# Online Pattern Recognition in Intensive Care Medicine

**Dr. rer. nat. Roland Fried, Prof. Dr. rer. nat. Ursula Gather,**
**Department of Statistics, University of Dortmund, D-44221 Dortmund, Germany**
**Dr. med. Michael Imhoff,**
**Surgical Department, Community Hospital Dortmund, D-44123 Dortmund, Germany**

## Abstract

In intensive care physiological variables of the critically ill are measured and recorded in short time intervals. The existing alarm systems based on fixed thresholds produce a large number of false alarms. Usually the change of a variable over time is more informative than one pathological value at a particular time point. Intelligent alarm systems which detect important changes within a physiological time series are needed for suitable bedside decision support. There are various approaches to modeling time-dependent data and also several methodologies for pattern detection in time series. We compare several methodologies designed for online detection of measurement artifacts, level changes, and trends for a proper classification of the patient's state by means of a comparative case-study.

## Introduction

In intensive care reliable detection of critical states and of intervention effects is important for adequate bedside decision support. Clinical information systems (CIS) can acquire and store physiological variables and device parameters online at least every minute. A physician can be confronted with more than 200 variables in a critically ill patient during a typical morning round[1], while an experienced physician may not be able to develop a systematic response to any problem involving more than seven variables[2]. The existing automatic alarm systems based on fixed thresholds produce a large number of false alarms due to measurement artifacts, patient movements or minor problems such as transient fluctuations past the set alarm limit[3]. On the other hand, changes of a variable over time are often more important than one pathological value at the time of observation. Various approaches have been suggested for the detection of qualitative patterns like outliers, level changes, and trends. Data abstraction obtained by online classification of the observations into these patterns can be used as input for rule-based decision systems developed using methods of artificial intelligence[4].

An intuitive approach is qualitative data abstraction by measuring deviations of measurements from a target range. However, application of this approach to real data showed that physiological variables oscillate considerably and that pattern recognition is difficult this way[5]. Moreover, these techniques do not capture the correlations between subsequent measurements of the variables. These temporal correlations can be included in trend templates[6]. Trend templates consist of sets of low order polynomial regression models describing qualitative characteristics. Pattern abstraction is done based on the fit of these templates to the observed data. The major drawbacks of this method are the demand for predefined expected behavior and absolute value thresholds. However, time series in intensive care often show irregular behavior like patchy outliers, or outliers and level changes occurring in short time lags. Such behavior is difficult to specify in advance. Moreover, thresholds should be dynamically depending on the patient's status in the past.

Statistical time series analysis allows to model temporal correlations and to track the temporal development of the patient's status. In general, it has been shown that time series techniques are suitable for retrospective analysis of physiological variables[7,8,9]. For pattern recognition, we can compare incoming observations to confidence bounds calculated by autoregressive models[10]. Or we can apply multivariate outlier identifiers after transforming the time series into a suitable m-dimensional Phase Space[11]. Alternatively, the influence of new observations on the parameters of a dynamic linear model can be calculated[12].

Since usually multiple variables are monitored, e.g. several blood pressures, we can apply logical rules to combine and validate the patterns detected in a single variable. Rules for multivariate monitoring might even detect patterns earlier than simpler univariate rules since a joint critical value for multiple variables can be exceeded because of simultaneous changes in several variables even when all single variables still look non-critical[13]. In the present paper we concentrate on a basic discussion and comparison of several methods, which are based on statistical time series analysis, to online pattern detection in dynamic processes of physiological variables.

## Methods

**Data set.** On the surgical intensive care unit of a tertiary referral center 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.

1067—5027/01/$5.00 

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 capital letters, e.g. $X_1, \ldots, X_N$ .

**Autoregressive models.** In autoregressive models (AR) each variable is expressed as a finite, weighted linear aggregate of previous observations plus a random error[14]. An AR model for a time series formally resembles a multiple regression. A stochastic process $\{X_t : t \in \aleph\}$ is called an autoregressive process of order, denoted by AR(p), if

$$X_t = \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p} + \varepsilon_t \text{ for all } t \in \aleph$$

where $\phi_1,\ldots,\phi_p$ are unknown weights measuring the influence of preceding values on $X_t$. The random errors $\varepsilon_t$, $t \in \aleph$ , 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 applications normality is assumed for the errors[14].

Several authors have successfully applied AR models in the field of critical care[15], in longitudinal physiological experiments[16], as well as in studies on laboratory data of the chronically ill[17]. It has been shown that usually autoregressive processes of low order are suitable to describe physiological variables.

After some preliminary experiments we choose second order autoregressive models, where the last two measurements directly influence the current observation, for all cases. An extensive interactive model selection process has to be avoided since this would not be feasible in online monitoring. Each time series is split into two segments, an estimation period containing the first n observations (average length 173 minutes) and a prediction period containing the remaining observations (average length 123). An AR(2)-model is fitted to the data from the estimation period by conditional least squares. Prediction intervals are 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 is 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 is classified into the different categories. Values outside the PI are classified as an outlier, if less than 5 consecutive observations (= minutes) are outside the PI, while a level change is identified by 5 or more consecutive observations outside the PI. Trend patterns are 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 is analyzed for typical trend patterns. If corresponding signs can be found, an AR(2) model is fitted to the first differences of the series. If this model shows a sufficient goodness of fit, a significant trend is assumed.

**Phase space models.** Phase space (PS) models are based on a transformation of the time series into an m-dimensional Euclidean space by constructing phase space vectors $\vec{x}_t$:

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

with $m \in \aleph\backslash\{0\}$, $t = 1, \ldots, N-(m-1)$ and $m \ll N$. Here, $m$ is called the embedding dimension. By this technique, which is derived from the theory of nonlinear dynamic systems[18,19], the dynamic information of a series is transformed into a spatial information.

For linear Gaussian processes, Bauer et al. recommended to choose $m$ similarly to choosing the order of an AR(p) model[11]. They defined the components of the phase space vectors to be chronological observations with a time delay (lag) of always one, as dependencies between consecutive observations should be considered for pattern identification. 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 and the covariance matrix of the vectors. For estimating these parameters either the classical or robust estimators of the mean and the autocovariances of a time series can be used. 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[11].

In our case-study the first 60 observations are taken and analyzed retrospectively (i.e., outlying regions are estimated and patterns in this time interval are identi-

fied). Thereafter a time window of length 60 is moved through the data. That means, that at time point $t = 61,...,$ it is determined, if the phase space vector is in an outlying region. If not, then no pattern is detected, and the ellipse is replaced by a new one, which is estimated from the last 60 observations. If the new phase space vector is found in a distant region, it is concluded that the system is not in a steady state, and after analyzing the consecutive vectors at times $t+1$ and $t+2$ it is decided which pattern is present.

**Dynamic Linear Models.** In dynamic linear models (DLM) the current value $X_t$ of the observed process is a linear transform of an unobservable state parameter and an added random observation error[20]. The value of this state parameter is assumed to depend on its value at time $t-1$ via a linear model.

In an early application, Smith and West used a multiprocess version of the linear growth model, which can be formulated as DLM, for monitoring patients after renal transplantation[21,22]. This linear growth model reads

$$X_t = (1\ 0)\vec{\theta}_t + \varepsilon_t$$

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

Here, the states are 2-variate parameters $\vec{\theta}_t = (\mu_t, \beta_t)'$, where the first component is the process level and the second component is the slope, i.e., the change in level at time $t$. Furthermore, $\varepsilon_t \sim N(0, V_t)$ is the random observation error and $\vec{\delta}_t \sim N(\vec{0}, W_t)$ is the random change in "evolution" at time $t$, respectively. The proper a-priori specification of the variances is important since the model is sensitive w.r.t. these hyperparameters. Smith and West modeled the occurrence of outliers, level shifts and trends via different variances of the error terms for each of the states and calculated the a-posteriori probabilities of the states[21]. This procedure is not very reliable in pattern identification since it is highly parameterized[23].

Alternatively, we can use a simpler single-process model and assess the influence of groups of observations on the parameter estimates. Peña[24] and De Jong and Penzer[25] suggested Cook[26]-type influence statistics of the smoothed parameters for the retrospective detection of structural changes. These statistics are based on deletion diagnostics, where the standardized difference between the parameter estimates using all data and the estimates derived after deletion of an interesting subgroup of the data is calculated.

For online detection of structural changes we use standardized differences between the parameter estimates calculated using all data available at time $t$ and the forecasts of the parameters calculated at time $t-d$. A level change can be detected by a large standardized difference between the predicted and the estimated level parameter at time $t$. We choose $d=5$ to consider the possibility of a level change lasting several time units. Similarly, we can detect a trend by a slope change, i.e., by a large standardized difference between the predicted and the estimated slope at time $t$. For the time delay, we choose $d=20$ since reliable detection of a slow monotone trend takes some time anyway. An outlier implies a large difference between the observation and the predicted level. Here, we choose $d=5$ again because of the possibility of outlier patches disturbing the estimate of the current level. Using large sample asymptotics we can compare the standardized differences to the percentiles of the standard Gaussian distribution. We consider several strategies w.r.t. the estimation intervals. Interval lengths of 30 and 60 minutes are 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) are tried out.

## Results

With autoregressive models all series with outliers, level changes and without a change were correctly identified, i.e., the classification was identical to that of the intensivist. The phase space approach always identified series without any change and with outliers, too. Identification of level changes failed, where the decrease or increase of the observed values was rather slow. Table I summarizes the results of our case-study.

Table I: Comparison of the methods. Number of correct classifications among the total number of patterns given in brackets. Using AR models all patterns with the only exception trends could be correctly identified, but sometimes the significance level had to be adjusted. PS models showed the best automatic performance without any adjustments, but cannot detect trends. DLMs allow trend detection, too.

| | AR | PS | DLM |
|---|---|---|---|
| **No change (23)** | 23 | 23 | 23 |
| **Outlier (35)** | 35 | 35 | 35 |
| **Temp. LC (24)** | 24 | 20 | 15 |
| **Perm. LC (42)** | 42 | 37 | 28 |
| **Trend (10)** | 0 | 0 | 6 |

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 turned out to be not very reliable when the changes do not have an ideal form. Moreover, the parameter estimates are 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 occurring stepwise and with patterns of outliers at short 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.

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 at 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%. 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.

The possibility of direct trend detection is the main advantage of DLMs since trend detection cannot be done directly neither with AR nor with PS models. 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.

## Discussion

The individual statistical evaluation of a single patient constitutes an important task in critical care monitoring.

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 possibility 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 automated level adjustment has already been included into the PS procedure and has lead to significant improvements.

DLMs demand suitable specification of the hyperparameters, and for any classification rule formulated by influence statistics there are patterns of outliers which can corrupt the analysis.

For DLMs robust Kalman filter procedures[27], 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 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.

From a clinical application perspective our results show that there is not one single statistical methodology that can gracefully handle all requirements for univariate pattern detection in physiologic time series. Therefore, it may be best to use different methods in combination and fine-tune each method to a specific set of patterns.

## References


1. Morris A, Gardner R. Computer applications. In: Hall J, Schmidt G, Wood L, editorsPrinciples of Critical Care. New York: McGraw-Hill Book Company; 1992. p. 500-14.
2. Miller G. The marginal number seven, plus or minus two: Some limits to our capacity for processing information. Psychol Rev 1956;63:81-97.
3. O'Carrol TM. Survey of alarms in an intensive therapy unit. Anaesthesia 1986;41:742-4.
4. Morik K, Brockhausen P, Joachims T.Combining statistical learning with a knowledge-based approach - A case study in intensive care monitoring. Bratko I, Dzeroski S, editors. Procs. Int. Conf. Machine Learning. Morgan Kaufman Publishers; 1999. p. 268-77.
5. Miksch S, Horn W, Popow C, F. Paky F. Utilizing Temporal Abstraction for Data Validation and Therapy Planing for Artificially Ventilated Newborn Infants. Artif Intell Med 1996;8:543-76.
6. Haimowitz IJ, Kohane IS. Managing temporal worlds for medical trend diagnosis. Artif Intell Med 1996;8:299-321.
7. Gordon K, Smith AFM. Modeling and monitoring biomedical time series. JASA 1990;85:328-37.

8. Hepworth JT, Hendrickson SG, Lopez J. Time series analysis of physiological response during ICU visitation. West J Nurs Res 1994;16:704-17.
9. Hill DW, Endresen J. Trend recording and forecasting in intensive care therapy. Brit J Clin Equip 1978:5-14.
10. Imhoff M, Bauer M, Gather U, Löhlein D. Time series analysis in intensive care medicine. Applied Cardiopulmonary Pathophysiology 1997;6:263-81.
11. Bauer M, Gather U, Imhoff M. The Identification of Multiple Outliers in Online Monitoring Data. Technical Report, SFB 475, Department of Statistics, University of Dortmund, Germany; 1999.
12. Normand SL, Tritchler D. Parameter Updating in a Bayes Network. JASA 1992;87:1109-20.
13. Gather U, Fried R, Lanius V, Imhoff M. Online Monitoring of High-dimensional Physiological Time Series – A Case-study. Estadistica (in press)
14. Box GEP, Jenkins GM, Reinsel GC. Time series analysis, forecasting and control. 3rd ed. Englewood Cliffs: Prentice Hall; 1994.
15. Imhoff M, Bauer M. Time series analysis in critical care monitoring. New Horizons 1996;4:519-31.
16. Lambert CR, Raymenants E, Pepine CJ. Time-series analysis of long-term ambulatory myocardial ischemia: effects of beta-adrenergic and calcium channel blockade. Am Heart J 1995;129:677-84.
17. Lim KL, Muir K, Powell RJ. Urine neopterin: a new parameter for serial monitoring of disease activity in patients with systemic lupus erythematosus. Ann Rheum Dis 1994;53:743-8.
18. Packard NH, Crutchfield JP, Farmer JD, Shaw RS. Geometry from a time series. Physical Review Letters 1980;45:712-6.
19. Takens F. Detecting strange attractors in turbulence. In: Rand DA, Young LS, editors. Dynamical systems and turbulence. Lecture Notes in Mathematics Vol. 898. Berlin: Springer; 1980. p. 366-81.
20. West M, Harrison J. Bayesian Forecasting and Dynamic Models. New York: Springer; 1989.
21. Smith AFM, West M. Monitoring Renal Transplants: An Application of the Multiprocess Kalman Filter. Biometrics 1989;39:867-78.
22. Smith AFM, West M, Gordon K, Knapp MS, Trimble IM. Monitoring kidney transplant patients. Statistician 1983;32:46-54.
23. Gordon K. The multi-state Kalman Filter in medical monitoring. Comput Methods Programs Biomed 1986;23:147-54.
24. Pena D. Influential Observations in Time Series. J. Business Economic Statistics 1990;8:235-41.
25. De Jong P, PENZER J. Diagnosing Shocks in Time Series. JASA 1998;93:796-806.
26. Cook RD. Detection of Influential Observations in Linear Regression. Technometrics 1977;19:15-8.
27. Schick IC, Mitter SK. Robust recursive estimation in the presence of heavy-tailed observation noise. Ann Stat 1994;22:1045-80.

*Supported, in part, by the DFG (SFB475)*