Online Classification of States in Intensive Care

Ursula Gather*, Roland Fried*, Michael Imhoff**

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

Abstract

In modern intensive care physiological variables of the critically ill can be reported online by clinical information systems. Intelligent alarm systems are needed for a suitable bedside decision support. The existing alarm systems based on fixed treshholds produce a great number of false alarms, as the change of a variable over time very often is more informative than one pathological value at a particular time point. What is really needed is a classification between the most important kinds of states of physiological time series. We aim at distinguishing between the occurence of outliers, level changes, or trends for a proper classification of states. As there are various approaches to modelling time-dependent data and also several methodologies for pattern detection in time series it is interesting to compare and discuss the different possibilities w.r.t. their appropriateness in the online monitoring situation. This is done here by means of a comparative case-study.
Key words: Online monitoring, time series analysis, state classification, change point detection, ARIMA models, phase space models, dynamic linear models

1 Introduction

In intensive care prompt detection of critical states and of intervention effects is of utmost importance. Most of the bedside decisions are still based on subjective judgement and experience and do not rely on statistical data analysis. Currently a physician may be confronted with more than 200 variables of each critically ill during his morning round (Morris and Gardner (1992)), while an experienced physician may not be able to develop a systematic response to any problem involving more than seven variables (Miller (1956)). Furthermore the existing alarm systems based on fixed treshholds produce a great number of false alarms due to measurement artefacts, patient movements or minor problems such as transient fluctuations past the set alarm limit (O'Carrol (1986)). Usually changes of a variable with time are more important than one pathological value at the time of observation. Hence, the online detection of qualitative patterns like outliers, level changes, or trends in physiological monitoring data is an important goal in medical time series analysis. In this paper we compare several statistical methods which could reach this goal.

In autoregressive models (AR) (Box et al. (1994)) each variable is expressed as a finite, linear aggregate of previous observations plus a stochastic term. Several authors have successfully applied AR models in the field of critical care (Imhoff and Bauer (1996)), in longitudinal physiological experiments (Lambert et al. (1995)), as well as in studies on laboratory data of the chronically ill (Imhoff et al. (1997)). It has been shown that usually autoregressive processes of low order are suitable for physiological variables. Pattern detection can be accomplished by comparing new observations with prediction bounds calculated by an AR model, which has to be specified automatically.

The phase space (PS) approach was originally introduced for nonlinear systems. Complex deterministic, especially chaotic systems can be analyzed by transforming the observed series into a Euclidean space. Several authors judge the complexity of heart rate dynamics by measures which are based on such a phase space embedding (Faustmann and Ganz (1994), Garfinkel et al. (1992), Khadra et al. (1997)). In the context of intensive care phase space models were introduced by regarding the phase space embedding as a multivariate sample of dependent observations (Bauer et al. (1998)). Consequently multivariate outlier identifiers (Becker and Gather (1999)) can be used for pattern detection then.

In dynamic linear models (DLM) (West and Harrison (1989)) the current value of the process is a linear transform of an unobservable state parameter and a random observation error. The state parameter is assumed to follow a first order AR model. In an early application Smith and West (1983) used a multiprocess version of the linear growth model, which can be formulated as DLM, for monitoring patients after renal transplant (see also Smith et al. (1983)). This procedure requires high computational power and is not very reliable in pattern identification (Gordon (1986), Stronegger (1991), Daumer and Falk (1998)). Alternatively de Jong and Penzer (1998) suggest pattern detection by assessing the influence of observations on the parameter estimates.

In general it has been shown that time series techniques are suitable for retrospective analysis of physiological variables (see the references mentioned above or also Hill and Endresen (1978), Gordon and Smith (1990), Hepworth et al. (1994)). In the following we extend a case-study described in Imhoff et al. (1998) by the inclusion of dynamic linear models. After describing the data set, we give some background information for each of the models and a description of their use for state classification. Finally we present and discuss the results of the case-study.

2 The data

On the surgical intensive care unit of the Community Hospital Dortmund, a 2000 bed teaching hospital, online monitoring data was acquired from 19 critically ill patients (eight female, eleven male, mean age 65 years) with extended hemodynamic monitoring requiring pulmonary artery catheters, in one minute intervals from a standard clinical information system. These data were transferred into a secondary SQL database and exported into standard statistical software for further analysis.

From a total of 550,000 single observations of seven variables (heart rate and invasive blood pressures), segments of 150 to 500 observations for each variable were visually classified by a senior intensivist into five clinically relevant patterns: no change, presence of outlier, temporary level change, permanent level change, and trend. The intensivist had not to define any objective criteria, why he chose a specific classification. From a total of 134 time series 23 were classified as without change, 35 as containing outliers, 10 as showing a trend pattern, and 24 and 42 as containing temporary and permanent level changes respectively. The time series were presented to the intensivist a second time in different order for reclassification without any different result. The same segments were analyzed with second order autoregressive (AR(2)), phase space (PS) and dynamic linear models (DLM).

In the following let $x_1, \ldots, x_N$ be a time series consisting of observations of a physiological variable at equidistant time points $t = 1, \ldots, N$. As usual we denote the corresponding random variables by capitals $X_1, \ldots, X_N$.

3 Autoregressive models

An autoregressive model for a time series formally resembles a multiple regression. A stochastic process $\{X_t : t \in \mathbb{Z}\}$ is called an autoregressive process of order p, denoted by AR(p), if

$$X_t = \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p} + \epsilon_t \text{ for all } t \in \mathbb{Z},$$

where $\phi_1, \ldots, \phi_p$ are unknown weights measuring the influence of preceding values on X_t. The variables $\epsilon_t, t \in \mathbb{Z}$, are assumed to stem from a white noise process, which is a sequence of uncorrelated variables from a fixed distribution with mean zero and time invariant variance. In most cases ϵ_t is assumed to be normally distributed (see Box et al. (1994) for more details).

Typically, physiological variables can be modeled by AR processes with short memory $p \leq 2$ in a satisfactory manner (Lambert et al. (1995), Imhoff and Bauer (1996), Imhoff et al. (1997)). For our data preliminary tests with classical interactive model selection showed, that either first or second order models were statistically appropriate. Second order autoregressive models were chosen for all cases as slight overdetermination is better than underdetermination if there are not too few observations. An extensive model selection process is not possible in online monitoring and has to be avoided.

Each time series was split into two segments, an estimation period (observations $x_1, \ldots, x_n$, average length 173 minutes) and a prediction period ($x_{n+1}, \ldots, x_N$, average length 123). An AR(2)-model with weights $\hat{\phi}_1$ and $\hat{\phi}_2$ was fitted to the data from the estimation period by conditional least squares. Prediction intervals were constructed for both the estimation period (one-step predictions) and the prediction period (h-step predictions) on the basis of the estimated weights.

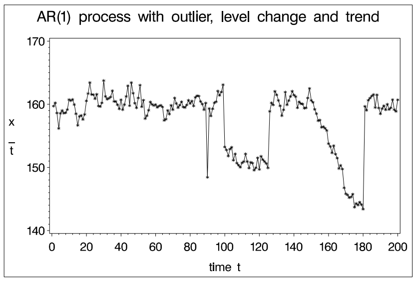
Pattern detection was done by comparing the actual observations to the prediction intervals (PI) for the prediction period. According to the number of values outside the PI, the pattern was classified into the different categories. Values outside the PI were classified as an outlier, if less than 5 consecutive observations (= minutes) were outside the PI, while a level change was identified by 5 or more consecutive observations outside the PI. A level change was called temporary, when less than 50% of the observations of the prediction period were outside the PI, permanent otherwise.

A simulated time series is shown in Figure 1. The data represent an AR(1) process, where one large and some small outliers have been inserted as well as a temporary level change and a trend period. The first 60 minutes are used for estimation and predicting the subsequent observations. Obviously the outliers and the level change are rapidly detected by comparison with the prediction bounds, while the detection of the trend takes longer in spite of the steep slope.

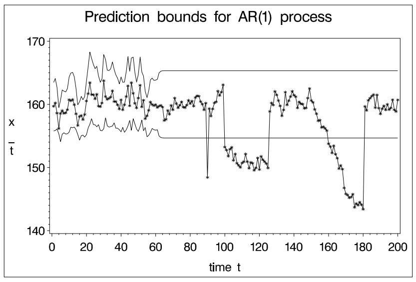
Hence, trend patterns were identified indirectly by deviations of the autocorrelation function (ACF) of the residuals and the Durbin-Watson-statistics. In this case, the ACF of the original series was analyzed for typical trend patterns. If corresponding signs could be found, an AR(2) model was fitted to the first differences of the series. If this model showed a sufficient goodness of fit, a significant trend was assumed.

A suitable choice of the best PI width for pattern detection was unknown before the study. In some cases a difference between the visual classification and the percentage of observations identified as outliers by the model occurred, if a 95% PI was chosen. One possible reason for this difference is a temporary violation of model assumptions like stationarity or the Gaussian distribution of the observations. The other important reason is that a fixed level for the PI cannot adapt to changes of the process variability. For a fixed level, small variability implies a small PI. Observations outside such a PI may be clinically irrelevant. Therefore, the PI level was adjusted a-posteriori from an initial 95% to 99% (strategy L1), 99.9%, 99.99% (together:

Figure 1: Simulated Time Series



AR(2) process with large outlier at $t = 80$, temporary level change between $t = 100$ and $t = 125$ and steep trend between $t = 150$ and $t = 180$.



The same series with upper and lower prediction bounds for $\alpha = 5\%$. The outlier and the temporary level change are rapidly detected by the PI, while for the trend it takes some time.

strategy L2), or in some cases to 90% (strategy H). The analysis was run again and pattern identification was done one more time with this adjusted PI.

4 Phase space models

Phase space models are based on a transformation of the series $x_1, \ldots, x_N$ into an m-dimensional Euclidean space R^m, the so-called phase space, by constructing phase space vectors $\vec{x}_t$:

$$\vec{x}_t := (x_t, x_{t+1}, x_{t+2}, \ldots, x_{t+(m-1)})' \in \mathbb{R}^m,$$

with $m \in \mathbb{N}\backslash\{0\}$, $t = 1, \ldots, N - (m-1)$ and $m \ll N$. Here, m is called the embedding dimension. By this technique, which is derived from the theory of nonlinear dynamic systems (Packard et al. (1980), Takens (1980)), the dynamical information of a series is transformed into a spatial information.

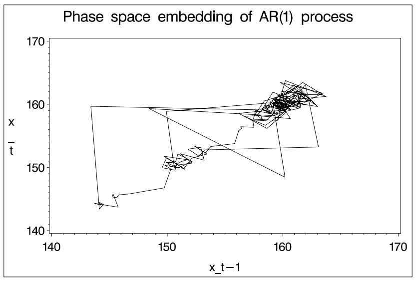
For choosing m there are numerous rules for nonlinear models, and in most applications the components of the phase space vectors are separated by a time delay. For linear Gaussian processes, Bauer et al. (1999) recommended to choose m similarly to choosing the order of an AR(p) model. They defined the components of $\vec{x}_t$ to be chronological observations with a time delay (lag) of always one, as dependencies between consecutive observations should be considered for pattern identification.

For $m = 2$ the set of vectors can be plotted in a two-dimensional space. The corresponding vector cloud, which is called *phase space embedding*, visualizes properties of the underlying dynamic. For stationary linear Gaussian processes (corresponding to a steady state) the vectors form an elliptic cloud. Its form reflects the dependency structure of the process. The centre and the shape of the ellipse are determined by the unknown mean vector $\vec{\mu}$ and the covariance matrix $\mathbf{\Sigma}$ of the vectors $(X_t, X_{t+1})'$. For estimating these parameters either the classical or robust estimators of the mean and the autocovariances of a time series can be used (Bauer et al. (1998, 1999)).

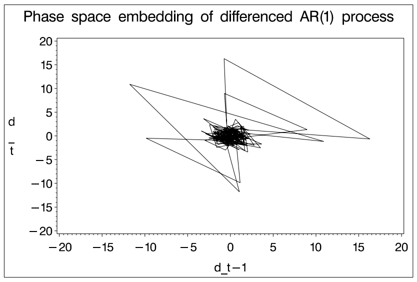
The distance of the vector (x_t, x_{t+1}) from the mean vector $\vec{\mu} = (\mu, \mu)$ gives information about its deviation from the steady state. The Mahalanobis distance (MD) can thus be used for identifying such spurious observations. Taking

$$MD_t = \sqrt{(\vec{x}_t - \vec{\mu})' \mathbf{\Sigma}^{-1} (\vec{x}_t - \vec{\mu})},$$

Figure 2: Simulated Time Series - II



Phase space embedding of the simulated series. The steady state corresponds to the big ellipse, the introduced outlier to the large triangle with points $(160, 145)$ *and* $(145, 160)$*, the temporary level shift to the small ellipse and the trend to the movement along the diagonal.*



After taking differences, the trend is no longer visible. The outlier and the return to steady state after the trend are represented by the large triangles with two points outside the ellipse and the temporary level change by two triangles with points extruding for change and return.

MD_t^2 is χ_2^2 distributed. The set of vectors $\vec{x}$ given by

$$\{\vec{x} \in \mathbb{R}^2 : (\vec{x} - \vec{\mu})'\mathbf{\Sigma}^{-1}(\vec{x} - \vec{\mu}) = \chi^2_{2,1-\alpha}\}$$

forms an ellipse around the mean vector $\vec{\mu}$, where $\chi^2_{2,1-\alpha}$ is the $1-\alpha$-quantile of a χ^2 distribution with two degrees of freedom for a given level α.

In practice the unknown parameters have to be replaced by estimators. By the choice of α, which determines the size of the critical ellipse, the probability for the procedure to falsely identify one or more values as outliers can be kept below a given probability.

If all observations lie inside the estimated ellipse, it can be said that the system is in a steady state. If one or more vectors extrude from the ellipse, a disturbance can be assumed. Disturbances can be distinguished by the movement of the affected vectors in the phase space. The left hand side of Figure 2 shows the phase space embedding of the simulated series from Figure 1. Subsequent vectors are connected by a line to visualize the movement through the phase space. The steady state corresponds to the large ellipse, the introduced outlier to a triangle leaving the ellipse, the level change to a second, smaller ellipse and the trend to the movement along the diagonal.

Our identification procedure uses the differenced time series $d_t = x_t - x_{t-1}, t = 2, \ldots, N$. The vectors $\vec{d_t}$ are analyzed in consecutive order whether they extrude from the cloud. On the basis of the movement of outlying phase space vectors a discrimination between different patterns is done. An abrupt level change in the original series results in one single outlier in the differenced series. Thus, if the vector $\vec{d_t}$ extrudes from the cloud, a distinction between an outlier and an abrupt level change is possible by the location of $\vec{d}_{t+1}$ (see Bauer et al. (1998) for a detailed description). The right hand side of Figure 2 shows the phase space embedding of the first differences of the simulated series.

In our case-study the first 60 observations were taken and analyzed retrospectively (i.e., outlying regions were estimated and patterns in this time interval were identified). Thereafter a time window of length 60 was moved through the data. That means, that at time point $t = 61, \ldots$, it was determined, if the phase space vector was in an outlying region. If not, then no pattern was deteced, and the ellipse was replaced by a new one, which was estimated from the last 60 observations $d_{t-59}, \ldots, d_t$. If the phase space vector $\vec{d_t}$ was found in a distant region, it was concluded that the system was not in a steady state, and after analyzing the consecutive vectors $d_{t+1}, d_{t+2}, \ldots$, it was decided which pattern was present. In this case only the aberrant observations have to be replaced by their predictions for continuing the monitoring procedure.

5 Dynamic Linear Models

Similar to autoregressive models, dynamic linear models (DLM) are defined by sequential parametric equations (West and Harrison (1989)). The idea is, that the observation X_t is a linear transformation of an unobservable state parameter $\vec{\theta}_t$ plus an error term. For these states an $AR(1)$ structure with possibly time-varying, but known parameters is assumed. Smith and West (1983) use a linear growth DLM for monitoring patients after renal transplant:

$$\begin{aligned} X_t &= \begin{pmatrix} 1 & 0 \end{pmatrix} \vec{\theta}_t + \epsilon_t \\ \vec{\theta}_t &= \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \vec{\theta}_{t-1} + \vec{\delta}_t \end{aligned}$$

Here, the states are 2-variate parameters $\vec{\theta}_t = (\mu_t, \beta_t)'$, where μ_t is the process level and β_t is the slope, i.e., the change in level at time t. Furthermore, $\epsilon_t \sim N(0, V_t)$ is the random observation error and $\vec{\delta}_t \sim N(\vec{0}, \mathbf{W}_t)$ is the random change in "evolution" at time t, respectively. An initialization $\vec{\theta}_0 \sim N(\vec{0}, \mathbf{W}_0)$ is needed to start the process.

The proper a-priori specification of the covariances is difficult. Smith and West deduced from theoretical considerations the general form

$$\mathbf{W}_t = \begin{pmatrix} \sigma_1^2 + \sigma_2^2 & \sigma_2^2 \\ \sigma_2^2 & \sigma_2^2 \end{pmatrix}.$$

They used a multiprocess model with different time invariant values $V^{(j)}$, $\sigma_{1,j}^2$ and $\sigma_{2,j}^2$, $j = 1, \ldots, 4$, for describing steady state, level change, slope change and outlier, which were defined by preliminary empirical trials. Classification was done by online calculation of the a-posteriori probabilities of the states for each time t. In a similar way Daumer and Falk (1998) used a conjugate sequential updating procedure. Experience shows that overparametrization and insufficient sensibility may occur in pattern detection since the model may turn out to be too flexible.

Other authors propose to assess the influence of groups of observations on the parameter estimates for (retrospective) diagnostics and detection of structural changes in a single-process model. Normand and Tritchler (1992) suggested the directed Kullback-Leibler divergence for diagnostic purposes. However, this is a summary measure for all changes, e.g. also changes in variability and, therefore, difficult to interpret. Harvey and Koopman (1992) proposed also in a retrospective setting to use auxiliary residuals of the observations, level and slope parameters, which are calculated via the smoothed parameters. De Jong and Penzer (1998) extended this approach by ideas of Peña (1990) and suggested Cook (1977) type influence statistics of the smoothed parameters, which are based on deletion diagnostics. Along the same lines we use the statistics

$$\begin{aligned} D_0^{t,h,k} &:= (x_t - \mu_{t|t+k}) V_{t|t-h}^{-1} (x_t - \mu_{t|t+k}) \\ D_1^{t,h,k} &:= (\mu_{t|t+k} - \mu_{t|t-h}) w_{t,1|t-h}^{-1} (\mu_{t|t+k} - \mu_{t|t-h}) \\ D_2^{t,h,k} &:= (\beta_{t|t+k} - \beta_{t|t-h}) w_{t,2|t-h}^{-1} (\beta_{t|t+k} - \beta_{t|t-h}), \end{aligned}$$

where $\mu_{t|\tilde{t}}$ and $\beta_{t|\tilde{t}}$ are the estimated parameter values at time t in the light of the information $D_{\tilde{t}}$ gained until time $\tilde{t}$. In the same way $V_{t|t-h}$, $w_{t,1|t-h}$ and $w_{t,2|t-h}$ are the predicted variances of X_t, μ_t and β_t respectively. $D_0^{t,h,k}$ is a normalized deviation from the process level, while $D_1^{t,h,k}$ and $D_2^{t,h,k}$ are normalized changes in the level and in the slope respectively.

The unknown covariances were fitted to the observations of an estimation interval using an ad hoc approach tested before with simulated data. We did one step of an EM-algorithm towards the ML-estimates from the starting values $\mu_0 = \overline{x}_n$, the mean of the estimation interval, $\beta_0 = 0$, $V = s_T^2$, the standard deviation in the estimation interval, and a non-singular covariance matrix $\mathbf{W}$ chosen by some preliminary trials. This procedure turned out to be rather robust against misspecification of $\mathbf{W}$. This is due to the large improvements of the estimates in the early steps of an EM-algorithm.

Several strategies were considered w.r.t. the estimation intervals. Interval lengths of 30 and 60 minutes were applied, and both possibilities of estimating the parameters only once for the whole series (as in the AR approach) and moving a time window of 30 (60) minutes through the series (like in the PS approach) were tried out.

A level change at time t induces a significant change in the level parameter and therefore large values of $D_1^{t,h,k}$ for $h,k > 0$. Similarly, trends can be characterized by slope changes and thus by large $D_2^{t,h,k}$. An outlier implies a large value of $D_0^{t,h,k}$. However, the distinction of the patterns is not easy. An outlier also results in a large $D_1^{t,1,0}$ and a level change typically also implies large $D_2^{t,h,k}$. Figure 3 illustrates the influence statistics for the simulated data. By preliminary trials with simulated and real data we found the following slightly complicated rules for pattern recognition, where $D_{i,j}^{h,k}$ is the j-largest value of $D_i^{t,h,k}$ in the estimation period. Classify the state at time t as

1. level change LC,
   if the following four conditions are all fulfilled:
   a) the one-step normalized level influence is large: $D_1^{t-1,1,0} \geq D_{1,q}^{1,0}$,
   b) during the last five minutes at least six out of eight multi-step level influences are large: $D_1^{t,i,5} \geq D_{1,3}^{i,5}$ or $D_1^{t+i,i+1,5} \geq D_{1,3}^{1,0}$ $(i = 1, \ldots, 4)$,
   c) neither x_{t-1} is an outlier nor x_{t+1} has returned to the steady state (see 4.),
   d) $|\mu_{t+5|0} - \mu_{t|0}| < |\mu_{t+5|0} - \mu_{t-1|0}|$.

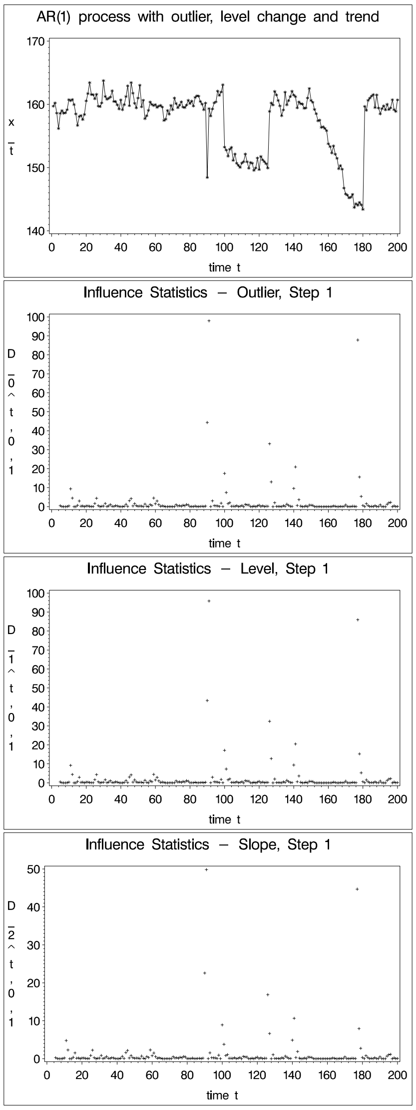
2. temporary level change TLC,
   if within half of the time after the change, say at $t+i$, another change or a trend occurs, such that $min\{|\mu_{t+i|t+i} - \mu_{t+j|t+j}|, j = 0,1\} > min\{|\mu_{t+i|t+i} - \mu_{t-j-1|t-j-1}|, j = 0,1\}$.

3. trend,
   if within the last 40 minutes there is no level change and in at least 80 % of the cases we have $D_2^{t-i,j,20} \geq D_{2,3}^{j,20}, i = 0, \ldots, 10, j = 0, \ldots, 4$.

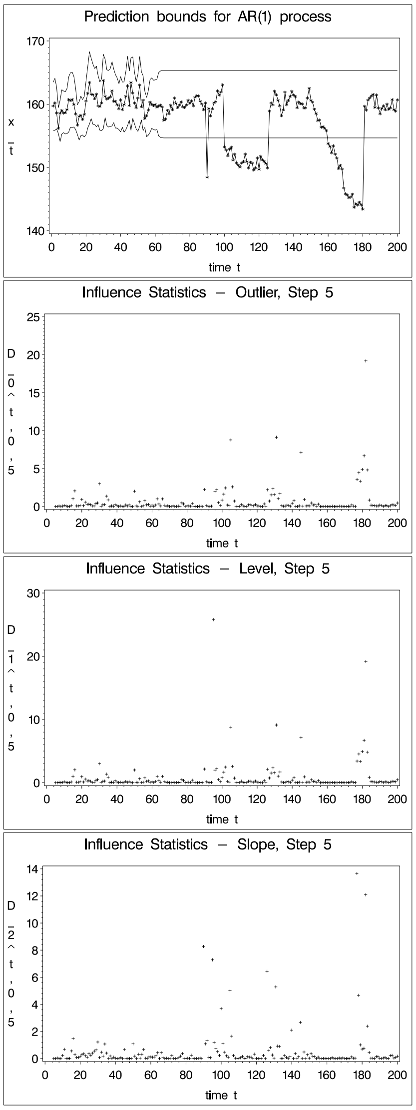
4. outlier,
   if $D_0^{t,1,0} \geq 9.0 \cdot D_{0,r}^{1,0}$ and at time t there is no level change or trend.
   If $D_0^{t,1,0} \geq 9.0 \cdot D_{0,r}^{1,0}$, but $D_1^{t,5,0} < D_{1,q}^{5,0}$ and x_{t-1} is an outlier, then we suppose that for time t the series has returned to the steady state after an outlier.

5. steady state:
   else.

Figure 3: Simulated Time Series – III



Influence statistics for outlier, level change and slope change one observation ahead. For both the outlier and the sudden end of the trend period all influence statistics are very large.



Influence statistics for outlier, level change and slope change five observations ahead. Now the influence of the level change is better visible, while the trend still is not obvious.

Basically we set $q = 2$, $r = 8$, and $q = 3$, $r = 16$ respectively for an estimation period of 30 or 60 minutes. These rules failed similarly as in the AR approach, when there was either very little variability in the estimation period or very high variability with some outliers. The former resulted in the detection of too many outliers and level shifts, as even small changes are "relatively big", while the latter rendered the detection of any change difficult. Thus, we tried several adjustments. In case of low variability, we multiplied the estimated variance V by the factor 10 (strategy L1), or multiplied all $D_{i,j}^{h,k}$ by 10 (L2). In case of very high variability, we either divided V by 10 and used $q = 4$ (H1) or we divided V by 100 and used $q = 5$ (H2). Standard factors were chosen as a routine adjustment is needed.

Under the normality assumption the statistics $D_0^{t,1,0}$ are independent (biased) estimators of the standardized squared deviation from the process level when the process is in a steady state. For $r = 8$ ($r = 16$) we have more or less the 32% percentile of the influence statistics for an estimation period of 30 (60) minutes (we started the calculation of the statistics at $t = 5$). As the expected 32% percentile of a squared $N(0, \sigma^2)$-distributed variable is an estimate of σ^2, $D_{0,r}^{1,0}$ can be treated as a robust variance estimator. Thus, our basic rule for outlier detection corresponds to the common "3σ rule".

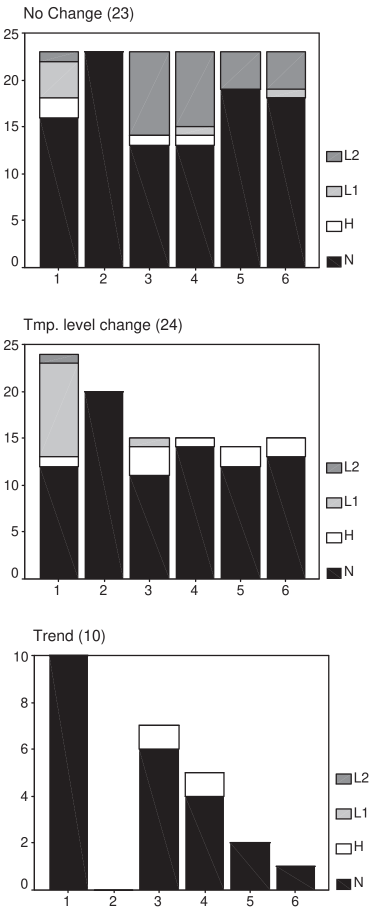
6 Results

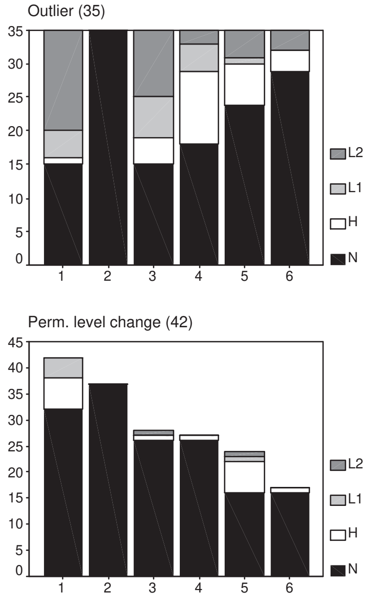
The results of our case-study are summarized in Figure 4. With autoregressive models all series with outliers, level changes and without change were correctly identified. The phase space approach always identified series without any change and with outliers, too. Identification of level changes failed, when the decrease or increase of the observed values was rather slow.

Dynamic linear models are at first sight very appealing as they allow to assess the distance of each observation from the current level as well as the changes in level and in slope over time. Nevertheless, classification with DLMs was more problematic since the influence statistics turn out to be not very reliable when the changes do not have an ideal form. Moreover, parameter estimates can be strongly affected by outliers. Series without change and with outliers could be identified more often with estimation intervals of 60 minutes. Level changes were detected best by moving an estimation interval of 30 minutes through the series. However, any of the results was worse than for the AR and the PS approach. Identification by influence statistics for the DLM parameters has severe problems with little variability during the estimation period, with level changes occuring stepwise and with patterns of outliers in small time lags. Little variability during the estimation period causes the detection of outliers and level changes to be too sensitive subsequently. Stepwise level changes are hard to detect since the smoothed level parameter adjusts step by step, possibly without any significant influence statistics. Several close outliers may either mask each other or be mistaken for a level change. Figure 5 shows an example of a stepwise level change which could not be detected.

All methods were more sensitive to outliers and level changes than clinically relevant. Especially with outlier detection, 95% prediction intervals for autoregressive models were too close. In a second run the prediction intervals were adjusted until clinically relevant results were found. This problem was most pronounced when the series had very small variability during the estimation period. For those series deviations from the mean are statistically significant on the 95% level which are clinically not meaningful, as the small prediction intervals do not reflect therapeutically important changes. In five cases of outlier detection, the PIs were adjusted to 99.99%. For a very sensitive detection of outliers in some instances the PI was reduced to 90% (Table 2). In PS-models an overall level of 99.99% was chosen for all series. For DLMs, standardized adjustments depending on the estimation period could improve classification in some cases.

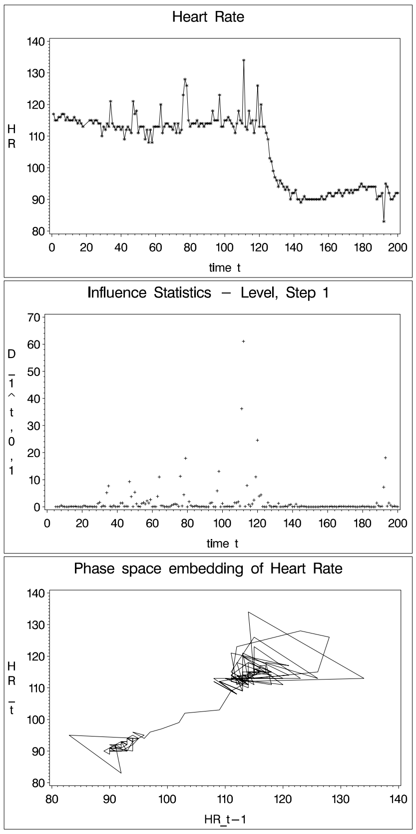
Figure 4: Numbers of correct classification results



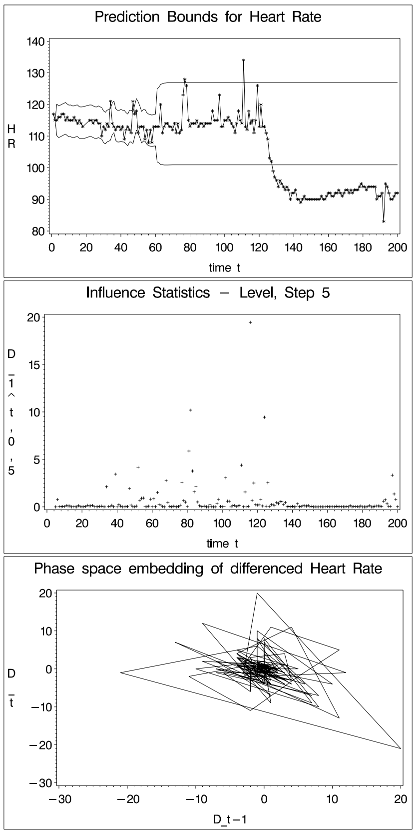


1: AR, 2: PS, 3-6: DLM,
3: 30 min, 4: 30 min MW,
5: 60 min, 6: 60 min MW;
the shading denotes
the applied correction strategy
(n means no correction
MW Moving Window)

Figure 5: Real Data Example – Heart Rate



Top: Series with a stepwise level change;
middle: Influence statistics for level change, one observation ahead;
bottom: Phase space embedding of the series;



Top: Upper and lower prediction bounds;
middle: Influence statistics for level change, five observations ahead;
bottom: Phase space embedding of first differences;

While the level change is detected by the AR and the PS approach, the influence statistics do not detect this stepwise change.

Trend detection cannot be done directly neither with AR nor with PS models. With AR models, it requires complete model diagnostics as described in section 3. After first order differentiation, the AR model was fitted to the time series in the estimation period. Because of the differentiation of the series the 95% PI widens rapidly after the start of the prediction period. This phenomenon precludes sensitive detection of changes during the prediction period. With PS models a trend can only be detected by the shape of the vector ellipsoid, which is a relatively insensitive method for the detection of slight trends, see Figure 2. The possibility of direct trend detection is the main advantage of DLMs. Trend detection with DLMs was best when the hyperparameters were fitted to an estimation interval of 30 minutes at the beginning and kept unchanged thereafter.

Comparison between precisely diagnosed AR models and over-determined models (AR order higher than necessary) showed that over-determined models allow a semi-automatic pattern detection without any trade-off in clinical sensitivity. PS models offer opportunities for fully automated time series analysis in this context. DLMs demand suitable specification of the hyperparameters, and for any classification rule formulated by influence statistics there are patterns of outliers which destroy the analysis. A more detailed discussion of AR and PS models can be found in Imhoff et al. (1998).

7 Discussion

The individual statistical evaluation of a single patient constitutes an important task in critical care monitoring. Reliable algorithms for artefact rejection have to be used before the series can be subjected to statistical analysis. Robustness against artefacts and outliers is still a major problem with most time series methods (Atkinson et al. (1997)).

In our study, patterns of univariate physiological time series could reliably be identified both with low order autoregressive models and phase space models. The only exception were trend patterns where both approaches have shortcomings. DLMs offer advantages for trend detection, but they are not as reliable as the other approaches for the detection of outliers and level changes. The phase space approach allows a meaningful application even with small sample sizes.

For most bedside decision problems the methods are too sensitive. AR models seem to be better in this regard than PS and DLMs. But a direct comparison is difficult because, in the estimation period of AR models, no pattern detection is performed. Thus, there is no possibility to misclassify patterns in this period, whereas PS models look for patterns from the onset.

A possible task for further research is to replace the assumption of normality by allowing for distributions with more weight in the tails (Adler et al. (1998)). This could reduce the sensitivity of the procedures. A disadvantage would be that estimation procedures are more complicated and demand more computational effort. This could be critical for an extension to a multivariate monitoring procedure for several physiological variables. Another way to reduce the sensitivity is to use an automatically adjusted level. A low level should be chosen, if the variability of the process is small and vice versa. Such an automatized level adjustment has already been included into the PS procedure and has lead to significant improvements. For DLMs robust Kalman filter procedures (Schick and Mitter (1994)), which are less sensitive against outliers, might improve classification. Influence statistics are based on non-reversible transformations, thus they imply a loss of information. This is much worse in online monitoring, where few information is available, than in a retrospective setting. As an alternative, the smoothed parameter estimates could be monitored directly.

Research has also to be done w.r.t. multivariate time series analysis since the underlying processes in intensive care are high dimensional with multiple dependencies between the physiological variables. This could improve classification of the physical state of a patient. Automatic classification of states in every situation and online application are not feasible at present. Methods for automatic online analysis of physiological variables would offer an option for a more reliable evaluation of the individual treatment. Such techniques could be employed to generate smart alarms, that may be more reliable and less error prone than currently used simple treshold alarms. On the other hand, time series analysis can be very helpful in the scientific off-line analysis of univariate intensive care monitoring data. They may complement the other analytical, data-based approaches towards medical diagnostics (e.g. Wernecke et al. (1988)).

References

ADLER, R., FELDMAN, R. and TAQQU, M.S. (1998): A Practical Guide to Heavy Tails. Statistical Techniques and Applications. Birkhauser Bosten, Cambridge.

ATKINSON, A.C., KOOPMAN, S.J. and SHEPHARD, N. (1997): Detecting shocks: Outliers and breaks in time series. *J. Econometrics*, *80*, 387–422.

BAUER, M., GATHER, U. and IMHOFF, M. (1998): Analysis of high dimensional data from intensive care medicine. In: R. Payne and P. Green (eds.): *Proceedings in Computational Statistics.* Springer-Verlag, Berlin, 185–190.

BAUER, M., GATHER, U. and IMHOFF, M. (1999): The Identification of Multiple Outliers in Online Monitoring Data. Technical Report, SFB 475, Department of Statistics, University of Dortmund, Germany.

BECKER, C. and GATHER, U. (1999): The Masking Breakdown Point of Multivariate Outlier Identification Rules. *J. Americ. Statist. Assoc.*, *94*, 947–955.

BOX, G.E.P., JENKINS, G.M. and REINSEL, G.C. (1994): *Time series analysis, forecasting and control.* Third edition. Prentice Hall, Englewood Cliffs.

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

COOK, R.D. (1977): Detection of Influential Observations in Linear Regression. *Technometrics*, *19*, 15–18.

DAUMER, M. and FALK, M. (1998): online change-point detection for state space models using multi-process Kalman filters. In: D. O'Leary (ed.): *Proceedings of the International Linear Algebra Society Symposium on "Fast algorithms for control, signals and image processing".* Elsevier, Amsterdam.

FAUSTMANN, P.M. and GANZ, R.E. (1994): Central cardio-autonomic disorganization in interictal states of epilepsy detected by phase space analysis. *Int. J. Neurosci.*, *78*, 43–47.

GARFINKEL, A., SPANO, M.L., DITTO, W.L. and WEISS, J.N. (1992): Controlling cardiac chaos. *Science*, *257*, 1230–1235.

GORDON, K. (1986): The multi-state Kalman Filter in medical monitoring. *Comput. Methods Programs Biomed.*, *23*, 147–154.

GORDON, K. and SMITH, A.F.M. (1990): Modeling and monitoring biomedical time series. *Journal of the American Statistical Association*, *85*, 328-337.

HARVEY, A.C. and KOOPMAN, S.I. (1992): Diagnostic Checking of Unobserved-Components Time Series Models. *J. Business & Economic Statistics*, *10*, 377–389.

HEPWORTH, J.T., HENDRICKSON, S.G. and LOPEZ, J. (1994): Time series analysis of physiological response during ICU visitation. *West J. Nurs. Res.*, *16*, 704–717.

HILL, D.W. and ENDRESEN, J. (1978): Trend recording and forecasting in intensive care therapy. *British Journal of Clinical Equipment*, January, 5–14.

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

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

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*, 24, 1305–1314.

DE JONG, P. and PENZER, J. (1998): Diagnosing Shocks in Time Series. *J. Americ. Statist. Assoc.*, *93*, 796–806.

KHADRA, L.M., MAAYAH, T.J. and DICKHAUS, H. (1997): Detecting chaos in HRV signals in human cardiac transplant recipients. *Comput. Biomed. Res.*, *30*, 188–199.

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.*, *129*, 677–684.

MILLER, G. (1956): The magical number seven, plus or minus two: Some limits to our capacity for processing information. *Psychol. Rev.*, *63*, 81–97.

MORRIS, A. and GARDNER, R. (1992): Computer applications. In: J. Hall, G. Schmidt, and L. Wood (eds.): *Principles of Critical Care*. McGraw-Hill Book Company, New York, 500–514.

NORMAND, S.-L. and TRITCHLER, D. (1992): Parameter Updating in a Bayes Network. *J. Americ. Statist. Assoc.*, *87*, 1109–1120.

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

PEÑA, D. (1990): Influential Observations in Time Series. *J. Business & Economic Statistics*, *8*, 235–241.

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

SMITH, A.F.M. and WEST, M. (1983): Monitoring Renal Transplants: An Application of the Multiprocess Kalman Filter. *Biometrics*, *39*, 867–878.

SMITH, A.F.M., WEST, M., GORDON, K., KNAPP, M.S. and TRIMBLE, I.M. (1983): Monitoring kidney transplant patients. *Statistician*, *32*, 46–54.

STRONEGGER, W.J. (1991): Kalmanfilter zur Online-Analyse von Verlaufskurven. In: U. Gilg and H. Seeber (eds.): *Multivariate Modelle*. Springer, Berlin.

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

WERNECKE, K.-D., KALB, G. and STUERZEBECHER, E. (1988): On Clasification Strategies in Medical Diagnostics (With Special Reference to Mixed Models). In: H.H. Bock (ed.): *Classification and Related Methods of Data Analysis*. North-Holland, Amsterdam.

WEST, M. and HARRISON, J. (1989): *Bayesian Forecasting and Dynamic Models*. Springer, New York.