Online Monitoring of High Dimensional Physiological Time Series - A Case-Study

Ursula Gather **Roland Fried** **Vivian Lanius**
Department of Statistics
University of Dortmund
D-44221 Dortmund, Germany
{Gather, Fried, Lanius}@statistik.uni-dortmund.de
Michael Imhoff
Community Hospital Dortmund
Surgical Department
D-44137 Dortmund, Germany
mike@imhoff.de

ABSTRACT

In modern statistical process control, intelligent alarm systems have to be constructed which extract the important information from multivariate time series and detect critical "out-of control" states of the underlying mechanism quickly and reliably. Regarding high-dimensional time series, statistical methods for dimension reduction can help to compress the data into a few relevant variables before characteristic patterns in the data are searched for. In this paper we apply graphical models as a preliminary step preceding a factor analysis of the vital signs of critically ill patients in intensive care. Then a procedure for the online-detection of change points in univariate time series is applied to the original series and to each of the factors and the results are compared to the judgment of an experienced physician.

Key Words

Time series analysis, statistical process control, dimension reduction, pattern recognition, intensive care

1 Introduction

Statistical process control aims at the extraction of important information from observed data and at the construction of alarm systems which detect critical "out-of control" states of the underlying mechanism quickly and reliably. Classical control charts consider mostly independent observations of one variable. Nowadays, increasing technical possibilities allow an online recording of many variables with high sampling frequencies. Often both autocorrelations and cross-correlations are found in the data which have to be regarded as high-dimensional time series.

In intensive care, for instance, clinical information systems (CIS) acquire and store physiological and device parameters online at least every minute. Physicians can be confronted with more than 200 variables of the critically ill patient during a typical morning round. Intelligent automatic monitoring procedures are needed to cope with this flood of information and to support decision-making at the bedside in time critical situations.

In process control of industrial manufacturing processes (weak) stationarity and the existence of a target value for the data may be assumed. It is often suggested to estimate the model parameters from past data. However, in the clinical application of monitoring vital signs neither stationarity can be assumed nor should a target value be specified because of natural changes of the data generating mechanism caused by the biorhythm for instance (Högel, 2000). Moreover, parameter estimation from past data is difficult regarding vital signs of human beings because of fundamental differences between individuals.

In intensive care, usually changes of a variable over time are more important than a single pathological value at the time of observation. Hence, the online detection of qualitative patterns such as outliers, level changes or trends in physiological variables is important for assessing the patient's state. In clinical practice, alarm systems based on fixed thresholds are used. However, their rate of false alarms is extremely high (between 70% and 99%, O'Carroll (1986), Wiklund et al. (1994)) and they have difficulties in detecting slow monotone trends. Alternatively, qualitative data abstraction has been developed using deviations of the measurements from the target range (Miksch et al. (1996)) or so-called trend templates (Haimowitz and Kohane (1996)). However, these methods do not consider autocorrelations or they demand a predefinition of expected behaviour, which is hard to specify in advance in

critical care.
Statistical time series analysis has shown to be useful for online detection of characteristic patterns in univariate time series. It 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 (Hill and Endresen (1978), Imhoff and Bauer (1996)). For pattern detection in single variables dynamic linear models (Gordon and Smith (1990)), ARIMA-models (Hepworth, Hendrickson and Lopez (1994), Imhoff et al. (1997)), and models based on a multivariate embedding (Bauer, Gather and Imhoff (1999)) have already been applied.
Pattern detection in multivariate time series is much more difficult than in univariate series since there are much more directions for deviations from the steady-state. Even an experienced physician is not able to develop a systematic response to any problem involving more than seven variables (Miller (1956)). Moreover, human beings are not able to judge the degree of relatedness between more than two variables (Jennings, Amabile and Ross (1982)). Further problems arise from the curse of dimensionality since we do not have sufficient data to estimate the model parameters reliably, particularly if the series can only be considered to be locally stationary (Dahlhaus (1997)). Moreover, in high dimensions the computational effort can exceed any available computational power (Huber (1999)). This problem becomes even more serious in online monitoring where fast and robust algorithms are needed. The necessity of robustness against disturbances such as sequences of patchy outliers is obvious since outliers can affect the correct classification of patterns. Therefore, reliable procedures for analyzing multivariate time series have to be developed and validated with real data.
In clinical practice, the physician typically selects the most important variables (according to his experience) and bases his decisions on the patterns found in these variables. Statistics offers alternative ways to reduce the dimension of the variables. Factor analysis can be applied for instance to find a few latent variables which capture most of the variability in the observed data. However, we have to ensure that the latent factors can be interpreted by the physician so that he is able to understand the alarm signals and to base his decision on them. This can be achieved by imposing a suitable structure on the loading matrices using physiological knowledge and the results obtained from data analysis with graphical models.
In Section 2 we summarize a procedure for the online-detection of patterns

in univariate time series and apply it to the measurements of ten vital signs (several blood pressures, heart rate, pulse, blood temperature) of a critically ill patient observed in intensive care. The detected patterns are compared to the judgment of an experienced senior physician. In Section 3 we apply dynamic factor analysis to the multivariate time series by analyzing the eigenstructure of the autocovariance matrices. In Section 4 we use graphical models for time series to detect associations between the vital signs and compare the results to the factors found before. In view of these results, in Section 5 we partition the variables into several subsets and analyze these subsets individually for common factors. These common factors are monitored and the patterns detected are compared to the judgment of the physician and the patterns found in the original series. We finally give a discussion of our results and an outline of future work.

2 Univariate Monitoring

Statistical time series methods so far have been mainly developed for retrospective applications. For the retrospective detection of patterns like outliers and level shifts Fox (1972) and Chang, Tiao and Chen (1988) propose likelihood ratio tests, while Peña (1990) and De Jong and Penzer (1998) apply influence statistics (see Hotta and Neves (1992) for an overview). Such methods are difficult to use online since they demand information on the future development of the time series as well. This problem is illustrated in Gather, Fried and Imhoff (2000).

For the online recognition of patterns of change in a dynamical system, we should estimate the dependence structure of the underlying process during the equilibrium or steady state from past data and find a measure to detect deviations from this steady state. Experience from earlier studies of physiologic variables (Lambert et al. (1995), Imhoff et al. (1997), Imhoff et al. (1998)) shows that physiological time series can typically be described adequately by low order AR(p)-models. 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 based on such models.

Bauer, Gather and Imhoff (1999) use an alternative approach to develop an automatic procedure for the online detection of outliers and level shifts in time series. They transfer rules for outlier identification in multivariate data to the time series context by modeling the marginal distribution of m-dimensional vectors

$$\underline{x}_t = \big(x(t), x(t-1), x(t-2), \ldots, x(t-m+1)\big)',$$
$$\underline{x}_t \in \mathbb{R}^m, \quad t = m, \ldots, N,$$

where the components are the time delayed elements of the observed time series $\{x(t)\}_{t\in\{1,\ldots,N\}}$ of length N with $m \in \mathbb{N}\backslash\{0\}, m \ll N$. In this way the dynamical information of the univariate time series is transformed into a spatial information within an m-dimensional space $\mathbb{R}^m$. 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.

For a linear stationary Gaussian process the time-delay vectors $\underline{x}_m, \ldots, \underline{x}_N$ are realizations of a sample of dependent random vectors $\underline{X}_m, \ldots, \underline{X}_N$ from a multivariate normal distribution. Hence they should form an m-dimensional elliptical cloud during the steady state and a control ellipsoid can be constructed based on the Mahalanobis distance

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

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

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

with $\hat{\gamma}_N(h) = \frac{1}{N}\sum_{t=1}^{N-h}(X(t) - \bar{X}_N)(X(t+h) - \bar{X}_N), \ h = 0, \ldots, m-1$, where $\bar{X}_N = \frac{1}{N}\sum_{t=1}^{N} X(t)$. It is obvious that $MDTS(t)$ can be replaced by its robust counterparts.

In a simulation study, Bauer, Gather and Imhoff (1999) compare this approach to forecast based detection rules for ARMA(p, q)-models as mentioned above. They find the latter approach to perform better in case of a single outlier, while the former approach seems to be preferable in case of a change affecting several subsequent observations because of a level shift, for instance. This is according to the intuition that using one-step ahead prediction means to consider only the possibility of the first outlier to occur at time t. In case of a patch of outliers or a level shift this approach fails with high probability if the first outlier is not detected and not replaced by a prediction. This deficiency will become even more serious in case of biological systems like the dynamic health state of a human being, which often shows a step-wise reaction to disturbances. In contrast to this, 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. For this reason the time-delay technique suggested in Bauer, Gather and Imhoff (1999) will be better than an approach based on ARMA(p,q) prediction for patchy outliers and level shifts at the slight expense of lower power against a single outlier, which is clinically a much less relevant phenomenon.

Similar to control charts from quality control one can either choose the level α using probability limits or some kind of control limits. Choosing probability limits means specifying a fixed value α. In practical monitoring situations it is difficult to keep the number of false alarms low when the variability of the process fluctuates and a fixed probability limit is used. Often only a deviation from the process level by more than $100k\%$, $k \in [0,1]$, is of interest, where k may be fixed by the physician, engineer, or operator. If the variability of the process is small then the probability limits are very sensitive and too many outliers are detected. On the other hand, if the variability is large the procedure is very insensitive. Hence, an alternative approach in such situations is to choose the level α depending on the process variance. Bauer, Gather and Imhoff (1999) suggest to use an adaptive significance limit by inscribing a control ellipsoid into the m-dimensional cuboid having side lengths 100k% which is centered at the current location. The current center and the current autocovariances can be estimated by moving a time window of length N through the time series assuming local stationarity (see Dahlhaus (1997) for a theoretical treatment of this concept). Concerning the choice of N, a balance between bias because of a too long window and variance because of a too short window should be searched for.

In any case the reliable distinction between patterns such as trends, level changes and outliers in physiological time series is difficult since often combinations of several patterns occur. In our case-study we use the differenced series $\{d(t)\}_{t\in\{2,\ldots,N\}}$, where $d(t) = x(t) - x(t-1)$. Differencing removes linear trends and allows simpler rules for distinguishing between outliers and level shifts. For a single outlier in the original series, we get two subsequent outlying differences, while for a level shift occurring in one step we get a single outlying difference. Hence, disturbances can be detected by the movement of the affected vectors in the m-dimensional space outside the control ellipsoid. However, the procedure described above is not suitable for the detection of slow trends. Fried, Gather and Imhoff (2000) adapt a rule for retrospective trend detection proposed by Abelson and Tukey (1963) for independent data and by Brillinger (1989) for time series. Under the assumption that there is an underlying signal $S(t)$, which is disturbed by a possibly autocorrelated noise $E(t)$, such that

$$X(t) = S(t) + E(t), \qquad t \in \mathbb{Z},$$

a weighted sum $\sum_{t=1}^{N} c(t)X(t)$ of the observations is used to test for any form of monotone increase of $S(t)$ during the time interval $t = 1, \ldots, N$, i.e., $S(1) \leq S(2) \leq \ldots \leq S(N)$ with $S(t) < S(t+1)$ for at least one $t \in \{1, \ldots, N-1\}$. The same considerations hold for a monotone decrease of $S(t)$. Since the weights $c(1), \ldots, c(N)$ must have mean $\overline{c} = 0$, the weighted sum has mean zero if $S(t)$ is constant over time. Writing $\overline{S} = N^{-1} \sum S(t)$, the weights are then determined to solve

$$\max_c \min_S \frac{|\sum (c(t) - \overline{c})(S(t) - \overline{S})|^2}{\sum (c(t) - \overline{c})^2 \sum (S(t) - \overline{S})^2},$$

i.e., to have a worst case discriminatory power for an extremely unfavorable trend which is as high as possible. This results in

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

and the corresponding worst case is a single step change. Thus, the hypothesis of a constant mean should be rejected in favor of a monotone increasing (decreasing) mean if $\sum_{t=1}^{N} c(t)X(t)$ is large (small) in comparison to its variance.

During the steady state the variance of $\sum_{t=1}^{N} c(t)X(t)$ is

$$Var\left(\sum_{t=1}^{N} c(t)X(t)\right) = \sum_{t=1}^{N} \sum_{s=1}^{N} c(t)c(s)\gamma(t-s)\,. \tag{2}$$

Hence, parameter estimation can be accomplished easily if we have reliable estimates of $\gamma(0), \ldots, \gamma(p)$ from a time window of length N. However, a trend has a serious impact on the usual sample autocovariances. For this reason Fried, Gather and Imhoff (2000) suggest to eliminate a (local) linear trend $at + b$ by usual regression methods first and to estimate the autocovariances from the residuals.

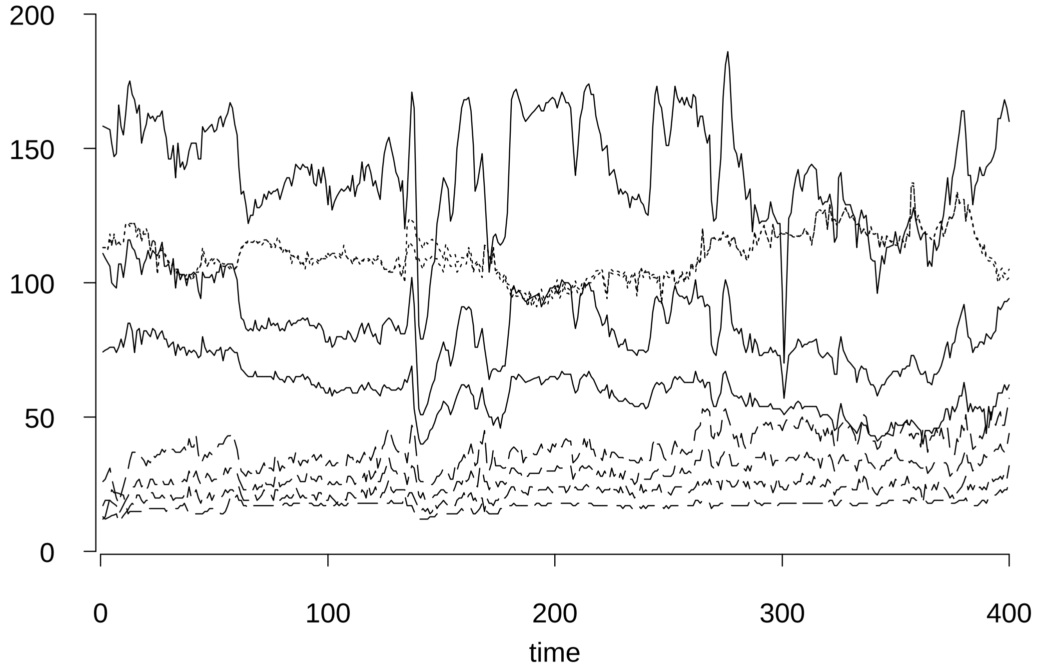

Figure 1: Vital signs of a critically ill patient, where —— : arterial pressures in mmHg (APD > APM > APS), − − − − − : pulmonary artery pressures and central venous pressure in mmHg (PAPS > PAPM > PAPD > CVP), ········· : heart rate and pulse in 1/min, first 400 observations.

In the following we apply a combined procedure for the online detection of change points in univariate time series to the following vital signs of a critically ill patient observed during extended hemodynamic monitoring: diastolic, systolic and mean arterial pressure (APD, APS, APM), diastolic, systolic and mean pulmonary artery pressure (PAPS, PAPD, PAPM), central venous pressure (CVP), heart rate (HR), pulse, blood temperature (temp). In the analysis 3804 subsequent observations measured in one-minute intervals are included. Figure 1 shows the first 400 observations of this multivariate time series.

The procedure combines the detection rule for outliers and level changes and the weighted sum chart of the original observations for the detection of monotone trends. We calculate both the Mahalanobis distance and the weighted sum from a time window consisting of $N = 30$ subsequent observations. Since experience tells us that most physiological time series can adequately be described by AR(2)-models, we choose the embedding dimension $m = 3$ for the detection of outliers and level shifts, and add a rule using the embedding dimension $m = 5$ to detect patterns such as level changes occurring in several steps during some subsequent observation times. For trend detection we choose the critical value $c = 5.0$ according to the insights gained in a simulation study (Fried, Gather, Imhoff (2000)), and consider a trend to happen if at least three subsequent values of the weighted sum exceed five times the estimated standard deviation of the test statistic since the weighted sum is not robust against outliers.

The results of this combined procedure are compared to the classifications given by an experienced senior physician who judged the individual time series plots and indicated outliers, level changes and trends which are clinically important or at least interesting. The Mahalanobis distance turns out to be much too sensitive in comparison to the judgment of the physician when a fixed significance level like $\alpha = 0.005$ is chosen. This reflects the problems with a fixed α which have already been mentioned. For this reason we use an adaptive significance level to detect changes of at least 10% of the current mean of the series. Furthermore, we call a level change critical if it is still detected when this adaptive significance level is divided by 10. A trend is called critical if it is detected even after increasing the critical value to 7.0.

Table 1 summarizes the results of our comparisons. The results for the blood temperature are not reported since there is only one pattern, an upward trend at the beginning of the series which is quickly recognized. Obviously, most of the patterns are detected correctly, and using a larger control ellipsoid and a higher critical value even allows to differentiate between relevant and interesting patterns in most of the cases. Almost all relevant patterns are detected using the smaller control bounds. However, these narrower bounds result in about 33% (45%) falsely detected outliers for arterial pressures, heart rate and pulse (for pulmonary artery pressures). In two cases the automatic procedure misspecified a level change as an outlier, and a few trends are not detected because of high variability with increasing and decreasing sequences during this period. The automatic procedure has some difficulties

		Outlier		Lev. Ch.		Trend	
AP	phys.	60	60	19	25	8	27
	det.	52	54	16	21	3	26
	fals.	11	30	2	10	2	27
HR	phys.	76	76	6	17	7	24
	det.	61	68	6	17	5	22
	fals.	23	41	3	5	7	6
PAP	phys.	106	125	10	29	34	69
	det.	104	125	8	29	27	68
	fals.	67	99	0	4	12	7

Table 1: Number of patterns labeled by the physician (phys.), number of cases detected correctly by our procedure (det.), and "falsely" identified patterns (fals.) summarized in blocks corresponding to the arterial pressures AP (APD, APM and APS), the pulmonary artery pressures PAP (PAPD, PAPM, PAPS, CV) and heart rate and pulse HR. The left column indicates the numbers of clinically relevant patterns, while the right column indicates all patterns.

in distinguishing between large outlier patches and level shifts (not reported in the table), which are both indicated by a signal from the detection rule for $m = 5$. This is an inherent problem of the online analysis of the irregular patterns which are found in the vital signs, since decisions are needed very quickly. We should point out that the classifications of the physician should not be treated as the gold standard, but as a subjective and so far only available judgment.

3 Multivariate Monitoring

In the following it is always assumed that we observe a multivariate stationary time series $\underline{X}(t) = (X_1(t), \ldots, X_k(t))'$, $t \in \mathbb{Z}$, of dimension k and with autocovariance function

$$\gamma_{ab}(h) = Cov(X_a(t+h), X_b(t)),\ 1 \leq a, b \leq k;\ h \in \mathbb{Z}$$

which is absolutely summable with respect to all time lags h for all pairs $a, b \in \{1, \ldots, k\}$. As pointed out by Tsay, Peña and Pankratz, the proper classification of outliers and level changes in multivariate time series is more difficult than in univariate time series, and multivariate detection rules are superior to simple combinations of univariate rules. The procedure for the detection of outliers and level shifts described in the last section can be extended to monitor a few, say $k = 3$, variables simultaneously. To do this we construct the $3m$-dimensional vectors

$$\underline{x}_t = (x_1(t), x_2(t), x_3(t), \ldots, x_1(t-m+1), x_2(t-m+1), x_3(t-m+1))'$$

and estimate the corresponding control ellipsoid. For instance, if we include $m = 2$ observations for each variable, we have to compute an estimate of the (6×6)–covariance matrix. Thereafter we move a time window of e.g. $N = 60$ observations through the data and compare the incoming observation at each point in time with the boundaries of the control ellipsoid. If the incoming observation is inside the ellipsoid, we include it in the time interval, reestimate the ellipsoid and move the time window to the next observation. If an observation is outside the ellipsoid, we compare the observations of each of the variables to the corresponding critical ellipsoid calculated for the individual variable at a time. All values which turn out to be univariate artifacts are replaced by the univariate prediction. After these replacements, the corrected observations are compared again to the multivariate critical ellipsoid. If they are inside the ellipsoid, we go on to the next observation. If they are outside an alarm should be given as there seems to be a change in structure.

If more than three variables are monitored simultaneously, we run into problems because of the curse of dimensionality. Possibly we do not have sufficient data to estimate the shape of the confidence ellipsoid reliably. However, the described procedure can still be applied if we first reduce the dimension of the variables.

Factor analysis aims at the reduction of the dimension of multivariate data by searching for unobservable, latent variables. Here it is assumed that there are a few, say l, latent variables called factors which drive the series and cause the correlations between the observable variables. In order to achieve good interpretability the factors can be rotated in the l-dimensional space.

Peña and Box (1987) suggest the following model for dynamic factor analysis of multivariate time series. For a k-dimensional time series $\{\underline{X}(t) : t \in \mathbb{Z}\}$ they assume

$$\underline{X}(t) = \mathbf{\Lambda}\underline{Z}(t) + \underline{\epsilon}(t) \tag{3}$$

to hold for each point in time t, where $\mathbf{\Lambda}$ is a $k \times l$-matrix of loadings, $\underline{Z}(t)$ are l-dimensional vectors of latent factors following a VARMA(p,q)-model, and $\{\underline{\epsilon}(t) : t \in \mathbb{Z}\}$ is a k-dimensional process of Gaussian white noise with zero mean and arbitrary covariance matrix $\mathbf{\Sigma}_\epsilon$, which is independent of $\{\underline{Z}(t) : t \in \mathbb{Z}\}$. To get identifiability of the parameters, $\mathbf{\Lambda}'\mathbf{\Lambda}$ can be restricted to be the identity.

If model (3) holds with independent factors, i.e, $\underline{Z}(t)$ follows a VARMA(p,q)-model where all coefficient matrices are diagonal, then the time-lagged autocovariance matrices $\mathbf{\Gamma}_X(h)$ of $\{\underline{X}(t) : t \in \mathbb{Z}\}$ are symmetrical for $h \geq 1$ and the columns of $\mathbf{\Lambda}$ will be the common eigenvectors of $\mathbf{\Gamma}_X(h)$ while the corresponding eigenvalues $\gamma_i(h)$, $i = 1, \ldots, l$, are the diagonal elements of the autocovariance matrices $\mathbf{\Gamma}_Z(h)$ of $\{\underline{Z}(t) : t \in \mathbb{Z}\}$. Peña and Box (1987) suggest to identify factor models using these findings.

An important criticism of this approach is that for time series with trend patterns the matrices $\mathbf{\Gamma}_X(h), h \geq 1$, are dominated by $\mathbf{\Gamma}_X(0)$, which causes them to be symmetric with similar eigenvalues such that actually a static principal component analysis is performed (Tiao and Tsay (1989)). Since our main interest is the detection of non-stationarities like sudden changes and trends, this is reasonable in our case. To understand this, let us assume that (3) is replaced by a deterministic analogue

$$\underline{X}(t) = \mathbf{\Lambda}\underline{z}(t) + \underline{\epsilon}(t), \tag{4}$$

where $\underline{z}(t)$ is an l-dimensional deterministic trend influencing the observable series via the matrix $\mathbf{\Lambda}$, $l \leq k$, and $\{\underline{\epsilon}(t) : t \in \mathbb{Z}\}$ is a stationary VARMA(p,q)-process of random deviations from this common trend. Assuming that the series is already centered to have the arithmetic mean zero, the sample autocovariance matrices result in

$$\begin{aligned}\hat{\mathbf{\Gamma}}_X(h) &= \frac{1}{N}\sum_{t=1}^{N-h}\Big[\mathbf{\Lambda}\underline{z}(t)\underline{z}(t+h)'\mathbf{\Lambda}' + \mathbf{\Lambda}\underline{z}(t)\underline{\epsilon}(t+h)' + \underline{\epsilon}(t)\underline{z}(t+h)'\mathbf{\Lambda}' \\ &\quad + \underline{\epsilon}(t)\underline{\epsilon}(t+h)'\Big] \end{aligned} \tag{5}$$

having expectation

$$E[\hat{\mathbf{\Gamma}}_X(h)] = \frac{1}{N}\sum_{t=1}^{N-h}\mathbf{\Lambda}\underline{z}(t)\underline{z}(t+h)'\mathbf{\Lambda}' + \frac{N-h}{N}\mathbf{\Gamma}_\epsilon(h)\,, \tag{6}$$

which is a sum of two symmetric positive-semidefinite matrices. The rank of the first summand is equal to the number of common trends. Hence, if we find $l < k$ factors when analyzing the autocovariance matrices and there are deterministic trends as we have assumed in model (4), then we can describe this deterministic structure by l common trends. Moreover, either the noise variances are negligible in comparison to the deterministic variation imposed by the trends, or the autocovariance matrices of the noise themselves show a similar factor model structure.

The physiologic time series considered in our case-study show very distinct variability. Therefore we first standardize them to have zero mean and sample variance 1. Then we try to identify a factor model by analyzing the eigenvalues and eigenvectors of the cross-covariance matrices $\mathbf{\Gamma}_Y(1), \ldots, \mathbf{\Gamma}_Y(H)$ of the standardized series $\{\underline{Y}(t) : t \in \mathbb{Z}\}$, or, equivalently, of the cross-correlation matrices of the original series for the time lags $h = 1, \ldots, H$. To get an idea about the appropriate choice of H we simulate a factor model (3) with ten observable variables and four factors of the same length as our original time series. The loading matrix and the error variances used in these simulations were chosen according to the first eigenvectors and eigenvalues obtained by an analysis of $\hat{\mathbf{\Gamma}}_X(1)$ for our data. We find the eigenvectors of the covariance matrices up to time lag 5 to be fairly stable and hence decide to set $H = 5$. However, it is also found that in our simulation setup we need as much as 1000 subsequent observations to guarantee the stability of the eigenvectors.

To check stationarity we analyze our time series data in four blocks of 951 subsequent observations each. By examining the corresponding eigenvalues we notice that four factors seem to be suitable for each period. The obtained eigenvectors are rotated by the 'varimax' procedure and the results are compared. Generally, the eigenvectors of the four periods are rather similar. For all time periods, we can identify one factor corresponding to the pulmonary artery pressures and the central venous pressure, one to the arterial pressures, one to the heart rate and the pulse, and one factor representing the blood temperature alone. These findings point at common behaviour within some subgroups of the variables. It should be noted, however, that in the fourth time period the loading vectors of the second and the third factor are interchanged in terms of size of the corresponding eigenvalues. For further investigation of the associations between the variables we apply graphical models before we proceed further.

Variable	fact. 1	fact. 2	fact. 3	fact. 4	fact. 1	fact. 2	fact. 3	fact. 4
	Time period 1				Time period 2			
PAPD	.5569	-.0679	-.0888	-.1145	.4507	.0318	-.0001	-.0908
PAPM	.4906	.0381	-.0033	.1345	.5111	-.0126	.0739	.0209
PAPS	.4485	.0609	.0538	.1402	.5752	-.0504	.0171	.3520
CVP	.4995	-.0246	.0343	-.1675	.4391	.0503	-.1266	-.2706
APD	-.0008	.5007	.0347	-.2967	-.0250	.5445	.0158	-.2339
APM	-.0245	.6144	-.0282	.0127	.0343	.5852	-.0128	.0204
APS	.0335	.6029	-.0079	.1611	.0010	.5918	.0013	.1950
HR	-.0051	.0041	.7105	.0107	-.0531	.0505	.7459	-.0527
PULS	.0115	-.0106	.6994	-.0127	.0681	-.0428	.6528	.0409
TEMP	.0140	.0227	.0034	.9024	-.0765	.0480	-.0312	.8389
	Time period 3				Time period 4			
PAPD	.5238	-.0052	.0013	-.1736	.4436	-.0901	.1052	-.0518
PAPM	.5115	.0099	-.0006	.0655	.5348	.0571	-.0245	-.0361
PAPS	.4499	.0417	-.0165	.2680	.4939	.1357	-.1065	-.0429
CVP	.5085	-.0329	.0162	-.0528	.5241	-.0948	.0271	.1487
APD	-.0489	.5311	.2075	.0785	.0532	.0103	.5261	-.1112
APM	.0329	.6051	-.0058	-.0030	-.0156	-.0018	.6006	-.0753
APS	.0017	.5923	-.1771	-.0696	-.0411	.0106	.5975	.2560
HR	.0004	.0035	.6820	-.0011	-.0136	.7031	.0116	-.0031
PULS	.0120	-.0040	.6800	-.0149	-.0028	.7020	.0109	.0086
TEMP	-.0338	-.0136	-.0129	.9394	-.0021	.0102	-.0506	.9435

Table 2: Factor loadings calculated from four subsequent blocks of 951 observations each after varimax rotation. For each block, the factor found for the largest eigenvalue corresponds to the intrathoracic pressures (PAPD, PAPM, PAPS, CVP) and the factor found for the smallest of the four eigenvalues corresponds to the temperature. The arterial pressures characterize the second factor in the first three blocks, but the third one in the fourth block. The remaining factor corresponds to the heart rate and the pulse. The factor loadings are rather similar for all time periods.

4 Graphical Models

In clinical practice physicians typically select a few vital signs to obtain a manageable number of variables. For instance, they concentrate on the arterial mean pressure and neglect the other arterial pressures since these variables are closely related.
Graphical models allow a statistical investigation of the associations within a multitude of variables (Cox and Wermuth (1996), Lauritzen (1996)). Dahlhaus (2000) extends this concept to find linear, possibly time-lagged associations in multivariate time series by analyzing the partial spectral coherence. Gather, Imhoff and Fried (2000) appraise the practical value of this new technique in a clinical study, where known associations within the hemodynamic system are reliably identified by graphical models for multivariate time series. Separate analysis of patients in different clinical states such as congestive heart failure or pulmonary hypertension even results in characterizations of the states by distinct association structures. However, for the reliable estimation of the partial spectral coherences many observations are needed and therefore this method can be used for online detection of critical states only if the sampling frequency is much faster than one observation per minute.
First we explain the concept of graphical models for time series. Under the assumptions stated at the beginning of section 3, the *cross-spectrum* between the time series $\{X_a(t) : t \in \mathbb{Z}\}$ and $\{X_b(t) : t \in \mathbb{Z}\}$ is defined as the Fourier-transform of their covariance function $\gamma_{ab}(h), h \in \mathbb{Z}$,

$$f_{ab}(\lambda) = f_{X_a X_b}(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma_{ab}(h) \exp(-i\lambda h)$$

(see Brillinger, 1981, p. 232ff). This defines a decomposition of the covariance function γ_{ab} into periodic functions of frequencies λ. The variables X_a and X_b are uncorrelated at all time lags h iff $f_{ab}(\lambda)$ equals zero for all frequencies. To distinguish between direct and induced linear relations between two series $X_a(t)$ and $X_b(t)$, the linear effects of the remaining variables on $X_a(t)$ and $X_b(t)$ have to be eliminated. Let $\underline{Y}(t) = (X_1(t), \ldots, X_k)'$, $j \neq a, b$, $t \in \mathbb{Z}$, denote the series of the other components. The *partial cross-spectrum* between $X_a(t)$ and $X_b(t)$ is defined as the cross-spectrum between the series $\{\epsilon_a(t) : t \in \mathbb{Z}\}$ and $\{\epsilon_b(t) : t \in \mathbb{Z}\}$,

$$f_{X_a X_b \cdot \underline{Y}}(\lambda) = f_{\epsilon_a \epsilon_b}(\lambda)\,,$$

where $\epsilon_a(t)$ and $\epsilon_b(t)$ are the residuals obtained by subtracting the linear influences of $\{\underline{Y}(t) : t \in \mathbb{Z}\}$ from $X_a(t)$ and $X_b(t)$ respectively. In the same way the (partial) cross-spectrum between two vector time series can be defined. Brillinger (1981, Theorem 8.3.1) shows that the partial cross-spectrum can be calculated by

$$f_{X_a X_b \cdot \underline{Y}}(\lambda) = f_{X_a X_b}(\lambda) - f_{X_a \underline{Y}}(\lambda)[f_{\underline{Y}\,\underline{Y}}(\lambda)]^{-1} f_{\underline{Y} X_b}(\lambda)\,, \tag{7}$$

where the components of the vectors $f_{X_a \underline{Y}}(\lambda)$ and $f_{\underline{Y} X_b}(\lambda)$ and the matrix $f_{\underline{Y}\,\underline{Y}}(\lambda)$ are cross- spectra between the corresponding variables.
The *partial spectral coherency* is a standardization of the partial cross-spectrum

$$R_{X_a X_b \cdot \underline{Y}}(\lambda) = \frac{f_{X_a X_b \cdot \underline{Y}}(\lambda)}{\left[f_{X_a X_a \cdot \underline{Y}}(\lambda) f_{X_b X_b \cdot \underline{Y}}(\lambda)\right]^{1/2}}\,. \tag{8}$$

In a *conditional correlation graph* for a multivariate time series we draw a vertex for each of the components $a = 1, \ldots, k$ of the time series and connect two vertices a and b by an edge whenever their partial spectral coherency $R_{X_a X_b \cdot \underline{Y}}(\lambda)$ is not identical to zero for all frequencies $\lambda \in \mathbb{R}$. Hence, a missing edge indicates that the linear relation between these two variables given all the others is zero. This is the pairwise Markov property for undirected graphical models in case of Gaussian disturbances.
For the empirical analysis of multivariate time series data, one can first estimate the cross-spectra from the data and then use versions of the equations (7) and (8) for the empirical functions to estimate the partial spectral coherencies. Thereafter we have to decide whether the partial spectral coherency may equal zero because sampling variability always causes estimates to be distinct from zero. For our calculations we used the program "Spectrum" (Dahlhaus and Eichler (2000)) which estimates the cross-spectrum by a nonparametric kernel estimator.
It is well-known that different associations between physiological variables may have distinct strengths. Therefore, we decide not to use the approximate joint $\alpha\%$ test that the partial spectral coherence $|R_{X_a X_b \cdot \underline{Y}}|^2$ equals zero at all frequencies since this allows a "yes - no" judgment only. Instead we classify the associations in high, medium, low and zero correlation on the basis of the area under the estimated partial spectral coherence. This area can be measured by the partial mutual information between the time series $\{X_a\}$ and $\{X_b\}$, which is defined by

$$-\frac{1}{2\pi} \int \log\{1 - |R_{X_a X_b \cdot \underline{Y}}(\lambda)|^2\} d\lambda$$

(Granger and Hatanaka (1964), Brillinger (1996)) or by variants of this.
The graphical correlation model which results from the analysis of the multivariate time series of vital signs already considered in the last sections is shown in Figure 2. As expected, strong partial correlations exist between the systolic, diastolic and mean arterial pressure (APS, APD and APM), between the heart rate and the pulse (HR and Puls), as well as between the systolic, diastolic and mean pulmonary artery pressure (PAPS, PAPD, PAPM). The central venous pressure is mainly related to the pulmonary artery pressures. Furthermore, many weak associations exist. The blood temperature seems to be rather isolated. Neglecting the weak associations we can identify the same groups from the graph as we have found in our factor analysis in the last section. The factor loadings calculated in a factor analysis of all variables "identify" each of the factors to belong to one of these groups, while the temperature is somewhat isolated. Hence, the results of both analyses coincide.
These results do also agree with medical knowledge. In order to obtain a manageable number of variables, physicians often select the mean pressures APM and PAPM as well as the heart rate. Hence, they select exactly that variable out of each of the subgroups identified by our graphical correlation model, which has the strongest association to the other variables. This is due to the nature of the mean pressures, which are "in between" the diastolic and systolic pressures.

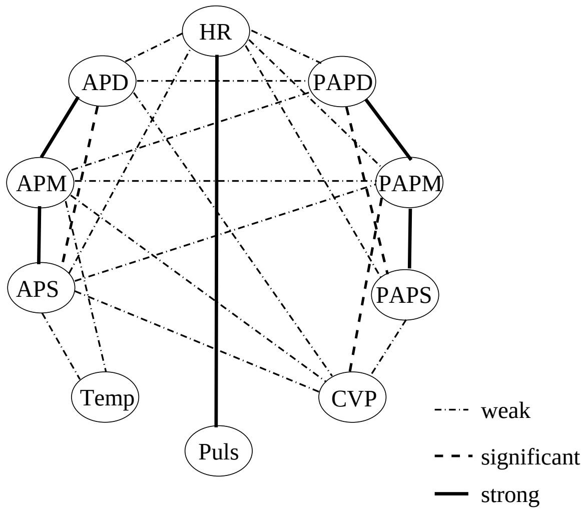

Figure 2: Partial correlation graph for the hemodynamic system.

5 Factor Analysis after Partitioning

For the construction of an intelligent alarm system based on a dynamical factor analysis of the vital signs it is important that the factors are interpretable for the physician and that they contain the important patterns found in the observable variables. Then we can simply monitor the factors and a summary measure of the model errors like the sum of the squared residuals for each time point.

The simulations in Section 3 show that in order to extract common factors from a high-dimensional time series we need a large number of observations. Peña and Box (1987) extract two common factors from a five-dimensional time series with 126 observation times. In Section 3 we find four interpretable factors for a ten-dimensional time series of vital signs using about 900 time points. Hopefully this number can be reduced if we use medical knowledge and analyze existing data by graphical models to subdivide the variables into strongly associated subsets. For instance, we can simplify the task of factor extraction in the hemodynamic system if we consider the subsets consisting of the arterial pressures, of the pulmonary artery pressures including the central venous pressure, and of the heart rate and the pulse respectively, compare Figure 2. In the following the blood temperature is neglected since it is not strongly related to the other variables, cf. sections 3 and 4.

This justifies a separate analysis of common factors for each group of variables. Since the variables of each group are measured on the same scale, we can use the sample covariance matrices up to time lag $H = 4$ and calculate the eigenvalues and the eigenvectors for each group. We find one factor to be sufficient for each group. The resulting factor loadings are provided in Table 3. Very similar loadings are obtained if we analyze the joint sample covariance matrices of all variables. Comparing the factors with each of the variables in the same group by taking differences reveals that the factors are really mixtures of all variables and cannot be identified with one of them. Figure 3 shows the factor and the observed variable which typically is selected as a representative for the corresponding group, i.e., the arterial mean pressure, the pulmonary artery mean pressure and the heart rate.

Now we apply the combined online-monitoring procedure described in Section 2 to the factors and compare the results to the judgment of an experienced senior physician. Almost all patterns which the physician considered to be important in the observed variables are visible in the factor series, too. The

Variable	factor 1	factor 2	factor 3
PAPD	0.3671294	0	0
PAPM	0.5413648	0	0
PAPS	0.6849306	0	0
CVP	0.3209671	0	0
APD	0	0.2449040	0
APM	0	0.4775610	0
APS	0	0.8437787	0
HR	0	0	0.6964061
PULS	0	0	0.7177000

Table 3: Factor loadings after partitioning the variables into closely related subgroups.

only exceptions are two slow trends detected only in the central venous pressure CVP and the systolic pulmonary artery pressure PAPS respectively, which were barely visible in the corresponding factor for the intrathoracic pressures. Except for a few (clinically not relevant) cases, outliers occurring only in a single variable are not visible any longer.
Table 4 compares the patterns detected by the physician and those detected by our combined procedure. Again, almost every pattern judged to be clinically relevant is also detected by our automatic procedure, and most of the interesting patterns are detected, too. Often we can even classify correctly whether a pattern is clinically relevant or interesting only. We get fewer false-positives when monitoring the factors than we find when monitoring the individual variables. This might be due to some smoothing effects since the noise is diminished somewhat by constructing linear combinations of similar variables.

6 Conclusion

Patterns in univariate physiological time series can be detected reliably using models from statistical time series analysis with corresponding detection rules. Outliers and level changes can be detected using rules based on a Mahalanobis distance for the marginal distribution of subsequent observations. Slow monotone trends can be identified using a weighted sum of a larger

		Outlier		Lev. Ch.		Trend	
AP	phys.	60	60	19	25	8	23
	det.	52	54	16	21	3	17
	fals.	11	30	2	10	0	4
HR	phys.	30	51	9	12	6	19
	det.	18	40	8	12	2	15
	fals.	8	3	4	2	2	2
PAP	phys.	26	39	8	15	14	25
	det.	25	37	8	14	10	21
	fals.	14	19	0	0	2	1

Table 4: Number of patterns labeled by the physician (phys.), number of cases detected correctly by our procedure (det.), and "falsely" identified patterns (fals.) for each of the factors corresponding to the arterial pressures AP, the intrathoracic pressures PAP (PAPD, PAPM, PAPS, CV) and heart rate and pulse HR. The left column indicates the numbers of clinically relevant patterns, while the right column indicates all patterns.

number of subsequent observations. Automatic procedures for online monitoring of physiological time series are too sensitive if a fixed significance level is used because of changing variability. Substantial improvements can be accomplished if an automatically adjusted level is used. Many of the patterns which have not been detected by our automatic procedure occurred during the first few hours of online monitoring. The patterns within this time period immediately after the operation are rather complicated and often several patterns occur immediately one after the other. Hence they are difficult to distinguish by any method. This does not pose a great problem for the real world application since during this time period the staff typically very closely observes the new patient. On the other hand, the outliers which were detected by the procedure and not by the physician stem from patterns which do not contain extreme values but are unusual given the immediate past of the process. Hence, they possibly contain information about irregularities in the physiological mechanism, too.

A major difficulty in online monitoring of physiological time series is the proper distinction between patchy outliers and level shifts. In our case-study, very often an alarm was given by the detection rule based on a five-dimensional embedding of the time series. Such an alarm points at an unusual pattern with a duration of more than one observation, which can either be a level shift occurring in several steps or a patch of outliers. Indeed, the physician decided retrospectively that a lot of these patterns are patchy outliers having a duration of about four or five observations. That this means four or five minutes for a sampling rate of one observation per minute, which is a long time needed anyway for distinguishing between a minor fluctuation and an important systematic change. A possibility to diminish the number of false alarms could be to use distinct classification rules for distinguishing between the several patterns based on the amount of change found for the first outlying time-delay vector: If the p-value of the Mahalanobis-distance is very low a sudden decision should be taken considering the next observation only, while for a significant, but not extreme p-value corresponding to a moderate change some subsequent observations should be considered to avoid unnecessary false alarms.

For multivariate monitoring, statistical methods for dimension reduction may be applied to compress the information into a few important variables. Graphical models explore the associations between the variables. They can be used to divide the variables into subgroups consisting of closely related variables. Then we either can select one variable from each subgroup, which is considered to be most important, or we can analyze the subgroups separately further using other statistical techniques.

Factor analysis uses the associations between the variables to extract a set of latent variables which captures most of the variability in the original data. Hence, we can try to replace the observed variables by these latent factors in the monitoring process. In our analysis, these factors have been calculated retrospectively from a large number of observations to get interpretable results. This yields a problem for the online calculation of common factors where only the observations up to the current time point are known. Therefore we use a graphical correlation model to guarantee identifiable factors calculated from subgroups of closely related variables. Then the calculation of factors should require fewer observations and hence a much shorter initial sequence for the estimation of a loading matrix. Like this we only lose the information about the associations between the groups of variables, which could even change over time depending on clinical states (Gather, Imhoff and Fried (2000)). Another possibility to overcome this problem could be to analyze existing data of former patients and to calculate an 'average' loading matrix. However, in this way we cannot achieve an individual assessment of

patients.
In our case-study, almost all patterns which are deemed to be relevant or at least interesting in the individual variables by the physician are visible in the corresponding factor, too. Hence the factors found in our analysis can truly be considered to be a lower dimensional summary of the multivariate time series. The observable variable which contained the largest number of patterns found in the other variables, i.e., which would be best for monitoring because of the highest "coverage" of important patterns (i.e. level changes and trends), are the mean pressures, which are usually also chosen for monitoring by the physician. However, PAPM covered "only" about 70 % of the patterns found in PAPS, PAPD or CVP (in comparison to more than 90% coverage of all patterns found in any variable of this group for the factor). Similarly, about 80% of the important patterns detected in APS and APD could also be found in APM, while all important patterns detected in this group are also visible in the factor.
For this case-study, we have chosen the variables representing the hemodynamic system of the patient since it gives very important information about the patient's state and because it is the basic set of variables for detecting life-critical situations. Moreover, there is very profound medical knowledge about this set of variables, about its associations and its basic reactions for distinct clinical states. This allows validation of statistical tools for online monitoring. The same partitioning of the variables into subgroups, which is gained by graphical correlation models here, could also be done by the physician without statistical analysis. Nevertheless, because of this agreement between statistical analysis and medical knowledge there is an indication that our approach can be employed to other systems where we have less background knowledge.
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 could be based on techniques of statistical time series analysis 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 (Morik et al. (2000)) the physician in charge might then be advised 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

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.*, 34, 1347-1369.

BAUER, M., GATHER, U., and IMHOFF, M. (1999). The Identification of Multiple Outliers in Online Monitoring Data. Technical Report 29/1999, SFB 475, University of Dortmund, 44221 Dortmund, Germany.

BRILLINGER, D.R. (1981). *Time Series. Data Analysis and Theory.* San Francisco: Holden Day.

BRILLINGER, D. R. (1989). The Consistent Detection of a Monotonic Trend Superposed by a Stationary Time Series. *Biometrika*, 76, 23-30.

BRILLINGER, D.R. (1996). Remarks Concerning Graphical Models for Time Series and Point Processes. *Revista de Econometria*, 16, 1-23.

CHANG, I., TIAO, G.C., and CHEN, C. (1988). Estimation of Time Series Parameters in the Presence of Outliers. *Technometrics*, 30, 193-204.

COX, D.R., and WERMUTH, N. (1996). *Multivariate Dependencies.* London: Chapman & Hall.

DAHLHAUS, R. (1997). Fitting Time Series Models to Nonstationary Processes. *The Annals of Statistics*, 25, 1-37.

DAHLHAUS, R. (2000). Graphical Interaction Models for Multivariate Time Series. *Metrika*, 51, 157-172.

DAHLHAUS, R., and Eichler, M. (2000). SPECTRUM, a fortran program to calculate and test partial spectral coherences. Available via http://www.statlab.uni-heidelberg.de/projects/.

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

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

FRIED, R., GATHER, U., and IMHOFF, M. (2000). The Online Detection of a Monotonic Trend in a Time Series. Preprint, Department of Statistics, University of Dortmund, Germany.

GATHER, U., FRIED, R., and IMHOFF, M. (2000). Online Classification of States in Intensive Care. In: *Festschrift in honor to Hans-Hermann Bock's 60th birthday, Data Analysis, Classification, and Applications*, eds. W. Gaul, O. Opitz, M. Schader, RWTH Aachen, Springer, 413-428.

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.

GORDON, K., and SMITH, A.S.M. (1990). Modeling and Monitoring Biomedical Time Series. *J. Americ. Statist. Assoc.*, 85, 328-337.

GRANGER, C.W.J. and HATANAKA, M. (1964). *Spectral Analysis of Economic Time Series.* Princeton: Princeton Press.

HAIMOWITZ, I.J., and KOHANE, I.S. (1996). Managing Temporal Worlds for Medical and Trend Diagnosis. *Art. Int. Med.*, 8, 299-321.

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. *Br. J. Clin. Equipment*, 1, 5-14.

HÖGEL, J. (2000). Applications of statistical process control techniques in medical fields. *Allg. Stat. Arch.*, 84, 337-359.

HOTTA, L.K., and NEVES, M.M.C. (1992). A Brief Review on Tests for Detection of Time Series Outliers. *Estadistica*, 44, 103-148.

HUBER, P.J. (1999). Massive Datasets Workshop: Four Years After. *J. Comp. Graph. Stat.*, 8, 635-652.

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.

JENNINGS, D., AMABILE, T., and ROSS, L. (1982). Informal Covariation Asessments: Data-Based Versus Theory-Based Judgements. *Judgment Under Uncertainty: Heuristics and Biases,* eds. Kahnemann, D., Slovic, P., Tversky, A., Cambridge: Cambridge University Press, pp. 211-230.

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.

LAURITZEN, S. L. (1996). *Graphical Models.* Oxford: Clarendon Press.

MIKSCH, S., HORN, W., POPOW, C., and PAKY, F. (1996). Utilizing Temporal Abstraction for Data Validation and Therapy Planning for Artificially Ventilated Newborne Infants. *Art. Int. Med.*, 8, 543-576.

MILLER, G. (1956). The Magical Number Seven, Plus or Minus Two: Some

Limits to Our Capacity for Processing Information. *Psychol. Rev.*, 63, 81-97.

MORIK, K., IMHOFF, M., BROCKHAUSEN, P., JOACHIMS, T., and GATHER, U. (2000): Knowledge Discovery and Knowledge Validation in Intensive Care. *Art. Int. Med.*, 19, 225-249.

O'CARROLL, T. (1986): Survey of alarms in an intensive therapy unit. *Anesthesia*, 41, 742-744.

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

PEÑA, D., and BOX, G.E.P. (1987). Identifying a Simplifying Structure in Time Series. *J. Americ. Stat. Assoc.*, 82, 836-843.

TIAO, G.C. and TSAY, R.S. (1989). Model Specification in Multivariate Time Series (with discussion). *J. Roy. Stat. Soc.*, B, 51, 157-213.

TSAY, R.S., PEÑA, D., and PANKRATZ, A.E. (2000). Outliers in Multivariate Time Series. *Biometrika*, 87, 789-804.

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.*, 6, 182-188.

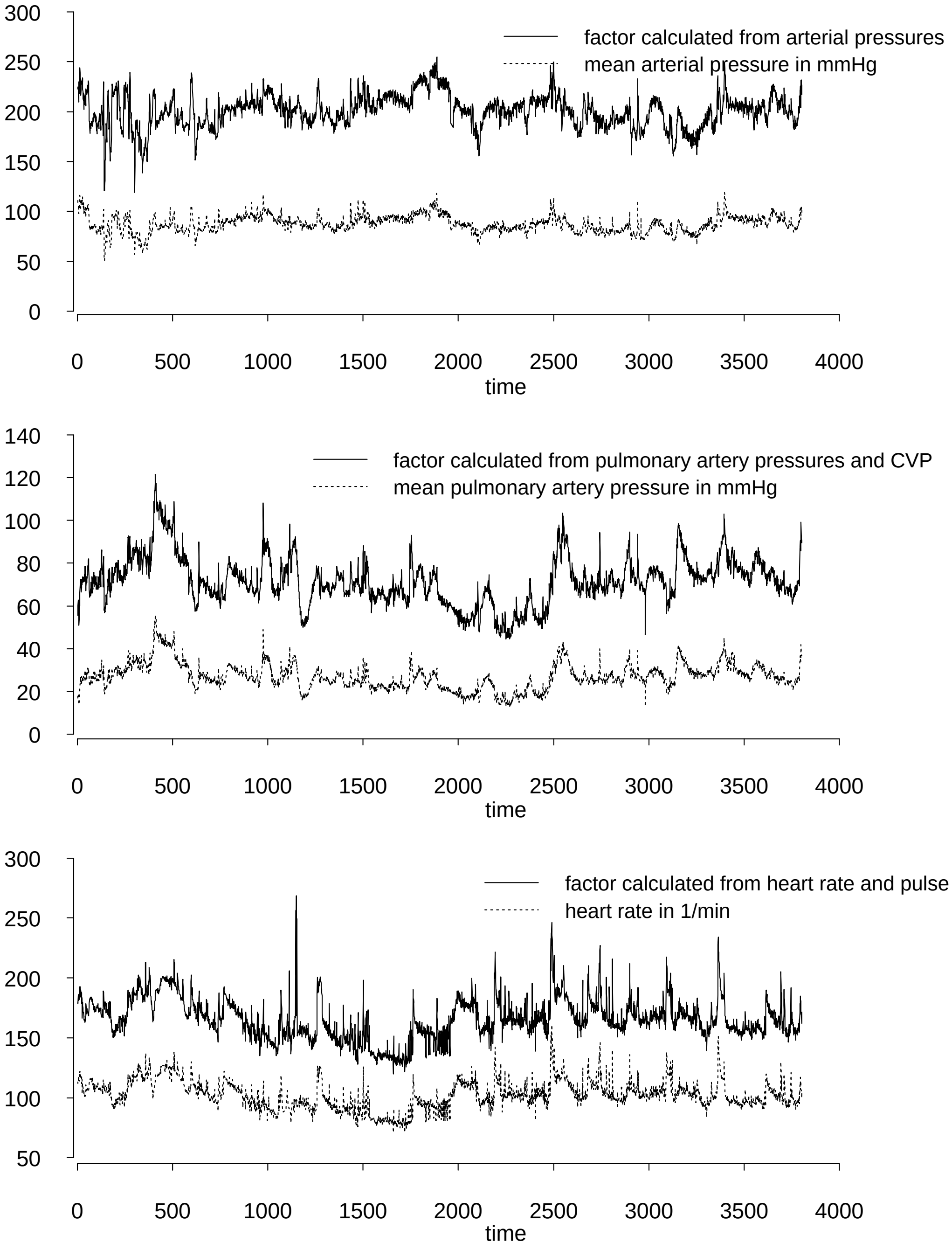

Figure 3: Extracted common factors of the analyzed subsets and one "representative" observed time series.