruce, F. and Wollenberg, B.F. (1992) Intelligent alarm processing in power systems. *Proceedings of IEEE* 80 (5), 663-672.

[37] Kleiner, B. and Martin, D. (1979) Robust estimation of power spectra. *JRSS B* 41 (3), 313-351.

[38] Larsson, J.E. and Hayes-Roth, B. (1998) Guardian: An intelligent autonomous agent for medical monitoring and diagnosis. In: *IEEE Intelligent Systems and their applications*. New York: IEEE.

[39] Lawless, S.T. (1994) Crying wolf: False alarms in a pediatric intensive care unit. *Critical Care Medicine* 22 (6), 981-985.

[40] Le, M.D., Raftery, A.E. and Martin, R.D. (1996) Robust bayesian model selection for autoregressive processes with additive outliers. *JASA* 91, 433, 123-131.

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

[42] 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 (2), 347-363.

[43] Ljung, G.M. (1993) On outlier detection in time series. *JRSS B* 55 (2), 559-567.

[44] MacGregor, J.F. (1997) Using on-line process data to improve quality: challenges for statisticians. *International Statistical Review* 65 (3), 309-323.

[45] 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.

[46] Martin, R.D. (1981) Robust methods for time series. In: *Applied Time Series Analysis II*. Ed. D.F. Findley, pp. 683-759. New York: Academic Press.

[47] Martin, R.D. and Yohai, V.J. (1985) Robustness in time series and estimating ARMA models. *Handbook of Statistics* 5, 119-155.

[48] —— (1986) Influence fuctionals for time series. *The Annals of Statistics* 14 (3), 781-818.

[49] Navendra, K.S. and Parthasarathy, K. (1990) Identification and control of dynamical systems using neural networks *IEEE Transactions on Neural Networks* 1 (1), 4-27.

[50] Nyschka, D., Ellner, S., McCaffrey, D. and Gallant, A.R. (1992) Finding chaos in noisy systems. *JRSS Series B* 54 (2), 399-426.

[51] Oppenheim, A.V. and Schafer, R.W. (1995) Discrete-time signal processing. Englewood Cliffs: Prentice-Hall.

[52] 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.

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

[54] Peña, D. and Guttman, I. (1988) Bayesian approach to robustifying the Kalman filter. In: *Bayesian Analysis of Time Series and Dynamic Models*. Ed. J.C. Spall, pp. 227-258. New York: Decker.

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

[56] 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.

[57] Schick, I.C. and Mitter, S.K. (1994) Robust recursive estimation in the presence of heavy-tailed observation noise. *The Annals of Statistics* 22, 1045-1080.

[58] Schmid, W. (1986) The multiple outlier problem in time series analysis. *Australian Journal of Statistics* 28 (3), 400-413.

[59] Schwager, S.J. and Margolin, B.H. (1982) Detection of multivariate normal outliers. *Annals of Statistics* 10, 943-954.

[60] Smith, A.F.M. and West, M. (1983) Monitoring renal transplants: an application of the multiprocess Kalman filter. *Biometrics* 39, 867-878.

[61] Stockinger, N. and Dutter, R. (1987) Robust time series analysis: a robust survey. *Kybernetica*, Supplement, 23, 1-90.

[62] Takens, F. (1980) Dynamical systems and turbulence. In: *Vol. 898 of Lecture Notes in Mathematics*. Berlin: Springer-Verlag.

[63] Takens, F. (1981) Detecting strange attractors in turbulence. In: *Dynamical Systems and Turbulence*. Eds. D.A. Rand and L.S. Young. New York: Springer-Verlag.

[64] Tatum, L.G. and Hurvich, C.M. (1993) A frequency domain approach to robust time series analysis. In: *New Directions in Statistical Data Analysis and Robustness*. Eds. S. Morgenthaler, B. Ronchetti, W.A. Stahel. Basel: Birkhäuser.

[65] Tsay, R.S. (1986) Time series model specification in the presence of outliers. *JASA* 81, 132-141.

[66] −− (1988) Outliers, level-shifts and variance changes in time series. *Journal of Forecasting* 7, 1-20.

[67] West, M. (1981) Robust sequential approximate Bayesian estimation. *JRSS B* 43, 157-166.

[68] West M. and Harrison, P.J. (1986) Monitoring and adaption in Bayesian forecasting models. *JASA* 81, 741-750.

[69] Yao, Q. and Tong, H. (1994) On prediction and chaos in stochastic systems. *Phil. Trans. R. Soc. Lond. A* 348, 357-369.