Robust and Adaptive Filtering of Multivariate Online-Monitoring Time Series

M. Borowski[1], M. Imhoff[2], K. Schettlinger[1], and U. Gather[1]

[1]Fakultät Statistik, Technische Universität Dortmund, Germany
[2] Department for Medical Informatics, Biometrics and Epidemiology, Ruhr-University Bochum, Germany

Abstract— **We propose a new regression-based filter for multivariate time series that separates signals from noise and outliers in real time. The new method merges the advantageous properties of two existent filtering procedures for online-monitoring time series. Our multivariate and robust procedure yields signal estimations at the right end point of a moving time window whose width is adapted to the current data situation. Since the proposed filter works in real time, it can be used, e.g., to lower the rate of false positive threshold alarms of intensive care online-monitoring systems.**

Keywords— **Online-monitoring, multivariate time series, signal extraction, robust regression, window width adaptation.**

I. Introduction

Intensive care online-monitoring systems are used to supervise the condition of a patient. Those systems provide measurements of several physiological variables at a certain frequency, e.g., once per second. The resulting multivariate time series are noisy and non-stationary with patterns such as level shifts or trend changes. The data frequently contain outliers that are technically related or measurement artifacts. Furthermore, the physiological variables are correlated. Figure 1 shows a multivariate online-monitoring time series measured on an intensive care unit at a frequency of once per second.

Common online-monitoring stations include a threshold alarm system: an alarm sounds once a measurement exceeds the preassigned upper or lower threshold. Due to frequent outliers many false alarms occur, leading to a desensitization of the clinical staff. Under the assumption that the observations consist of relevant signals which are overlaid by noise and outliers, the number of alarms can be decreased if the alarm thresholds are not compared to the raw measurements but to the separated signals.

We propose a new procedure for extracting signals from multivariate time series in real time. Our method combines the advantageous properties of two filters that are based on robust regression in moving time windows. Therefore, we firstly introduce those two filters in the following section. Afterwards the new method is explained. In the third section the proposed procedure is applied to a multivariate time series from intensive care to demonstrate its use. The final section provides a summary.

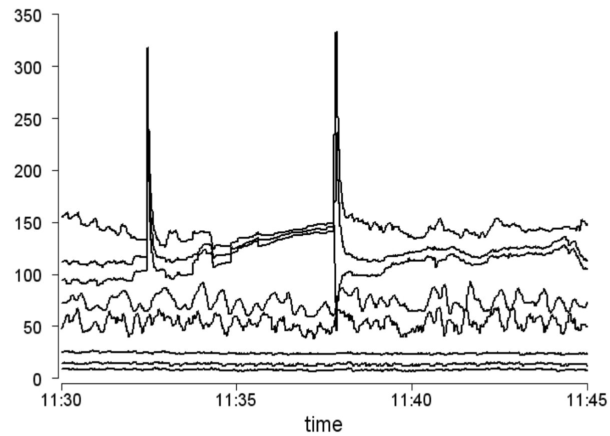

Fig. 1 Multivariate online-monitoring time series. From the top down: systolic, mean, and diastolic arterial blood pressure, pulse, heart rate, and systolic, mean, and diastolic pulmonary artery blood pressure. For presentational reasons the particular time series are shifted up- or downwards, respectively, by a fixed amount.

II. The filtering procedure

For k-variate online-monitoring time series $\mathbf{y}(t) \in \mathbf{R}^k$, $t = 1,...,T,$ we assume that the observed data can be decomposed into a true but unknown signal which is overlaid by noise and outliers:

$$\mathbf{y}(t) = \boldsymbol{\mu}(t) + \boldsymbol{\varepsilon}(t) + \boldsymbol{\eta}(t), \ t = 1,...,T. \quad (1)$$

In this model $\boldsymbol{\mu}(t) = (\mu_1(t),...,\mu_k(t))' \in \mathbf{R}^k$ denotes the k-dimensional signal at time *t*. The noise term is $\boldsymbol{\varepsilon}(t) \in \mathbf{R}^k$, where $\varepsilon_1(t),...,\varepsilon_k(t)$ are errors from a symmetric distribution with zero median and time-dependent variances $\sigma_1(t),...,\sigma_k(t)$. The errors may be correlated, i.e., $Cov\left(\varepsilon_i(t), \varepsilon_j(t)\right), \ i, j = 1,...,k, \ i \neq j,$ may be unequal to zero. An outlier generating mechanism that produces impulsive spiky noise is denoted by $\boldsymbol{\eta}(t) \in \mathbf{R}^k$.

Due to frequent outliers the k-dimensional signal $\boldsymbol{\mu}(t)$, $t = 1,...,T,$ should be estimated robustly, e.g., by applying the running median [1] to each univariate component of the multivariate time series. However, online-monitoring time series from intensive care often contain enduring trends. Hence, robust *regression filters*, that approximate *k*-variate time series by *k* regression lines within a moving time window, lead to better results [2].

We consider two kinds of regression filters: *delayed* filters approximate the signal by the level of the regression line at the midmost position *t* in a moving time window $(t-w,...,t,...,t+w)$ of odd width $n=2w+1$, $w \in \mathbf{N}$. Hence, the signal is estimated with a delay of *w* time units. Unlike delayed filters, *online* filters estimate the signal within a time window $(t-n+1,...,t)$, $n \in \mathbf{N}$, by the level of the regression line at the rightmost position *t*. By this approach the signal is estimated without time delay (except computing time), and also even window widths are eligible.

Our new procedure is based on *online Repeated Median* (oRM) regression which is an online modification of the delayed *Repeated Median* (RM) regression [3]. Both filters are highly robust since they resist up to $\lfloor n/2 \rfloor / n$ outliers within a sample of length *n*. In a simulation study, [2] compare several delayed filtering techniques w. r. t. their suitability in online-monitoring and find the RM filter to offer best compromise results. Moreover, in [4] it is shown that the oRM regression also outperforms other online filters in the special situation of online-monitoring.

A. *The adaptive online Repeated Median filter (aoRM)*

The output of a moving-window filter is strongly affected by the window size: large windows induce ‘smooth’ signal estimations with little variability whereas small windows lead to small biased signal estimations. Since the data structure is not known beforehand, an optimal choice of the window width is not possible.

One basis of our proposed method is the *adaptive online Repeated Median* filter (aoRM) [5] that approaches this bias-variance trade-off problem by adapting the size of the window sample to the current data situation. However, the aoRM is a *univariate* filter, i.e., it must be applied separately to each univariate component of a multivariate time series. That means the aoRM filter does not account for the correlation structure of the variables, in contrast to *multivariate* filters.

In [5] a univariate online-monitoring time series $y(t) \in \mathbf{R}$, $t=1,...,T$, is assumed to be locally linear. Hence, it can be approximated by a straight line within a small moving window $(t-n+1,...,t)$ of length *n*:

$$y(t-n+s) = \mu(t) + \beta(t)\cdot(s-n) + \varepsilon(t,s) + \eta(t,s), \qquad (2)$$

where $s=1,...,n$ and $n \le t \le T$. Here $\mu(t) \in \mathbf{R}$ is the level of the straight line at the recent time point *t*, and $\beta(t) \in \mathbf{R}$ is the associated slope indicating the general direction of the trend in the time window; $\varepsilon(t,s) \in \mathbf{R}$ is a noise component with zero median and time-dependent variance, and $\eta(t,s) \in \mathbf{R}$ an outlier generating process.

The aoRM filter is based on oRM regression. The oRM estimates of $\mu(t)$ and $\beta(t)$ are given by

$$\hat{\beta}^{oRM}(t) = \underset{s\in\{1,...,n\}}{\text{med}} \{z(s)\} \qquad (3)$$

with $z(s) = \underset{v\neq s, v\in\{1,...,n\}}{\text{med}} \left\{ \frac{y(t-n+s)-y(t-n+v)}{s-v} \right\}$ and

$$\hat{\mu}^{oRM}(t) = \underset{s\in\{1,...,n\}}{\text{med}} \left\{ y(t-n+s) - \hat{\beta}^{oRM}(t)\cdot(s-n) \right\}, \qquad (4)$$

where $\hat{\mu}^{oRM}(t)$ is used to estimate the signal at time *t*.

The aoRM has basically the same signal estimation output as the oRM filter with the difference that for the oRM the window width *n* is fixed while for the aoRM the window width is adapted to the current data situation at each time *t* and hence is denoted by *n*(*t*). We demand $n(t) \in \{n_{\min},...,n_{\max}\} \subset \mathbf{N}$: the minimum bound $n_{\min}$ guarantees robustness against a certain number of outliers while $n_{\max}$ limits the computing time. The inputs $n_{\min}$ and $n_{\max}$ must be set beforehand by the user. However, previous knowledge is not required.

The aoRM algorithm can be described as follows:

[start] set $t = n(t) \in \{n_{\min},...,n_{\max}\}$
[1] perform the oRM regression in the time window $(t-n(t)+1,...,t)$ and obtain $\hat{\mu}^{oRM}(t)$
[2] if $n(t) = n_{\min} \Rightarrow$ store $\hat{\mu}^{oRM}(t)$ and go to step [4]
[3] if $\hat{\mu}^{oRM}(t)$ is *appropriate* $\Rightarrow$ store $\hat{\mu}^{oRM}(t)$ and go to step [4];
if $\hat{\mu}^{oRM}(t)$ is *not appropriate* $\Rightarrow$ set $n(t)$ to $n(t)-1$ and go back to step [1]
[4] **update**: set $n(t+1) = \min\{n(t)+1, n_{\max}\}$; set $t+1$ to t and go to step [1]

The main step of this algorithm is step [3] whereat the oRM signal estimation (4) from step [1] is tested to be appropriate or not appropriate. The test is based on the fact that oRM regression results in an equal number of positive and negative residuals. The test procedure is explained in detail in [5]. If $\hat{\mu}^{oRM}(t)$ is not appropriate, the window sample is reduced by discarding the leftmost observation. Then the oRM regression is performed within the smaller time window, and it is tested again. These steps are repeated until either $n(t)$ is equal to $n_{\min}$ (compare step [2]) or the signal estimation is appropriate. Then $\hat{\mu}^{oRM}(t)$ is stored.

Afterwards the window sample is updated for the next time point $t+1$ by incorporating the next observation $y(t+1)$ (step [4]). Hence, the sample size is increased by one. If it exceeds the maximum bound $n_{\max}$, the leftmost or oldest window observation, respectively, is excluded from the window sample. Thus, $n(t)$ cannot be greater than $n_{\max}$. Finally, $t+1$ is set to *t* and the next iteration starts.

B. The multivariate online Trimmed Repeated Median-Least Squares filter (oTRM-LS)

The multivariate *Trimmed Repeated Median-Least Squares* filter (TRM-LS) [6] yields robust and efficient signal estimations and is able to detect multivariate outliers that are possibly not detected by univariate filters. However, the TRM-LS is a delayed filter that does not approach the bias-variance trade-off problem. Our new procedure includes an online modification of the TRM-LS filter that works in a moving window $(t-n+1,...,t)$ of fixed width *n* and is denominated as *online* TRM-LS (oTRM-LS) filter.

Similarly to (2) we assume each univariate component of a multivariate time series to be locally linear. Under this assumption a multivariate time series can be described by *k* straight lines in a short time window $(t-n+1,...,t)$:

$$\mathbf{y}(t-n+s) = \boldsymbol{\mu}(t) + \boldsymbol{\beta}(t)\cdot(s-n) + \boldsymbol{\varepsilon}(t,s) + \boldsymbol{\eta}(t,s), \qquad (5)$$

$s = 1,...,n$ and $n \le t \le T$. Here $\boldsymbol{\mu}(t) \in \mathbf{R}^k$ is the level vector at time *t* and $\boldsymbol{\beta}(t) \in \mathbf{R}^k$ the vector of the *k* slopes. An outlier generating process is denoted by $\boldsymbol{\eta}(t,s) \in \mathbf{R}^k$, and the components $\varepsilon_1(t,s),...,\varepsilon_k(t,s)$ of $\boldsymbol{\varepsilon}(t,s) \in \mathbf{R}^k$ are errors from a symmetric distribution with zero median and time-dependent variance, where $Cov\left(\varepsilon_i(t,s), \varepsilon_j(t,s)\right)$, $i \ne j$, $i, j = 1,...,k$, may be unequal to zero.

The oTRM-LS signal estimate at time *t*, $n \le t \le T$, is obtained by the following computing steps:

[a] perform the oRM regression in the time window $(t-n+1,...,t)$ for each of the *k* univariate components of the multivariate time series
[b] determine the oRM residuals for each component and combine them to vectors $\mathbf{r}^{oRM}(t-n+s) \in \mathbf{R}^k$, $s = 1,...,n$
[c] utilize the n residual vectors to estimate the local error covariance matrix $\boldsymbol{\Sigma}(t) \in \mathbf{R}^{k\times k}$
[d] determine the set $S_t \subset \{1,...,n\}$ of window positions corresponding to residual vectors $\mathbf{r}^{oRM}(t-n+s)$ with $\mathbf{r}^{oRM}(t-n+s)'\,\hat{\boldsymbol{\Sigma}}(t)^{-1}\mathbf{r}^{oRM}(t-n+s) \le d, \quad s \in S_t$
[e] perform the multivariate LS regression based on $\left\{\mathbf{y}(t-n+s) \,\middle|\, s \in S_t\right\}$; store the levels of the *k* LS lines at the rightmost position *t*, denoted by $\hat{\boldsymbol{\mu}}^{oTRM-LS}(t) \in \mathbf{R}^k$

At step [c] the error covariance matrix $\boldsymbol{\Sigma}(t)$ is estimated. Therefore, a robust estimator is required to avoid a masking effect. In [6] it is recommended to use the fast computable *orthogonalized Gnanadesikan-Kettenring* estimator [7]. At step [d] those time points within the window are determined whose corresponding residual vectors are 'not too large' with regard to the local correlation structure. [6,7] suggest some adequate choices for the upper bound *d*, e.g., the $\alpha-$quantile of a chi-square distribution with *k* degrees of freedom. Based on the trimmed window sample a multivariate LS regression is performed (step [e]) to obtain the signal estimate $\hat{\boldsymbol{\mu}}^{oTRM-LS}(t) \in \mathbf{R}^k$ which consists of the levels of the *k* LS regression lines at time *t*.

C. The adaptive online Trimmed Repeated Median-Least Squares filter (aoTRM-LS)

Our new method arises from a combination of the aoRM and the oTRM-LS filter and therefore is denominated as *adaptive online Trimmed Repeated Median-Least Squares* filter (aoTRM-LS). The working assumption (5) of the oTRM-LS filter is also used here, and the inputs $n_{\min}$ and $n_{\max}$ of the aoRM filter must be set beforehand by the user.

The aoTRM-LS algorithm is defined as follows:

[start] set $t = n(t) \in \{n_{\min},...,n_{\max}\}$
[A] apply the aoRM procedure in the time window $(t-n(t)+1,...,t)$ to each of the *k* univariate components of the multivariate time series to obtain *k* adapted *individual window widths* $n_i(t) \le n(t)$, $i = 1,...,k$
[B] set the *overall window width* $n_{ov}(t) = \min_{i=1,..,k}\{n_i(t)\}$;
[C] apply the oTRM-LS filter to the multivariate sample in the time window $(t-n_{ov}(t)+1,...,t)$ and store the signal estimation, denoted by $\hat{\boldsymbol{\mu}}^{aoTRM-LS}(t) \in \mathbf{R}^k$
[D] **update**: set $n(t+1) = \min\{n_{ov}(t)+1, n_{\max}\}$; set $t+1$ to *t* and go to step [A]

At step A the aoRM procedure is applied to obtain k *individual window widths* $n_1(t),...,n_k(t)$ that are used to get an *overall window width* $n_{ov}(t) \in \{n_{\min},...,n(t)\}$ (step B). Since we assume each univariate component of the multivariate time series to be locally linear (5), we must set $n_{ov}(t)$ to the minimum of the individual window widths. Then the oTRM-LS filter is applied to the multivariate sample in the time window $(t - n_{ov}(t)+1,...,t)$ to obtain the signal estimate $\hat{\boldsymbol{\mu}}^{aoTRM-LS}(t) \in \mathbf{R}^k$ (step C). Finally, the window sample of width $n_{ov}(t)$ is updated for the next time point $t+1$ (step D). This is done similarly to the update step 4 of the aoRM algorithm.

III. Application

Figure 2 shows the online-monitoring measurements from Figure 1 (gray) and the corresponding aoTRM-LS signal extraction (black) with $n_{\min}$ = 50 and $n_{\max}$ = 100. The aoTRM-LS filter has been applied retrospectively to this data sample. However, it would have yielded the same signal extraction if it had been applied online.

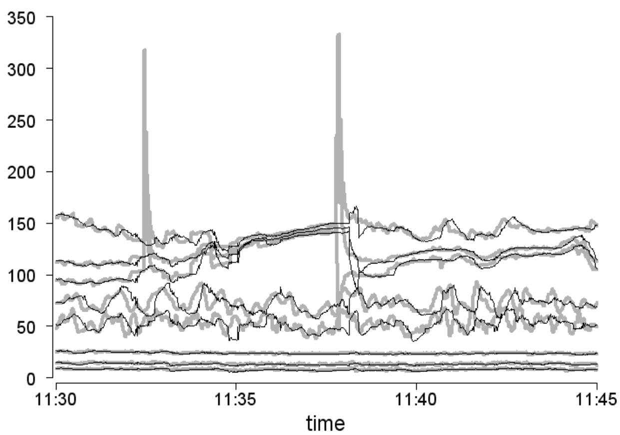

Fig. 2 Multivariate online-monitoring time series from Figure 1 (gray) and aoTRM-LS signal extraction (black).

The aoTRM-LS filter reproduces the information given by the online-monitoring system while observational noise is neglected. Moreover, the two conspicuous peaks of the arterial blood pressures around time 11:32 and 11:38 are ignored. An experienced physician annotated both peaks as being clinically irrelevant outliers. Hence, two false alarms would have been suppressed within these fifteen minutes if the filter had been applied online.

IV. Summary

Our new robust *adaptive online Trimmed Repeated Median-Least Squares* procedure serves for filtering multivariate time series in real time. It separates relevant signals from noise and outliers within a moving time window at the rightmost or current time point, respectively. The bias-variance trade-off problem for the optimal choice of the window width is approached since the size of the time window is adapted to the current data situation at each point in time. Furthermore, the filter considers the local correlations between the variables in order to estimate the signal and to detect multivariate outliers which are possibly not detected by univariate filters.

Our proposed procedure can be used, e.g., to lower the rate of false positive threshold alarms of intensive care online-monitoring systems. However, the problem of missing measurements must be solved for an online application in practice since the procedure does not work for 'incomplete' data.

Acknowledgment

We gratefully acknowledge the financial support of the Deutsche Forschungsgemeinschaft (Collaborative Research Center 475, 'Reduction of Complexity in Multivariate Data Structures').

References

1. Tukey J (1977) Exploratory data analysis. Addison Wesley, Reading, Massachusetts
2. Davies P, Fried R, Gather U (2004) Robust signal extraction for online monitoring data. J. Statistical Planning and Inference 122:65-78
3. Siegel A (1982) Robust regression using repeated medians. Biometrika 69:242-244
4. Gather U, Schettlinger K, Fried R (2006) Online signal extraction by robust linear regression. Computational Statistics 21:33-51.
5. Schettlinger K, Fried R, Gather U (2008) Real time signal processing by adaptive repeated median filters. Submitted
6. Lanius V, Gather U (2007) Robust online signal extraction from multivariate time series. Technical Report 38/07, Collaborative Research Center 475, Technische Universität Dortmund
7. Maronna A, Zamar R (2002) Robust estimates of location and dispersion for high-dimensional data sets. Technometrics 44(4):307-317

- Corresponding Author: M. Borowski
- Institute: Fakultät Statistik, Technische Universität Dortmund
- City: 44221 Dortmund
- Country: Germany
- Email: borowski@statistik.tu-dortmund.de