PATTERN RECOGNITION IN INTENSIVE CARE ONLINE MONITORING

R. Fried
Department of Statistics
University of Dortmund
D-44221 Dortmund
Germany
(fried@statistik.uni-dortmund.de)

U. Gather
Department of Statistics
University of Dortmund
D-44221 Dortmund
Germany
(gather@statistik.uni-dortmund.de)

M. Imhoff
Surgical Department
Community Hospital Dortmund
D-44137 Dortmund
Germany
(mike@imhoff.de)

ABSTRACT

Clinical information systems can record numerous variables describing the patient's state at high sampling frequencies. Intelligent alarm systems and suitable bedside decision support are needed to cope with this flood of information. A basic task here is the fast and correct detection of important patterns of change such as level shifts and trends in the data. We present approaches for automated pattern detection in online-monitoring data. Several methods based on curve fitting and statistical time series analysis are described. Median filtering can be used as a preliminary step to reduce the noise and to remove clinically irrelevant short term fluctuations.

Our special focus is the potential of these methods for online-monitoring in intensive care. The strengths and weaknesses of the methods are discussed in this special context. The best approach may well be a suitable combination of the methods for achieving reliable results. Further investigations are needed to further improve the methods and their performance should be compared extensively in simulation studies and applications to real data.

1 Introduction

In the last three decades there has been a rapid development of the equipment used for monitoring of critically ill patients in intensive care. Clinical information systems (CIS) allow an comprehensive acquisition and storage of this data with high sampling frequencies. Up to 2000 physiological variables, laboratory data, device parameters, medication parameters etc. are recorded in the course of time. The volume of these data frequently exceeds the human ability to assimilate, identify and conceptually relate the observations [1]. While a physician can be confronted with more than 200 variables of the critically ill patient during a typical morning round [2], it is well-known that human beings are not able to develop a systematic response to any problem involving more than seven variables [3]. Moreover, humans are limited in their ability to judge the degree of relatedness between only two variables [4]. Another serious problem is the detection of slow trends in the data. Additionally, personal bias, subjective experience, and a certain expectation toward the respective intervention may distort an objective judgement [5]. Thus electronic bedside decision support offers large potential benefit.

The automatic alarm systems which are currently used are based on fixed thresholds which have to be chosen by the health care professional. They produce a huge number of false alarms due to measurement artifacts, patient movements or transient fluctuations past the set alarm limit. Coughing, turning of and movements of the patient, therapy, blood sampling and flushing of the catheters cause transient artifacts. Most of the alarms, about 90%, are irrelevant in terms of patient care [1], [6], [7]. O'Carrol [8] reports that in a study of 1455 alarms that occurred during three weeks only eight cases were actually life-threatening. While the large number of false alarms could be reduced by choosing sufficiently wide thresholds, this precludes the detection of intervention effects and the problems in detecting slow monotone trends increase even more. The unreliability of fixed threshold alarms may even lead to critical or life-threatening situations.

Clinical decision support aims at providing physicians with therapy guidelines directly at the bedside. The best recommendation possible should be supplied under all circumstances [9]. To achieve this goal quantitative measurements and qualitative reasoning have to be integrated in a system that recommends interventions in real time. One step to reach this aim is to abstract the numerical measurements of the patient's vital signs into qualitative patterns, because usually changes of a variable over time are more important than a single pathological value at the time of observation. The online detection of qualitative patterns such as artifacts, level changes and trends is important for assessing the patient's state.

Several approaches for qualitative data abstraction have been suggested. Haimowitz and Kohane [10], [11] fit trend templates to the data, which are predefined functional forms of relevant patterns. Miksch et al. [12] (1996) propose to measure deviations of measurements from a given target range. Mäkivirta [1] suggests to preprocess the data by a median filter, which is common practice in signal processing.

Statistical intervention analysis is useful for retrospective assessment of the effectiveness of therapeutical actions. In statistical process control, Shewhart, CUSUM- and EWMA-charts are applied. Smith and West [13] and Gather et al. [14] use methods based on statistical time series analysis.

In the following sections we describe several approaches to pattern recognition in a sequence $y_1, \ldots, y_n$ of subsequent observations of a physiological variable measured at equidistant time points with a high sampling frequency such as one observation per minute. We neglect approaches like RESUME [15], [16] which are designed for low-frequency data.

2 Curve fitting

Several curve fitting methods have been developed and applied to clinical problems, the most obvious being the first order approximation with a straight line. Although higher order polynomials have also been used, their applications are limited to cases where such a relation is assumed to exist.

Haimowitz and coworkers [10], [11] developed TrenDx using the concept of trend templates for diagnosing pediatric growth disorders and detecting clinically significant trends in hemodynamics and blood gases in intensive care units. A trend template denotes a time-varying pattern in multiple variables common to a diagnostic population. Predefined patterns of normal and abnormal trends represent disorders as typical patterns of relevant parameters. Trends are diagnosed by matching the observed data to the trend templates.

Each pattern contains representations of landmark events and a set of phases. The trend templates are temporally linked to the patient history. The anchor points used for the trend templates are not necessarily identical to particular events in the patient's history, but may lay within a certain time range around the time point of an event. The phases are represented by a partially ordered set of time intervals with variable endpoints to consider uncertainties. Low-order polynomials constrain the observed values of the variables during each time interval. Hence, these value constraints are parameterized linear regression models describing variation in data assigned to an interval. Haimowitz and coworkers use seven qualitatively distinct elementary regression models corresponding to constant, linear and quadratic models. They consider these seven models to be sufficient to roughly distinguish between different behaviors.

The trend templates are organized in monitor sets which belong to a certain clinical state ("context") of the patient, i.e., all trend templates within a monitor set are assumed to belong to the same patient's state described by the context. One of the monitor sets is the expected or normal model, while the others should warrant the attention of the physician. All members of a monitor set are concurrently matched to the observed data and a matching score is calculated for each template. This score is defined to be the mean absolute percentage error between the observations and regression model estimations. An overall score is obtained for each monitor set using a weighted average of these error scores.

If multiple hypothesized diseases and disease chronologies have to be matched to the data and compared to one another this repeated calculation can become prohibitively expensive. Therefore some simplifying strategies are implemented in the system.

The major disadvantages of TrenDx are the necessary predefinition of the expected normal behavior of the variables during the whole time considered and the usage of absolute value thresholds matching a trend template, which do not take into account the different degrees of parameters' abnormalities. Moreover, the thresholds should be dynamically derived according to the patient's status in the past [12]. TrenDx is designed for trend detection only and does not cover the activation of therapeutic actions and the assessment of the effectiveness of therapeutic actions.

Miksch et al. [12] developed VIE-VENT. In this system knowledge-based monitoring and therapy planning for artificially ventilated newborn infants are integrated. First, the incoming data are validated to arrive at reliable values. Then these values are abstracted into qualitative descriptions, i.e., temporal patterns. The procedure for temporal data-abstraction consists of transformation of quantitative measurements to qualitative values, smoothing of data oscillating near thresholds, smoothing of schemata for data-point transformation, context-sensitive adjustment of qualitative values and transformation of interval data.

The measurements are transformed into qualitative values by dividing the numerical range of each variable into regions of interest and attainable goals. Each of these regions corresponds to a qualitative value. For blood-gas measurements e.g. they use seven qualitative categories of blood-gas abnormalities. The corresponding regions are smaller the nearer the target range since they represent different degrees of abnormalities. The definition of these regions requires specific predefined target values depending on different attainable goals.

Since oscillation of the measurements near a threshold causes rapid oscillation of the qualitative categories smoothing is applied to keep the categories stable. Miksch et al. predefine neighborhoods of the regions of interest and a maximum smoothing activation time period. The size of the neighborhood depends on the size of the region. Smoothing starts if the category for the incoming observation y_t is not the same as the category for the previous measurement and y_t lies within the neighborhood of the previously observed category. In this case the qualitative values of the new observations are set to the category before the change. Smoothing stops if the new observation leaves the neighborhood or the maximal smoothing time is reached.

For different contexts, i.e., different clinical states, different target values are used for the categories. Changing the context would result in a sudden shift of the category resulting possibly in a recommendation for drastic therapeutic intervention. Therefore the thresholds of the schemata are smoothed within a predefined time period, which may be three to eight hours, after changing the context.

Schemata for trend-curve fitting are defined describing the dynamics of different degrees of parameter abnormalities. The qualitative categories are combined with qualitative trend descriptions expected by the physician. In this way qualitative descriptions resulting in an explicit categorization of the change of the variables over time are obtained. Changes of the parameters are modelled by exponential functions, which are piecewise linearized to reduce the complexity. For instance, Miksch et al. consider four different durations of trends. In this way, e.g. the results of therapeutic actions can be assessed.

Miksch et al. tested their approach on simulated and real data of blood-gas measurements. They found the therapeutic recommendations to be consistent and reasonable, except when invalid measurements occurred, but they also considered some limitations of their approach. Information about the frequency of therapeutic interactions in the past would be useful for future reasoning. Data abstraction should include a memory which weights the data since more recently observed data are more important for the reasoning process than data observed in older time periods. Moreover, they found more noise in the real data than expected because of measurement errors, online transmission problems and input from different people in different environments and in different experimental settings.

Both methods presented in this section do not consider correlations within the data and they demand predefinition of expected behavior, which may be hard to specify in advance in critical care. Curve fitting methods behave poorly in the presence of large noise and short term transients which are typical problems in intensive care monitoring [1]. Median filtering helps to overcome this problem.

3 Median filtering

Mäkivirta and coworkers [1], [17] suggest the use of median filters for preprocessing the observations. Median filtering is a basic non-linear method for signal processing which helps to remove noise and transients from the signal without distortion of the signal's baseline. For subsequent observations $y_1, \ldots, y_N$ the median filter of length $2l+1$, $l \in \mathbb{N}$, is defined by

$$\tilde{y}_t = median\{y_{t-l}, \ldots, y_{t+l}\}, \quad t = l+1, \ldots, N-l\,.$$

This filter responds to a change in the signal with a time delay of l observations. On the other hand, this filter tolerates up to l artifacts within a time span of $2l+1$ observations without breaking down completely. Therefore, there is a trade-off between noise attenuation and the time delay. Particularly, these filters have excellent attenuation of impulsive noise since their response to an impulse is zero. Median filters preserve sudden level changes in the signal and diminish the strong noise-like variability with frequent transients. The bias of the filter is related to the variance of the noise and the height of the edge. Therefore, in the presence of excessive noise the ability of a median filter to preserve edges deteriorates.

Mäkivirta [1] noted that increasing the length of the median filter causes a radical decrease of the incidence of both false and true alarms. Therefore, a dual limit alarmer was proposed [17] based on a short and a long median filter. For the short median filter, wide control margins are used to detect sudden critical changes in the patient's state while for the long median filter narrower margins are used to assess therapy effects. Trends can be detected by fitting a linear regression model to the filtered data. In a clinical study, a decrease of about 64 % of the false alarms could be achieved in comparison to usual threshold alarming performed from the unfiltered data.

Several variations of the basic principle of median filtering can be found in the literature. Justusson [18], [19] suggests a weighted median filter giving more weight to the current observation. Nieminen et al. [20], [21] develop median filter based algorithms for real-time trend detection. Their methods are based on so-called FIR median hybrid (FMH) filters which are constructed from several linear subfilters using median operations. These algorithms gradually refine trends when new data becomes available. A particular concern is the detection of slope changes in a trend. This is important because of the frequent transients and sudden changes in the physiological variables of a critically ill patient. Nieminen et al. state that their methods have excellent noise attenuation properties with respect to the delay of the system. However, all these approaches do not include correlations within the data into the analysis.

4 Statistical time series analysis

Endresen and Hill [22] consider methods which are based on the independence of the observations to be not appropriate for the analysis of an observed series from variables such as the heart rate and the blood pressures. When analyzing such time series they found large positive correlations between subsequent measurements. Woodward and Gray [23] point out that high positive correlations within subsequent observations of a time series do often lead to the false detection of a trend. This supports the demand to incorporate autocorrelations within a time series into the analysis as stated by Endresen and Hill [22].

Statistical time series analysis allows to consider autocorrelations, leads to interpretable descriptions of complex underlying dynamics, provides forecasts, gives confidence bounds and allows the assessment of the clinical effects of therapeutic interventions [24], [25]. It has also been shown to be useful for online detection of characteristic patterns in univariate time series. For pattern detection in single variables dynamic linear models [26], ARIMA-models [27], [28] and models based on a phase space type approach [29] have been applied.

In this section we give a brief, but general introduction into statistical time series analysis. Here, the measured sequence of observations (time series) $y_1, \ldots, y_N$ is assumed to be generated from a stochastic process $\{Y_t : t \in \mathbb{Z}\}$, i.e., y_t is regarded as a realization of a random variable Y_t, $t = 1, \ldots, N$. We assume mean and variance to exist for the random law describing the distribution of Y_t and denote them with μ_t and σ_t^2 respectively. Often joint normality of the random variables is assumed. Under this assumption all information about the dependencies between the variables is contained in their correlations. Statistical time series analysis incorporates these dependencies into the reasoning process and uses them to improve decision making based on the observed values. Therefore, the main focus of basic time series models are usually the correlations between the variables while the mean of the random variables Y_t is assumed to be constant over time, i.e., $\mu_t = \mu$ for all $t \in \mathbb{Z}$. This property is called mean-stationarity, while we call the series variance stationary if the variances of the random variables are constant over time, i.e., $\sigma_t^2 = \sigma^2, t \in \mathbb{Z}$. If not only the variances $\sigma_t^2 = Cov(Y_t, Y_t)$, but all covariances $Cov(Y_t, Y_{t+h})$ of the stochastic process are independent of t for all time lags $h \in \mathbb{N}$ we call the series second order stationary and define the autocorrelation function $\rho : \mathbb{N} \to \mathbb{R}$ by

$$\rho(h) = \frac{Cov(Y_t, Y_{t+h})}{Var(Y_t)} .$$

A series which is both mean stationary and second order stationary is called weakly stationary. For the ease of notation, the constant mean μ is usually assumed to be zero. In a practical application we can estimate the mean by the sample mean and subtract it from the series if we assume the underlying stochastic process to be at least mean stationary.

Autoregressive integrated moving average (ARIMA) models are perhaps the most often used time series models, see [30], [31]. An ARIMA(p, d, q) model assumes that the d-times differenced time series follows a weakly stationary ARMA(p, q) model. Differencing means to calculate the increments, i.e. the differences $d_t = y_t - y_{t-1}$ between subsequent observations.

An ARMA(p, q) model assumes that Y_t can be represented as

$$Y_t - \phi_1 Y_{t-1} - \ldots - \phi_p Y_{t-p} = a_t - \theta_1 a_{t-1} - \ldots - \theta_q a_{t-q},\ t \in \mathbb{Z},$$

where $\{a_t : t \in \mathbb{Z}\}$ is a white-noise process of independent identically $N(0, \sigma_a^2)$-distributed random variables ("shocks") and $\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q$ are unknown weights measuring the influence of past observations and past shocks on the current observation. For $q = 0$ we simply call the process AR(p) process, while it is called MA(q) process if $p = 0$. The statistical properties of these models can be exploited using the algebraical theory for polynomials. For this reason the backshift operator B is defined by $By_t = y_{t-1}$. Then $\Phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$ and $\Theta(B) = 1 - \theta_1 B - \cdots - \theta_q B^q$ are polynomials in B of degrees p and q respectively. In ARMA models it is assumed that both polynomials have all their roots outside the unit circle. The former demand follows the assumption of (weak) stationarity of the time series. For an AR(1) model for instance this means that the absolute value $|\phi_1|$ must be smaller than one. An AR(1) model with $\phi_1 = 1$, which has a root located at 1, is a non-stationary ARIMA(0,1,0) model. In this simple model, which is called a random walk, the variance of the observation Y_t increases over time.

Any ARMA model which fulfills the assumptions stated above can also be written in the so-called $AR(\infty)$ form as

$$\Pi(B)Y_t = a_t$$

where $\Pi(B) = \Theta(B)^{-1}\Phi(B) = 1 - \pi_1 B - \pi_2 B^2 - \ldots - \ldots$, or in the $MA(\infty)$ form as

$$Y_t = \Psi(B)a_t$$

where $\Psi(B) = \Phi(B)^{-1}\Theta(B) = 1 - \psi_1 B - \psi_2 B^2 - \ldots$.

ARMA models for a given time series $y_1, \ldots, y_N$ are often constructed using an iterative procedure described in [31]. In this procedure the steps identification, estimation and diagnosis are repeated. An investigator using this procedure should have some experience with it since active interaction is needed.

In the identification step, the orders p and q of the ARMA model are determined. For this reason the sample autocorrelation function SACF

$$\hat{\rho}(h) = \frac{\sum_{t=1}^{N-h}(y_t - \overline{y})(y_{t+h} - \overline{y})}{\sum_{t=1}^{N}(y_t - \overline{y})(y_t - \overline{y})}, \quad \overline{y} = \frac{1}{N}\sum_{t=1}^{N} y_t,$$

and the sample partial autocorrelation function SPACF are analyzed. The former estimates the unknown autocorrelation function of the process, while the latter estimates the autocorrelations between Y_t and Y_{t+h} after elimination of the linear influences of $Y_{t+1}, \ldots, Y_{t+h-1}$ on Y_t and Y_{t+h}. The SPACF can be calculated from the SACF using some recursions [30], [31]. For an AR(p) process the SACF should be exponentially declining and the SPACF should be about zero at all time lags $h > p$. Conversely, for an MA(q) process the SACF should be about zero from time lag $q + 1$ on, while the SPACF should be exponentially declining. For mixed ARMA(p, q) processes the SACF and SPACF show more complex patterns. Their identification often takes several cycles of the procedure.

In the estimation step, the unknown parameters $\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q$ are estimated from the data. While usage of exact maximum likelihood estimators is preferable for short time series with $N \leq 50$ observations, for long time series with $N > 100$ simpler techniques such as conditional maximum likelihood, which is equivalent to conditional least squares under the assumption of normality, usually provide almost identical results. A detailed description of these algorithms can be found in [30] and [31].

In the diagnosis step we check whether the model provides an adequate description of the time series. Commonly the estimated residuals are compared to white noise. The hypothesis of white noise can be tested with the Box-Ljung Q-statistics for instance, while the Durbin-Watson test gives information about possible non-stationarities of the time series. If the model turns out to be satisfactory we can use it for further analysis, otherwise we should modify it according to the impressions gained in the diagnosis stage and then iterate the three steps of the procedure until a satisfactory model is derived.

5 Intervention analysis

The basic assumption of (weak) stationarity is not fulfilled for the physiological time series observed in intensive care. In the contrary, the detection of changes in the patient's state, which may be caused by possibly life-threatening complications or by a change in medicamentation, are the main reason to analyze these data. Moreover, biorhythm causes systematic long-term oscillations, and many measurement artifacts occur in clinical time series. Intervention analysis has been proposed to incorporate patterns of change in statistical time series analysis, and appropriate methods for the retrospective detection of such extraordinary events have been developed. The effects of such events are modelled by using deterministic functions of time for describing a time-varying mean structure. Although these methods cannot be applied online since past and future observations of the variable are used in the analysis, their retrospective application can be useful to assess the effectiveness of therapeutical methods and to construct a knowledge base for future bedside decision support.

Intervention analysis is accomplished via an iterative procedure for detection and removement of patterns of change. Dummy variables are used to model changes in the mean of a time series [32]. Chang et al. [33] propose an iterative procedure based on repeated likelihood ratio tests for outlier detection and parameter estimation which is generalized in Tsay [34] to include level shifts and temporary changes. Nowadays, the procedure of Chen and Liu [35] for outlier detection and parameter estimation in ARIMA models seems to be widely used. Commonly four patterns of change in the mean are considered in statistical time series analysis. These are additive outliers AO, innovational outliers IO, level shifts LS and transient changes TC [36].

An additive outlier AO represents an isolated spike in the time series, which can be caused by an external error, e.g. a measurement artifact, changing the observed value at one particular time point τ without further effects on the future values of the time series. In this case, we observe a modified series $z_1, \ldots, z_N$ instead of $y_1, \ldots, y_N$, which is related to the latter by

$$z_t = \begin{cases} y_t & t \neq \tau \\ y_t + \omega_A & t = \tau \end{cases}$$

Here, ω_A is the unknown effect of the AO. Using the $AR(\infty)$-representation of the ARMA model, we can write the AO model as

$$\Pi(B)(z_t - \omega_A I_t^{(\tau)}) = a_t \, .$$

The dummy variable $I_t^{(\tau)}$ is identical to zero for all t with the only exception $t = \tau$ where $I_\tau^{(\tau)} = 1$. An additive outlier can have serious effects on the estimated residuals of the future observations after time point τ and on the estimated model parameters. The residuals e_t of the observed process after time point τ are related to the residuals of the underlying ARMA process via

$$e_{\tau+j} = a_{\tau+j} - \pi_j \omega_A, \quad j \geq 0\,,$$

where $\pi_0 = -1$. Moreover, it can be proved that a single additive outlier pushes all sample autocorrelations toward zero [36].

An innovational outlier IO can be due to an internal change or endogenous effect which causes an extraordinary shock at some time point τ. The model for an IO is

$$z_t = y_t + \Psi(B)\omega_I I_t^{(\tau)}\,,$$

where ω_I is the unknown size of the IO. Hence, the effects of an IO on the observed time series depend on the ARIMA model. The IO model can be written equivalently as

$$\Pi(B)z_t = a_t + \omega_I I_t^{(\tau)}\,.$$

In case of known model parameters, an IO only affects the residual of the observed process at the time point τ

$$e_\tau = a_\tau + \omega_I,$$

and $e_t = a_t$ otherwise. It is well-known that IOs usually have minor effects on the sample autocorrelations and on the parameter estimates [37]. For a time series which forms white noise AO and IO are equivalent, while an IO in a random walk is equivalent to a level shift.

A level shift LS corresponds to a step change in the mean level of the process at a time point τ. The model for a LS is

$$z_t = \begin{cases} y_t, & t < \tau \\ y_t + \omega_L, & t \geq \tau \end{cases}\,,$$

where ω_L is the size of the step. This model can also be written as [36]

$$\Pi(B)z_t = a_t + \omega_L \Pi(B)(1-B)^{-1} I_t^{(\tau)}\,.$$

A LS affects all residuals after the time point when the shift occurs:

$$e_{\tau+j} = a_{\tau+j} + \ell_j \omega_L, \quad j \geq 0,$$

where $\ell_j, j = 0, 1, \ldots$ are the coefficients of $\ell(B) = \Pi(B)(1-B)^{-1}$. A LS pushes the sample autocorrelations at all time lags to one if there are many observations before and after the shift.

A transitory change TC is a temporary LS that dies out exponentially with rate δ and initial impact ω_T, i.e., it is a level shift that decreases with time and fades to zero, see Tsay [34]. Alternatively we could describe it as a spike that takes a few time periods to disappear. The model for a TC is

$$\Pi(B)z_t = a_t + \omega_T \Pi(B)(1 - \delta B)^{-1} I_t^{(\tau)}.$$

For $\delta = 1$, a TC is identical to a LS, while for $\delta = 0$, a TC is an AO. Therefore, a TC with $0 < \delta < 1$ can be seen as an intervention in between a LS and an AO. Generally, the parameter δ specifies how fast the effect of a TC decreases. Tsay [34] chose the values $\delta = 0.8$ and $\delta = 0.6$ and stated that the results were only slightly different in his applications. In case of high positive autocorrelations within the time series, the TC model is close to an IO model. This can result in some misclassifications [34].

Furthermore, in the physiological variables observed in intensive care slow monotone trends can be found which lead to a modified level of the process. A linear trend can be modelled by a ramp shift outlier RS

$$z_t = \Psi(B)a_t + \omega_R R_t^{(\tau)} \tag{1}$$

$$R_t^{(\tau)} = \begin{cases} 0 & t < \tau \\ t - \tau & t \geq \tau \end{cases} \tag{2}$$

$R_t^{(\tau)}$ is called a ramp effect. This model means a temporary LS over the first differences after time point τ

$$(1 - B)z_t = \Psi(B)(1 - B)a_t + \omega_R (1 - B)^{-1} I_t^{(\tau)}.$$

Piecewise linear trends, i.e., trends with a slope which changes occasionally, can be modelled by several subsequent ramp shift outliers.

For retrospective estimation and testing of intervention effects we first assume that the parameters of the ARMA model for y_t are known. Let $e_t = \Psi(B)z_t$ be the residuals from the observed series given the true model parameters. For an AO, IO, LS and TC the model for the residuals can be written in regression form as

$$e_t = \omega x_t + a_t$$

where we have $\omega = \omega_A$ and $x_t = \Pi(B)I_t^{(\tau)}$ for AO, $\omega = \omega_I$ and $x_t = I_t^{(\tau)}$ for IO, $\omega = \omega_L$ and $x_t = \Pi(B)(1 - B)^{-1}$ for LS, and $\omega = \omega_T$ and $x_t = \Pi(B)(1 - \delta B)^{-1}$ for TC. As the model parameters and therefore x_t are assumed to be known and the residuals a_t are independent, ω can be estimated by ordinary least squares, leading to $\hat{\omega} = \sum_t e_t x_t / \sum_t x_t^2$, with variance $\sigma_a^2 (\sum x_t^2)^{-1}$ [36]. Estimation and testing of an RS can be done by applying the methods designed for an LS to the differenced residuals.

In the AO case this approach leads to

$$\hat{\omega}_A = \rho_A^2 \Pi(F) e_\tau,$$

where F is the forward shift operator defined by $Fz_t = z_{t+1}$, and $\rho_A^2 = (1+\pi_1^2+\ldots+\pi_{N-\tau}^2)^{-1}$. This result corresponds to the fact that all residuals after τ are affected by an AO and therefore all of them carry some information. The variance reads $Var(\hat{\omega}_A) = \rho_A^2\sigma_a^2$.

For an IO only the residual at time τ carries information on ω_I. The estimate is

$$\hat{\omega}_I = e_\tau$$

and the variance is $Var(\hat{\omega}_I) = \sigma_a^2$.

In case of an LS, all residuals after the change point τ are informative. The estimate

$$\hat{\omega}_L = \rho_L^2 \ell(F) e_\tau$$

with $\rho_L^2 = (1+\ell_1^2+\ldots+\ell_{N-\tau}^2)^{-1}$ combines this information in a linear way. It can be shown that this statistic measures the difference of the levels before and after time point τ [36]. The variance of the estimate is $Var(\hat{\omega}_L) = \rho_L^2\sigma_a^2$.

For a TC we have

$$\hat{\omega}_T = \rho_\tau^2 \beta(F) e_\tau$$

where $\rho_T^2 = (1+\beta_1^2+\ldots+\beta_{N-\tau}^2)^{-1}$ and the β_i are the coefficients of $\beta(B) = \Pi(B)(1-\delta B)^{-1}$. The variance is $Var(\hat{\omega}_T) = \rho_T^2\sigma_a^2$.

In order to test whether an outlier of known type $j \in \{AO, IO, LS, TC\}$ has occurred at time point τ, one typically tests

$$H_0 : \omega_j = 0 \text{ versus } H_1 : \omega_j \neq 0$$

using the likelihood ratio method. This criterion leads to comparing the parameter estimate to its standard error. The resulting test statistic

$$\lambda_{j,\tau} = \frac{\hat{\omega}_{j,\tau}}{\rho_{j,\tau}\sigma_a},$$

can be compared with a percentile of the student-t-distribution. We add the index τ to stress that the corresponding statistics are calculated for a particular time point τ.

On the other hand, if the type, but not the location of an outlier is known, a test statistic based on the maximized likelihood ratio $\lambda_j = \max_\tau\{\lambda_{j,\tau}\}$ can be used. The distribution of this maximum is complicated because of the correlations between the $\lambda_{j,\tau}$. Approximations of these distributions based on simulations can be found in Chang et al. [33] and Ljung [38].

For the realistic case that neither the model parameters nor the location or the type of the outliers are known, Tsay [34] suggested an iterative procedure for outlier detection. This procedure consists of specification, estimation, detection and removal cycles. Simulation studies revealed that this procedure seems to work very well in case of isolated outliers [34]. It can be summarized as follows:

1. An ARMA model is fitted to the observed time series using maximum likelihood under the assumption that no outlier has occurred; the model residuals are calculated.

2. The maximum of all likelihood ratio statistics $\lambda = \max_j \lambda_j$ is calculated. If λ is larger than a predetermined constant c we assume that an outlier of the corresponding type has occurred at this particular time point. We define new residuals by subtracting the estimated effects of this outlier from the model residuals.

3. Using these new residuals, a new estimate of the residual variance is obtained, and the likelihood ratios are computed again using the new residuals from step 2. Steps 2 and 3 are repeated until no further outliers are found.

4. In the second stage the sizes of the identified outliers and the parameters of the time series model are estimated jointly by fitting the corresponding model to the observed time series. Using the resulting residuals steps 1 to 4 are repeated until no further outliers are found.

Imhoff et al. [39] employed intervention analysis to assess the effects of therapeutic interventions at known times within short time series of pulmonary target variables. They only considered level shifts which describe relevant, long-term therapeutic effects measuring the change in mean level caused by an intervention. As a result, they could separate effective from non-effective interventions, which allows to reconsider therapeutic strategies.

6 Statistical process control

Statistical process control is a widely applied methodology for controlling industrial manufacturing processes for instance. Control charts such as the Shewhart-, the CUSUM- and the EWMA-chart aim at fast detection of systematic changes, particularly level shifts, in a process. The standard measure of the performance of a control chart is the average run length ARL, which is the expected time until an alarm is given. One has to consider both the in-control and the out-of-control ARL of a chart. The former is the ARL of a process which is under control, i.e., if no shift occurs within the time period considered. The latter is the ARL in the case that a shift by a certain amount occurs. It is a function of the size of this shift. For simplicity the shift is usually assumed to occur at the first observation considered. Obviously, large values of the in-control ARL and small values of the out-of-control ARL are good. One has to find a balance between these goals since increasing the in-control ARL by choosing larger control limits for a special chart means to increase the out-of-control ARL, too. To compare two control charts, usually the in-control ARL of both charts is set to a certain value by adjusting the control limits and then the out-of control ARLs of the charts are considered.

The performance of classical control charts suffers seriously from autocorrelations [40], [41], [42], [43]. Two approaches have been proposed to overcome this problem [42]: Modified control charts compare the original measurements to control limits which are appropriately changed to consider autocorrelations. Forecast-based monitoring schemes consist of a two-stage procedure. First a time series model is fitted to the data and then the one-step ahead forecast residuals are monitored using traditional control charts. If the time series model is correctly specified, the forecast residuals are independent and identically distributed with mean zero during the steady-state, and thus the traditional control chart used in the second step should be reliable. Forecast-based monitoring schemes often employ autoregressive AR(p)-models since they constitute a quite flexible model class describing a wide variety of autocorrelation functions, and since simple computational formulae for prediction and confidence bounds estimation exist. Simulation studies show that the properties of forecast-based monitoring schemes strongly depend on the parameter estimates [42], [43], [44], [45].

The individuals-chart is a special form of the Shewhart-chart for which the individual observations are compared to predetermined upper and lower control limits UCL and LCL. An observation outside the control limits results in an alarm suggesting that some action is required. These control limits are based on a certain allowable deviation from a target value, which may be an estimate of the mean during the steady state, for instance. The allowable deviation is often based on normal theory resulting in $UCL = \mu_0 + ks$ and $LCL = \mu_0 - ks$, where μ_0 is the target value, s is an estimate of the variability found in the undisturbed process, e.g. the standard deviation, and k can be chosen as a certain percentile of the standard normal distribution. Fixed threshold alarms can be regarded as classical individuals-charts where the original observations are monitored. In case of autocorrelated observations, the individuals chart can be applied to the forecast errors of an AR model [46]. Both the classical and the forecast-based individuals charts are not robust against measurement artifacts since an alarm is triggered in case of a single extreme value. Sometimes it is suggested to add a run rule to an individual chart, e.g. "if two out of the last three observations are between two and three standard deviations away from the centerline and on the same side of it, an alarm should be given" [45].

EWMA-charts use exponential weighting methods for suppressing large variability in the monitored variable. The EWMA-chart with weighting factor λ monitors the weighted sum of the observations

$$EM_t = \lambda Y_t + (1 - \lambda) EM_{t-1} \,.$$

An initialization EM_0 is needed which can be set to a target value or a mean value obtained from past data. Then EM_t is compared to appropriately chosen control limits. It is well-known that EM_t describes the best predictor if the observed variable follows an ARIMA(0,1,1) process with parameter $\theta_1 = 1-\lambda$. Although this property could be used to choose the value of λ by fitting an ARIMA(0,1,1) model to the data [47], usually a value $\lambda \leq 0.2$ is chosen [48], [49] since λ defines the influence of the latest measurement. An EWMA-chart with $\lambda = 1$ is an individuals chart. For small values of λ, EM_t is a low-pass filter which diminishes artifact noise and most of the fast fluctuations of the measured signal. Generally, EWMA-charts with large λ perform better for large level shifts, while EWMA-charts with small λ do better for small shifts [42]. Zhang [49] proposed an EWMA-chart with control limits modified to consider correlations within the data, while Montgomery and Mastrangelo [50] developed an EWMA-chart based on the forecast errors of an AR(1) model. EWMA-charts can also detect slow trends [48]. Nevertheless, Mäkivirta [1] stated that there is no evidence that exponential weighting is preferable to any other weighting method, apart from the fact that exponentially weighting allows to use a very convenient recursive structure.

Cumulative sum (CUSUM-)charts [22] accumulate differences between the actual and the expected values of a monitored variable. Timmer, Pignatiello and Longnecker [51] among others constructed a CUSUM-chart for an AR(1) process using a likelihood ratio approach. Two recursive filters, an upper and a lower one, are used for the detection of upward and downward shifts in the mean of an AR(1) process with mean μ_0 during the steady state. Choosing a certain out-of-control value μ_A of the mean, which should be detected with high probability, the upper filter is given by

$$U_t = \max[0, U_{t-1} + (W_t - m_{W_t} - k_t \sigma_{W_t})]\,,$$

where

$$\begin{aligned}
W_t &= \begin{cases} (1+\phi_1)Y_1, & t = 1 \\ Y_t - \phi_1 Y_{t-1}, & t > 1 \end{cases} \\
m_{W_t} &= \begin{cases} (1+\phi_1)\mu_0, & t = 1 \\ (1-\phi_1)\mu_0, & t > 1 \end{cases} \\
k_t &= \begin{cases} (1+\phi_1)k, & t = 1 \\ (1-\phi_1)k, & t > 1 \end{cases} \\
\sigma_{W_t} &= \begin{cases} \sqrt{\frac{1+\phi_1}{1-\phi_1}}\sigma_a, & t = 1 \\ \sigma_a, & t > 1 \end{cases}\,,
\end{aligned}$$

and $k = (\mu_0 + \mu_A)/2$ is the reference value for detection of a shift towards μ_A for an i.i.d. chart. This means that at each time point $t > 1$ the previous value is modified by adding the amount the forecast error is larger than an certain percentile of its steady-state distribution if the resulting value is larger than zero, and by restarting from zero otherwise. This percentile is chosen according to the size of a shift which should be detected very soon with high probability. Additionally, an initialization U_0 has to be chosen. The filter for detecting downward shifts is defined accordingly. By its construction, a CUSUM-chart is not robust against artifacts since an extreme observation has large influence on the filters. Hence, artifacts can either result in false alarms or hinder the fast detection of a level shift.

As pointed out by Lin and Adams [48], a forecast based individuals-chart has high probability of detecting a level shift immediately, but it has low probability of shift detection after the first observation because of the recovery of the forecasts. On the other hand, EWMA charts have low probability of fast signals, but higher probability of detection at subsequent observations. The performance of the individuals chart applied to forecast errors is better than the performance of EWMA-charts in case of large level shifts, while the performance of the individuals chart is inferior to the performance of EWMA- and CUSUM-charts in case of small or moderate shifts [48], [49], [51]. In a comparison of several Shewhart and EWMA-charts, Lu and Reynolds [42] recommended a Shewhart-chart of the observations for the detection of large shifts in case of positive autocorrelations, an EWMA-chart of the residuals for medium shifts and an EWMA-chart of the observations for small shifts. Zhang [49] considered his modified EWMA-chart to be preferable to several other individuals- and EWMA-based charts in case of small and moderate autocorrelations and medium-sized level shifts. In the case of very strong positive autocorrelations and a medium-sized or large level shift he supposed a residuals-chart to be superior. However, Kramer and Schmid [43] concluded that Shewhart-charts based on the original observations perform better than Shewhart-charts of the residuals in case of positive autocorrelations. Run rules can be misleading for autocorrelated data [42]. Adams and Tseng [45] found individuals charts to be more robust against misspecification of the model parameters than EWMA- and CUSUM-charts. Simulations in Bauer et al. [29] provide evidence that the individuals chart for forecast errors is powerful against single outliers, but that an approach described in the next section is better for the detection of multiple outliers and level shifts.

Schack and Grieszbach [52] suggest an adaptive trend estimator based on exponential weighting. However, they apply it for visualization only and do not construct a significance test based on their approach. Montgomery and Mastrangelo [50] constructed a cumulative tracking signal using the sum of forecast errors from an EWMA chart, which is standardized by the mean absolute deviation. They specify critical values for this tracking signal from their experience. However, a cumulative sum is very sensitive against artifacts because of its long memory. The tracking signal developed earlier by Trigg [53], which uses exponential weighting of the forecast errors, has some advantages in this respect. Simulation studies [54] show that Trigg's tracking signal TTS performs better than the cumulative tracking signal for small values of λ w.r.t. the average run length in case of independent measurements. TTS has difficulties to detect level shifts occurring in several steps [22] and does not distinguish between trends and level shifts [55]. Stoodley and Mirnia [56], who also use a cumulative sum of forecast errors provided by an EWMA chart, suggested rules based on the number of subsequent forecast errors having the same sign or being larger (smaller) than a certain value δ to recognize whether a trend, a level shift or an artifact has occurred.

The greatest problem of the usage of statistical control charts for online monitoring in intensive care is the implicit assumption of (weak) stationarity of the data generating mechanism and the existence of a target value for the data. However, in the clinical application of monitoring vital signs it is impossible to determine more than a region of acceptable values for a physiological variable. Moreover, physiological processes are not stationary, but change according to the clinical state and the biorhythm of the patient. The mean value of a physiological variable does not remain fixed all of the time [57]. Rather, the aim in intensive care is to detect and limit extreme variation. Moreover, parameter estimation from past data, which is often suggested for industrial processes, is difficult for vital signs because of fundamental differences between the individuals.

7 Online pattern recognition based on statistical time series analysis

The statistical control charts presented in the previous section are particularly designed for the detection of step changes and assume the existence of a target value the observations should be compared with. Statistical time series analysis, however, allows to estimate a time-varying mean value and to detect unusual changes in this mean. Intervention analysis does not work online since we need future observations of the variable to calculate the likelihood ratio statistics. Generally, pattern recognition will be more reliable the more observations we can use in the analysis. In the following we discuss several strategies for online pattern detection based on statistical time series analysis which can be found in the literature.

In an early attempt to apply statistical time series analysis to online monitoring data, Smith and West [13] used a multiprocess dynamic linear model to monitor patients after renal transplantation. In dynamic linear models (DLMs) [58] the observation X_t at time t is regarded as a linear transform of an unobservable state parameter. This state is assumed to change dynamically in time according to a simple regression model. Particularly, the linear growth model

$$\begin{aligned} X_t &= \mu_t + \epsilon_t \\ \mu_t &= \mu_{t-1} + \beta_{t-1} + \delta_{t,1} \\ \beta_t &= \beta_{t-1} + \delta_{t,2} \end{aligned}$$

is very appealing in the context of physiological variables since its state at time t consists of a level parameter μ_t and a slope parameter β_t which are easily interpretable. In the multiprocess version used by Smith and West different variances of the random observation error ϵ_t and the random changes in evolution $\delta_{t,j}$, $j = 1, 2$, at time t are used for describing the steady state, artifacts, level changes and trends. For pattern classification they calculated the posterior probabilities of these states in a Bayesian framework using a multiprocess Kalman filter. Routine application of this model has not been practiced yet because of its very strong sensitivity against misspecification of the hyperparameters, particularly of the error variances, and its insensitivity against moderate level shifts.

The number of the hyperparameters and the necessary computational effort can be significantly reduced by using a single-process model. Wassermann and Sudjianto [59] constructed a control chart by comparing the current mean estimate to a specified target value. An alarm is triggered in their approach if the abolute difference is larger than a fixed multiple of the estimated standard deviation of the mean. However, this rule does not allow early detection of trends.

All patterns can be detected by assessing the influence of recent observations on the parameter estimates via suitably chosen influence statistics [60], which compare estimates of the state parameters calculated with these recent observations to estimates calculated without them. For a level change and a trend the recent observations should have a large influence on the estimate of the level and the slope parameter respectively, while an outlier should be far from the estimated level.

This approach based on influence statistics was successfully applied for retrospective pattern detection [60, 61]. For online application, however, we have to estimate the hyperparameters from a rather short estimation period. Practical experience shows that difficulties arise if the variability during the estimation period is low, if level changes occur in several little steps and if patterns of outliers occur in short time lags. Little variability during the estimation period causes the detection of outliers and level changes to be too sensitive subsequently. Level changes in several steps are hard to detect since the smoothed level parameter adjusts in each step, so that possibly the influence statistics are not significant at any time. Several outliers in short time lags may either mask each other or may be mistaken for a level change. Nevertheless, all kind of patterns in hemodynamic time series could correctly be identified in most of the cases [62].

Alternatively, for the online recognition of patterns of change in a dynamical system we can model the underlying process during the equilibrium or steady state from past data and use a measure to detect deviations from this steady state. Experience from earlier studies of physiologic variables [28], [39], [63] shows that physiological time series can typically be described adequately by low order AR(p) models during the time interval in between extraordinary events. The choice $p = 2$ seems to be sufficient in most cases. An intuitive rule for the detection of an outlier is to compare the incoming observation to the one-step ahead prediction in such a model, i.e., to use a forecast based Shewhart chart.

Bauer et al. [29] used an alternative approach to develop an automatic procedure for the online detection of outliers and level shifts in time series. They modeled the marginal distribution of m-dimensional vectors of subsequent observations $\mathbf{y}_t = (y_t, y_{t-1}, \ldots, y_{t-m+1})'$. In this way the dynamical information of the univariate time series is transformed into a spatial information within an m-dimensional space and rules for outlier identification in multivariate data can thus be transferred into the time series context. The embedding dimension m should be chosen according to the dependence structure of the underlying process. Since most physiological time series can be described by AR(2) models, $m = 3$ is an obvious choice.

Under the assumption of joint normality of the random variables during the steady-state, the time-delay vectors $\mathbf{y}_m, \ldots, \mathbf{y}_N$ should form an m-dimensional elliptical cloud. A control ellipsoid can be constructed using the Mahalanobis distance

$$MDTS(t) = \sqrt{\big(\mathbf{Y}_t - \bar{Y}_{N-m+1}\big)'\mathbf{S}^{-1}_{Y,N-m+1}\big(\mathbf{Y}_t - \bar{Y}_{N-m+1}\big)},$$

$t = m, \ldots, N$. Here $\bar{Y}_{N-m+1} = \frac{1}{N-m+1}\sum_{t=m}^{N}\mathbf{Y}_t$ is the arithmetic mean of the time-delay vectors and $\mathbf{S}_{X,N-m+1}$ is the sample covariance matrix

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

with $\hat{\gamma}_N(h) = \frac{1}{N}\sum_{t=1}^{N-h}(Y_t - \bar{Y}_N)(Y_{t+h} - \bar{Y}_N)$, $h = 0, \ldots, m-1$, where $\bar{Y}_N = \frac{1}{N}\sum_{t=1}^{N} Y_t$.
Bauer et al. [29] compared this approach in a simulation study to forecast based detection for ARMA(p,q)-models as mentioned above. They find the forecast based detection to perform better for single outliers, while the approach based on the multivariate Mahalanobis distance is preferable for patterns such as level shifts which affect several subsequent observations. This is according to the forecast recovery of one-step ahead prediction. For a patch of outliers or a level shift the approach based on forecasts fails with high probability if the first outlier is not detected and not replaced by a prediction. This deficiency is even more serious in case of biological systems like the health state of a patient, which often shows a step-wise reaction to disturbances and interventions. On the other hand, using the marginal distribution means to judge m subsequent observations simultaneously. The power of the rule based on the Mahalanobis distance should be increasing with the number of subsequent outliers since they move the time-delay vector further out of the control ellipsoid than a single outlier does. Therefore, the rule based on the Mahalanobis distance will be better than an approach based on one-step ahead prediction for patchy outliers and level shifts at the slight expense of lower power against single outliers, which are clinically a much less relevant phenomenon. Robust estimators of the autocovariances and the mean could be used to overcome problems such as swamping and masking effects of outliers [64].

The approaches based on forecasts and the Mahalanobis distance are suitable for the detection of outliers and level shifts, but not for the detection of slow trends. In intensive care, many changes of actual interest have a duration ranging from some minutes to several hours [1].

For retrospective detection of slow trends in time series data regression based models can be used for instance [23]. Here, it is assumed that the observation Y_t at time point t is a measurement of the current process level $\mu_t = \alpha + \beta t$, which is disturbed by autocorrelated noise E_t. For the noise, often an AR(p) model is assumed. Thus, the model reads

$$\begin{aligned} Y_t &= \alpha + \beta t + E_t \,, \\ E_t &= \phi_1 E_{t-1} + \ldots + \phi_p E_{t-p} + a_t \,. \end{aligned}$$

We either can use simple least squares estimators for α and β and adjust their variances for the autocorrelations within the noise process estimated from the regression residuals, or we can use maximum likelihood or robust techniques to estimate all model parameters jointly. Neglecting the autocorrelations increases the probability of erroneous conclusions [23]. Such a regression based approach can be modified for online monitoring by using a moving time window. The test statistic can be calculated for the current time window including the last n observations. An inherent problem is the appropriate choice of the window width n since for trends with different length different window widths should be optimal.

Another problem is that linearity of the trend is assumed since trends which are nonlinear might not be detected this way. In [65] a rule for retrospective detection of any monotone trend which was proposed by Abelson and Tukey [66] for independent data and by Brillinger [67] for time series was adapted to the online monitoring situation. Let μ_t be the time-varying level of the process, which is disturbed by autocorrelated noise E_t, such that

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

Then a weighted sum $\sum_{t=1}^{N} w_t Y_t$ of the observations of the current time window is used to test for any form of monotone change of μ_t during the time interval $t = 1, \ldots, n$, i.e., $\mu_1 \leq \mu_2 \leq \ldots \leq \mu_n$ with $\mu_t < \mu_{t+1}$, or $\mu_1 \geq \mu_2 \geq \ldots \geq \mu_n$ with $\mu_t > \mu_{t+1}$ for at least one $t \in \{1, \ldots, n-1\}$. Since the weights $w_1, \ldots, w_n$ are restricted to have arithmetic mean $\overline{w} = 0$, the weighted sum has mean zero if μ_t is constant over time. The weights are then determined to solve

$$\max_{w} \min_{\mu} \frac{|\sum(w_t - \overline{w})(\mu_t - \overline{\mu})|^2}{\sum(w_t - \overline{w})^2 \sum(\mu_t - \overline{\mu})^2}$$

with $\overline{\mu} = n^{-1} \sum \mu_t$. This means, they are chosen to have a worst case discriminatory power for an extremely unfavorable trend which is as high as possible. This results in

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

and the corresponding worst case is a single step change. Comparing the mean of time delayed moving windows with length m, which is a standard approach to detect systematic differences [68], has lower worst case discriminatory power since it corresponds to using a weighted sum with weights $w_1 = -1/m, \ldots, -1/m, 0, \ldots, 0, 1/m, \ldots, 1/m$. The hypothesis of a constant mean should be rejected in favor of a monotone increasing (decreasing) mean if $\sum_{t=1}^{n} w_t Y_t$ is large (small) in comparison to its variance. During the steady state this variance is equal to

$$Var\left(\sum_{t=1}^{n} w_t Y_t\right) = \sum_{t=1}^{n}\sum_{s=1}^{n} w_t w_s \gamma(t-s)\,, \tag{3}$$

where $\gamma(h), h = 0, 1, \ldots,$ are the autocovariances of the noise process. Hence, parameter estimation can be accomplished easily if we have reliable estimates of these autocovariances. Since a trend has a serious impact on the usual sample autocovariances, we can try to eliminate a (local) linear trend $\alpha + \beta t$ by regression methods first and estimate the autocovariances from the residuals [65]. In a simulation study, suitable critical values for the standardized weighted sum were determined. This procedure seems reliable for the detection of both linear and non-linear trends which are not very slow. Nevertheless, very large positive autocorrelations may cause some problems. Since very large positive autocorrelations result in monotone sequences just like deterministic trends, these mechanisms are hard to distinguish within short time series anyway [23].

8 Conclusion

There are several approaches to pattern detection within physiological time series. Each of them has its own strengths and weaknesses. In intensive care, very complex combinations of several patterns may be observed within rather short time intervals. This complicates the online application of curve fitting methods, for instance, since they require predefined functional forms. Fixed target values can hardly be specified in advance because of fundamental differences between individuals and because of the uncertain temporal development of the patient's state. Therefore, usual control charts and methods developed within other clinical contexts such as mechanical ventilation can hardly be used. Autocorrelations within subsequent measurements should be considered since neglecting them may lead to false conclusions. Another problem is that methods which detect large changes with high probability and with a short time delay as is needed for life-threatening complications may be insensitive against small or moderate shifts. Reliable detection of the latter is important for assessing intervention effects and as an input for knowledge-based bedside decision support [69]. A particularly difficult problem is the fast and correct detection of a slow trend. Mäkivirta [1] stated that the trend detectors developed at that time had little practical use. Moreover, a useful system should not only detect a trend, but it should also be able to quantify it. In view of all these difficulties certainly the best approach for online pattern recognition within physiological time series is to search for a proper combination of several methods [14]. The individual methods should be further refined and improved using ideas from other approaches. For instance, the performance of the method for online detection of outliers [29] was improved for real time series by replacing the usual fixed significance levels by significance levels which adapt to the time-varying variability of the process. Such an adaptive significance level corresponds to a control limit, i.e., a certain allowable deviation from the process level which is fixed by the physician.

For multivariate monitoring, we either can combine the information contained in several variables using logical rules, or we can try to calculate joint control regions. The latter approach suffers from the "curse of dimensionality" [70], i.e., from the large number of unknown parameters which have to be estimated from the data. To fill a high-dimensional sample space we need huge sample sizes, which are rarely available in practice. Physicians usually select one variable out of a group of closely related variables and base their decisions on the patterns found in this variable only. Statistical methods for dimension reduction like factor analysis can be used to compress the relevant information into a few important variables [71]. Graphical models [72] provide additional information to guarantee that the results obtained from dynamic factor analysis are interpretable by the physician. In a case-study it was shown that latent factors can be more adequate for detection of patterns in the observed variables than each single variable [14].

In conclusion, methods for automatic online analysis of physiological variables offer an opportunity for a more reliable evaluation of the individual treatment and lead to intelligent alarm systems. A future task is the construction of intelligent bedside decision support systems. Such a system can be based on techniques for data abstraction as we have outlined here. These techniques could be combined with methods of artificial intelligence which use the patterns found in the statistical analysis to assess the current state of the patient. By classifying these patterns according to existing knowledge gained from physicians and former data analysis [69] the physician in charge might then be given options of how to respond properly.

Acknowledgements

The financial support of the Deutsche Forschungsgemeinschaft (SFB 475, "Reduction of complexity in multivariate data structures") is gratefully acknowledged.

References

[1] Mäkivirta, A. (1989), "Towards reliable and intelligent alarms by using median filters," Research Reports 660, Technical Research Centre of Finland.

[2] Morris, A., and Gardner, R. (1992), "Computer applications," *Principles of critical care*, Hall, J., Schmidt, G., Wood, L. (eds.), McGraw Hill, New York, pp. 500-514.

[3] Miller, G. (1956), "The marginal number seven, plus or minus two: Some limits to our capacity for processing information," *Psychol. Rev.*, vol. 63, pp. 81-97.

[4] Jennings, D., Amabile, T., and Ross, L. (1982), "Informal covariation asessments: Data-based versus theory-based judgements," *Judgment Under Uncertainty: Heuristics and Biases*, Kahnemann, D., Slovic, P., Tversky, A. (eds.), Cambridge University Press, Cambridge, pp. 211-230.

[5] Guyatt, G., Drummond, M., Feeny, D., Tugwell, P., Stoddart, G., Haynes, R., Bennett, K., and LaBelle, R. (1986), "Guidelines for the clinical and economic evaluation of health care technologies," *Soc. Sci. Med.*, vol. 22, pp. 393-408.

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

[7] Wiklund, L., Hök, B., Stähl, K., and Jordeby-Jönsson, A. (1994), "Postanaesthesia monitoring revisited: Frequency of true and false alarms from different monitoring devices," *J. Clin. Anesth.*, vol. 6, pp. 182-188.

[8] O'Carrol, T. (1986), "Survey of alarms in an intensive therapy unit," *Anesthesia*, vol. 41, pp. 742-744.

[9] Morris, A. (1998), "Algorithm-based decision making," *Principles and practice of intensive care monitoring*, J.A. Tobin (ed.), McGraw Hill, New York, pp. 1355-1381.

[10] Haimowitz, I.J., Le, P.P., and Kohane, I.S. (1995), "Clinical monitoring using regression-based trend templates," *Art. Intel. Med.*, vol. 7, pp. 473-496.

[11] Haimowitz, I.J., and Kohane, I.S. (1996), "Managing temporal worlds for medical trend diagnosis," *Art. Intel. Med.*, vol. 8, pp. 299-321.

[12] Miksch, S., Horn, W., Popow, C., and Paky, F. (1996), "Utilizing temporal data abstraction for data validation and therapy planning for artificially ventilated newborn infants," *Art. Intel. Med.*, vol. 8, pp. 543-576.

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

[14] Gather, U., Fried, R., Lanius, V., and Imhoff, M. (2001), "Online monitoring of high-dimensional physiological time series - a case-study," *Estadistica*, to appear.

[15] Shahar, Y., and Musen, M.A. (1993), "RESUME: A temporal-abstraction system for patient monitoring," *Computers and biomedical research*, vol. 26, pp. 255-273.

[16] Shahar, Y., and Musen, M.A. (1996), "Knowledge-based temporal abstraction in clinical domains," *Art. Intel. Med.*, vol. 8, 267-298.

[17] 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 Medicine*, vol. 34, pp. 134-149.

[18] Justusson, B.I. (1978), "Noise reduction by median filtering," *Pro. 4th Int. Joint Conf. Pattern recognition*, Kyoto, Japan, Nov. 1978, pp. 502-504.

[19] Justusson, B.I. (1981), "Median filtering: Statistical properties," *Topics in applied physics, two dimensional signal processing II*, Huang, T.S. (ed.), Springer-Verlag, Berlin.

[20] Nieminen, A., Neuvo, Y., and Mitra, U. (1988), "Algorithms for real-time trend detection," *ICASSP International Conference on Acoust., Speech and Signal Proc. 1988*, IEEE, New York, pp. 1530-1532.

[21] Nieminen, A., Neuvo, Y., and Mitra, U. (1989), "Algorithms for real-time trend detection," *Signal processing*, vol. 18, pp. 1-15.

[22] Endresen, J., and Hill, D.W. (1977), "The present state of trend detection and prediction in patient monitoring," *Intensive Care Medicine*, vol. 3, pp. 15-26.

[23] Woodward, W.A., and Gray, H.L. (1993), "Global warming and the problem of testing for trend in time series data," *Journal of Climate*, vol. 6, pp. 953-962.

[24] Hill, D.W., and Endresen, J. (1978), "Trend recording and forecasting in intensive care therapy," *British Journal of Clinical Equipment*, january, pp. 5-14.

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

[26] Gordon, K., and Smith, A.S.M. (1990), "Modeling and Monitoring Biomedical Time Series," *J. Americ. Statist. Assoc.*, vol. 85, pp. 328-337.

[27] Hepworth, J.T., Handrickson, S.G., and Lopez, J. (1994), "Time series analysis of physiological response during ICU visitation," *West J. Nurs. Res.*, vol. 16, pp. 704-717.

[28] Imhoff, M., Bauer, M., Gather, U., and Löhlein, D. (1998), "Statistical pattern detection in univariate time series of intensive care on-line monitoring data," *Intensive Care Medicine*, vol. 24, pp. 1305-1314.

[29] Bauer, M., Gather, U., and Imhoff, M. (1999), "The identification of multiple outliers in online monitoring data," Technical Report 29/1999, SFB 475, Department of Statistics, University of Dortmund, Germany.

[30] Brockwell, P.J., and Davis, R. A. (1987), *Time Series: Theory and Methods*, 2nd ed., Springer, New York.

[31] Box, G.E.P., Jenkins, G.M., and Reinsel, G.C. (1994), *Time Series Analysis. Forecasting and Control*, 3rd ed., Prentice-Hall, Englewood Cliffs.

[32] Fox, A.J. (1972), "Outliers in time series," *J. Roy. Statist. Soc. Ser. B*, vol. 34, pp. 350-363.

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

[34] Tsay, R.S. (1988), "Outliers, level shifts and variance changes in time series," *J. Forecasting*, vol. 7, pp. 1-20.

[35] Chen, C., and Liu, L. (1993), "Joint estimation of model parameters and outlier effects in time series," *J. Am. Stat. Assoc.*, vol. 88, pp. 284-297.

[36] Peña, D., Tiao, G.C., and Tsay, R.S. (2000), *A course in time series analysis*, Wiley, New York.

[37] Muirhead, C.R. (1986), "Distinguishing outlier types in time series," *J. R. Statist. Soc. Ser. B*, vol. 48, pp. 39-47.

[38] Ljung, G.M. (1993), "On outlier detection in time series," *J. Roy. Stat. Soc. B*, vol. 55, pp. 559-567.

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

[40] Johnson, R. A., and Bagshaw, M. (1974), "The effects of serial correlation on the performance of the CUSUM tests," *Technometrics*, vol. 16, pp. 103-122.

[41] Padgett, C. S., Thombs, L. A., and Padgett, W. J. (1992), "On the α-risks for Shewhart control charts," *Communications in Statistics - Simulation and Computation*, vol. 21, pp. 1125-1147.

[42] Lu, C.-W., and Reynolds, M.R. Jr. (1999), "EWMA control charts for monitoring the mean of autocorrelated processes," *Journal of Quality Technology*, vol. 31, pp. 166-187.

[43] Kramer, H., and Schmid, W. (2000), "The influence of parameter estimation on the ARL of Shewhart type charts for time series," *Statistical Papers*, vol. 41, pp. 173-196.

[44] Tseng, S., and Adams, B.M. (1994), "Monitoring autocorrelated processes with an exponentially weighted moving average forecast," *J. Statist. Comp. Simul.*, vol. 50, pp. 187-195.

[45] Adams, B.M., and Tseng, I.-T. (1998), "Robustness of forecast-based monitoring schemes," *Journal of Quality Technology*, vol. 30, pp. 328-339.

[46] Runger, G.C., and Willeman, T.R. (1995), "Model-based and model-free control of autocorrelated processes," *Journal of Quality Technology*, vol. 27, pp. 283-292.

[47] Hembree, G.B. (1994), "Recursive estimation of the weighting factor for EWMA control charts from autocorrelated data," *ASAProQlPr*, 39-43.

[48] Lin, W.S.W., and Adams, B.M. (1996), "Combined control charts for forecast-based monitoring schemes," *Journal of Quality Technology*, vol. 28, pp. 289-301.

[49] Zhang, N.F. (1998), "A statistical control chart for stationary process data," *Technometrics*, vol. 40, pp. 24-38.

[50] Montgomery, D.C., and Mastrangelo, C.M. (1991), "Some statistical process control methods for autocorrelated data," *Journal of Quality Technology*, vol. 23, pp. 179-193.

[51] Timmer, D.H., Pignatiello, J. Jr., and Longnecker, M. (1998), "The development and evaluation of CUSUM-based control charts for an AR(1) process," *IEEE Transactions*, vol. 30, pp. 525-534.

[52] Schack, B., and Grieszbach, G. (1994), "Adaptive methods of trend detection and their applications in analysing biosignals," *Biometrical Journal*, vol. 36, pp. 429-452.

[53] Trigg, D.W. (1964), "Monitoring a forecasting system," *Operational Research Quarterly*, vol. 15, pp. 271-274.

[54] Gardner, E.S. (1983), "Automatic monitoring of forecast errors," *Journal of Forecasting*, vol. 2, pp. 1-21.

[55] Kennedy, R.R. (1995), "A modified Trigg's tracking variable as an 'advisory alarm' during anaesthesia," *International Journal of Clinical Monitoring and Computing*, vol. 12, pp. 197-204.

[56] Stoodley, K.D.C., and Mirnia, M. (1979), "The automatic detection of transients, step changes and slope changes in the monitoring of medical time series," *The Statistician*, vol. 28, pp. 163-170.

[57] Högel, J. (2000), "Applications of statistical process control techniques in medical fields," *Allg. Stat. Archiv*, vol. 84, pp. 337-359.

[58] West, M., and Harrison, J. (1989), *Bayesian Forecasting and Dynamic Models*, Springer, New York.

[59] Wasserman, G.S., and Sudjianto, A. (1993), "Short run SPC based upon the second order dynamic linear model for trend detection," *Communications in Statistics - Computation and Simulation*, vol. 22, 1011-1036.

[60] Peña, D. (1990), "Influential observations in time series," *J. Business & Economic Statistics*, vol. 8, pp. 235-241.

[61] De Jong, P., and Penzer, J. (1998), "Diagnosing shocks in time series," *J. Americ. Statist. Assoc.*, vol. 93, pp. 796–806.

[62] Gather, U., Fried, R., and Imhoff, M. (2000), "Online classification of states in intensive care," *Festschrift in honor to Hans-Hermann Bock's 60th birthday, Data Analysis, Classification, and Applications*, W. Gaul, O. Opitz, M. Schader (eds.), Springer, Berlin, pp. 413-428.

[63] Lambert, C.R., Raymenants, E., and Pepine, C.J. (1995), "Time-series analysis of long-term ambulatory myocardial ischemia: effects of beta-adrenergic and calcium channel blockade," *Am. Heart J.*, vol. 129, pp. 677-684.

[64] Becker, C., and Gather, U. (2000), "The masking breakdown point of multivariate outlier identification rules," *J. Americ. Statist. Assoc.*, vol. 94, pp. 947-955.

[65] Fried, R., Gather, U., and Imhoff, M. (2001), "The online detection of a monotone trend in a time series," Preprint, Department of Statistics, University of Dortmund, Germany.

[66] Abelson, R.P., and Tukey, J.W. (1963), "Efficient utilization of non-numerical information in quantitative analysis: general theory and the case of simple order," *Ann. Math. Statist.*, vol. 34, pp. 1347-1369.

[67] Brillinger, D.R. (1989), "Consistent detection of a monotonic trend superposed by a stationary time series." *Biometrika*, vol. 76, pp. 23-30.

[68] Härdle, W. (1990), *Applied nonparametric regression*, Cambridge university press, Cambridge.

[69] Morik, K., Imhoff, M., Brockhausen, P., Joachims, T., and Gather, U. (2000), "Knowledge discovery and knowledge validation in intensive care," *Art. Int. Med.*, vol. 19, pp. 225-249.

[70] Friedmann, J.H. (1994), "An overview of predictive learning and function approximation," in *From Statistics to Neural Networks*, V. Cherkassky, J.H. Friedmann, H. Wechsler (eds.), Springer, Berlin et al., pp. 1-61.

[71] Fried, R., Gather, U., and Imhoff, M. (2000), "Some statistical methods in intensive care online monitoring - a review," *Medical Data Analysis*, R.W. Brause, E. Hanisch (eds.), Springer, Berlin, pp. 67-77.

[72] Gather, U., Imhoff, M., and Fried, R. (2000), "Graphical Models for Multivariate Time Series from Intensive Care Monitoring," Technical Report 33/2000, SFB 475, University of Dortmund, 44221 Dortmund, Germany.