INTERNATIONAL JOURNAL OF ADAPTIVE CONTROL AND SIGNAL PROCESSING
Int. J. Adapt. Control Signal Process. 2010; **24**:346–362
Published online 20 April 2009 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/acs.1105

Real-time signal processing by adaptive repeated median filters

K. Schettlinger*,†, R. Fried and U. Gather

Fakultät Statistik, Technische Universität Dortmund, 44221 Dortmund, Germany

SUMMARY

In intensive care, a basic goal is to extract the signals from very noisy time series in real time. We propose a robust online filter with an adaptive window width, which yields a smooth representation of the denoised data in stable periods and which is also able to trace typical patterns such as level shifts or trend changes with small time delay. Several versions of this method are evaluated and compared with a simulation study and on real data. Copyright © 2009 John Wiley & Sons, Ltd.

Received 28 March 2008; Accepted 10 November 2008

KEY WORDS: data-driven window width selection; online monitoring; robustness; time series

1. INTRODUCTION

Monitoring systems in intensive care are important tools for diagnosing and judging the state of the critically ill. These systems assess and display measurements of physiological variables in real time. As a decision support for the medical staff, they also contain alarm systems and specific diagnosis tools. Currently, alarms for variables such as heart rate, blood pressures or oxygen saturation are still based on simple threshold rules: an alarm is triggered if an upper or lower control limit is exceeded. Such systems are prone to measurement artefacts, which lead to a high number of false alarms.

Pre-processing the input data used in the alarm system by robust online filtering can be expected to reduce the number of false alarms, as it does not only remove single artefacts but also patches of them. Since a change in the patient's health status is often accompanied by a trend change or a level shift, the applied filtering methods also need to be able to retain such clinically relevant patterns with no or only a small time delay. We propose an adaptive filter for signal processing, which is capable of adapting quickly to the data structure.

As a simple and general working model for a univariate time series $(X_t)_{t\in\mathbb{Z}}$, we assume the following:

$$X_t = \mu_t + \varepsilon_t + v_t \tag{1}$$

where μ_t denotes the underlying signal at time t, which is assumed to run smoothly over time apart from some sudden trend changes or level shifts; $(\varepsilon_t)_{t\in\mathbb{Z}}$ is an error process of i.i.d. random variables with $E(\varepsilon_t)=0$ for all $t\in\mathbb{Z}$ and (possibly time-dependent) variance $\mathrm{Var}(\varepsilon_t)=\sigma_t^2$; $(v_t)_{t\in\mathbb{Z}}$ represents an outlier-generating process.

*Correspondence to: K. Schettlinger, Institut für mathematische Statistik und industrielle Anwendungen, Fakultät Statistik, Technische Universität Dortmund, 44221 Dortmund, Germany.
†E-mail: schettlinger@statistik.tu-dortmund.de

Contract/grant sponsor: Deutsche Forschungsgemeinschaft; contract/grant number: SFB475

In intensive care, methods are required that are computationally fast and also understandable for the medical staff. 'Classical' time series techniques, dynamic linear models, Kalman filters and methods in statistical process control rely on an underlying parametric model or data-generating process, respectively. Their sensitivity to misspecification of the model parameters provides a source of error, which is unacceptable in online monitoring. This is possibly the reason why most of these approaches never got implemented in commercial products [1]. Therefore, we focus on a simple non-parametric approach.

Moving window techniques offer an intuitive and simple approach to filtering. However, standard methods like moving averages can be strongly biased in the presence of outliers while median filters, as suggested by Tukey [2], are robust against artefacts but deteriorate in trend periods. A local linear approximation instead of local constant fitting improves the filter output [3, 4], but estimation of the signal in the centre of a time window implies a time delay of half a window width. In order to avoid this delay, we approximate the signal value at the end of the window ensuring online estimation.

Following the approach in [5], we assume that the underlying signal μ_t is approximately linear in a certain local neighbourhood. In particular, if t denotes the most recent point in time and n indicates the window width, we assume that

$$\mu_{t-n+i} \approx \mu_t - (n-i)\beta_t, \quad i = 1, \ldots, n \tag{2}$$

The signal can thus be approximated by a regression fit to the n most recent observations, with μ_t estimated by the fitted value at time t and β_t denoting the slope of the regression line in this window of width n. When using a robust regression technique, a good and reliable approximation of the true signal is possible without time delay—even in trend periods.

The investigations in [4, 5] prove that compared with other robust regression methods a filter based on repeated median (RM) regression [6] shows the best performance for application to intensive-care time series, in terms of smoothness of the estimated signal, efficiency and computation time.

One disadvantage of the simple RM filter is the fact that sudden level shifts are not traced well. Certain extensions to overcome this problem have been proposed in [7–10]; specific tests for shift detection in time series, which can be incorporated into filters can, for example, be found in [11, 12]. These filters are designed for delayed signal extraction and have not been investigated for full online analysis yet. A weighted version of the RM filter [13] also improves on shift preservation and permits full online analysis. However, all these methods require the specification of a fixed window width n.

Figure 1 shows a simulated time series and two signals extracted by the simple online RM filter using different fixed window widths. The use of a larger window width results in a smoother signal estimation, whereas the use of a smaller window width leads to a smaller time delay in tracing the level shift at time $t = 76$, also yielding a smaller bias at the expense of higher variability.

The choice of the 'optimal' window width or bandwidth, respectively, and the corresponding bias-variance trade-off is a problem that also arises in kernel density estimation and non-parametric regression. It is well-known that methods based on local bandwidth selection can adapt better to the structure of the underlying function than methods using a fixed bandwidth [14, 15]. Various *data-adaptive* choices of the 'optimal' window width have been described in [16–21], but none of these methods is robust against outliers. The approach described in [22, 23] can also be used for robust filters (see, [24, 25]) but it is computer-intensive and hence not reasonable for online signal extraction in intensive care.

Since the online RM filter outperforms other filters for intensive-care applications [5], we focus on a data-driven window width selection especially for this real-time filter, which can cope with trended data and is robust against outlying values. An exact straightforward RM algorithm would need $O(n^2)$ computation time [6], but by applying RM to moving time windows an update of the estimates can be achieved in linear time [26]. This makes the RM filter very attractive for online analysis in intensive care.

Section 2 focusses on simple RM regression within one time window and states some important properties. In Section 3, a version of the RM filter is proposed, which is able to chose the window width

DOI: 10.1002/acs

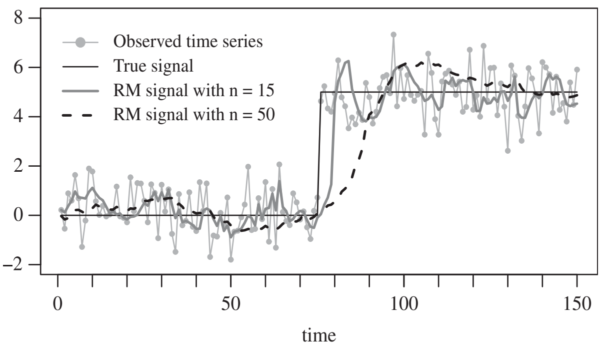

Figure 1. Online signal extraction by the simple RM filter with window widths $n=15$ and $n=50$.

n adaptively, depending on the underlying data structure. Several versions of this procedure are compared with a simulation study in Section 4. Section 5 introduces some modifications of the filter designed to meet practical demands. Section 6 presents some examples illustrating the performance of the modified adaptive filters for simulated and real time series and Section 7 gives concluding remarks and a brief outlook.

2. REPEATED MEDIAN REGRESSION

The simple RM filter for online signal extraction applies RM regression [6] within a sliding window of length n and estimates the level μ_t of the signal at the current time t by the RM fit at the last point in time within each window.

Let $\mathbf{x}_{t,n}=(x_{t-n+1},\ldots,x_t)'$ denote the data in the current window. The RM slope and level estimates at time t are then defined as

$$\hat{\beta}_t^{\mathrm{RM}} = \underset{i=1,\ldots,n}{\mathrm{med}} \left\{ \underset{j\neq i}{\mathrm{med}} \left\{ \frac{x_{t-n+i}-x_{t-n+j}}{i-j} \right\} \right\} \tag{3}$$

and

$$\hat{\mu}_t^{\mathrm{RM}} = \underset{i=1,\ldots,n}{\mathrm{med}} \{x_{t-n+i}+(n-i)\cdot\hat{\beta}_t^{\mathrm{RM}}\} \tag{4}$$

where the median $\mathrm{med}\{\cdot\}$ at an even sample size n is defined as the arithmetic mean of the $(n/2)$th and $(n/2+1)$st order statistic.

RM can reach the optimal breakdown point of $\lfloor n/2\rfloor/n$, meaning that the regression estimation still yields 'sensible' results when almost 50% of the data are contaminated by arbitrarily deviating values [27]. Since data from intensive care are very likely to be strongly contaminated by measurement artefacts, this *robustness* property is indeed required in this context.

Compared with least squares, the RM level estimate has a finite sample efficiency of about 70% at standard normal data and even higher efficiencies at heavy tailed or skewed data [5]. This means that it possesses a lower variance than many other robust estimates, causing the RM filter to result in smoother signal approximations than other robust regression filters.

Furthermore, RM possesses the 'exact fit property'

When at least $n-\lfloor n/2\rfloor+1$ of the n observations are collinear the RM regression line runs exactly through these observations (5)

nao entendi

Such a situation is likely to occur for the data we have in mind because of the discrete measurement scale for physiological time series extracted from an intensive-care online monitoring system.

The RM also has the property of the 'balance of residual signs', which is important for the procedure described in the remainder of the paper: in [28] it is pointed out that estimating the RM location hierarchically by the trend-corrected observations like in (4) causes the median of the residuals to be zero if both, the independent and the dependent variable, follow a continuous distribution. This means

$$\sum_{i=1}^{n} \mathrm{sign}(r_{t,i})=0 \tag{6}$$

where the RM residuals within one window are denoted by

$$r_{t,i}=x_{t-n+i}-(\mu_t^{\mathrm{RM}}-(n-i)\beta_t^{\mathrm{RM}}),\quad i=1,\ldots,n \tag{7}$$

and the sign function is defined as

$$\mathrm{sign}(r)=\begin{cases}-1 & \text{if } r<0\\ 0 & \text{if } r=0\\ 1 & \text{if } r>0\end{cases}$$

Since the regression within each window takes place against equidistant points in time, the assumption of

DOI: 10.1002/acs

a continuous carrier distribution is not fulfilled here. However, if we assume that the time series data come from a continuous distribution, the probability that condition (6) is not fulfilled is negligible.

3. ADAPTIVE RM FILTERS

In this section, we present an approach for a data-driven window width selection for the online RM filter. In [8] a procedure is proposed for local window width choice for a delayed RM filter, which uses (6) for assessing the model fit in each window: after fitting an RM regression to the data in the window, the total number of positive residuals T^+ and negative residuals T^- within the first and the last quarter of the window is evaluated. The window width is reduced and the RM fit is repeated if either T^+ or T^- is too small or too large compared with the expected number of residuals with the same sign in the described subset under a symmetric error distribution. However, this approach is based on a symmetric window estimating the signal in the centre of the current time window and, if necessary, reducing the window by discarding both, the first and the last value in the window. This is not reasonable for online estimation because it may cause elimination of the most recent observation.

3.1. Outline of the adaptive online procedure

The adaptive online filters presented here are based on the simple online RM filter estimating the signal in a sliding time window by (4) with the difference that the window width is not fixed but can vary over time. Therefore, we denote the window width for estimation at time t by n_t.

The fundamental idea of the adaptive window width selection is as follows: if the current fit *is not* considered adequate, then the window width used to determine the fit is reduced, and the RM fit is re-estimated at the smaller sample. If the current fit *is* considered adequate, then the window width can be enlarged for the estimation at the next point in time. Figure 2 shows an outline of the adaptive online RM filter.

For application of this filter, the minimum and maximum window width has to be chosen by the user. The minimum window width $n_{\min}$ controls how many outliers the filter can handle (within each window) and hence controls robustness. Furthermore, the smallest window should contain enough observations for a 'sensible' estimation. In intensive care, up to five subsequent outliers can occur without indicating a relevant change [29]. To ensure robustness against such situations we choose $n_{\min}=11$ in the following.

The maximum window width $n_{\max}$ should be chosen such that it can be assumed that our working assumption (2) of local linearity of the signal is still appropriate. The computation time is also limited by the specification of $n_{\max}$ because of the increasing computational demand with increasing window width. Moreover, considering that the proposed filter is intended for real-time application, the current time window should not contain information that seems 'too old' for the current estimation. With the background that a patient's health status can change within relatively short time, we choose $n_{\max}=121$ in the following, such that the largest window in applications to intensive-care time series evaluated per second includes the current time and the preceding two minutes.

Here, the first signal estimation takes place when at least $n_{\min}$ observations are present, i.e. the first time and the initial width for the first window is set to $t=n_t=n_{\min}$. However, it is also possible to start with the online signal estimation at a later time using a (possibly) larger window width $t=n_{\text{start}}\in\{n_{\min},\ldots,n_{\max}\}$.

For each time, the estimation of the current signal level is repeated, either until it cannot be rejected that the current estimate is appropriate, or until the window width cannot be reduced any more, i.e. $n_t=n_{\min}$. If the estimation is considered suitable in the first place, no further iterations need to be performed resulting in a considerable gain of computation time compared with methods requiring more than one pass over the data (see, e.g. [24]).

The core of this procedure consists of the goodness-of-fit test for checking the adequacy of the current signal estimate which raises two main questions:

1. Which test should be used?
2. Which subset of residuals should be regarded for the test?

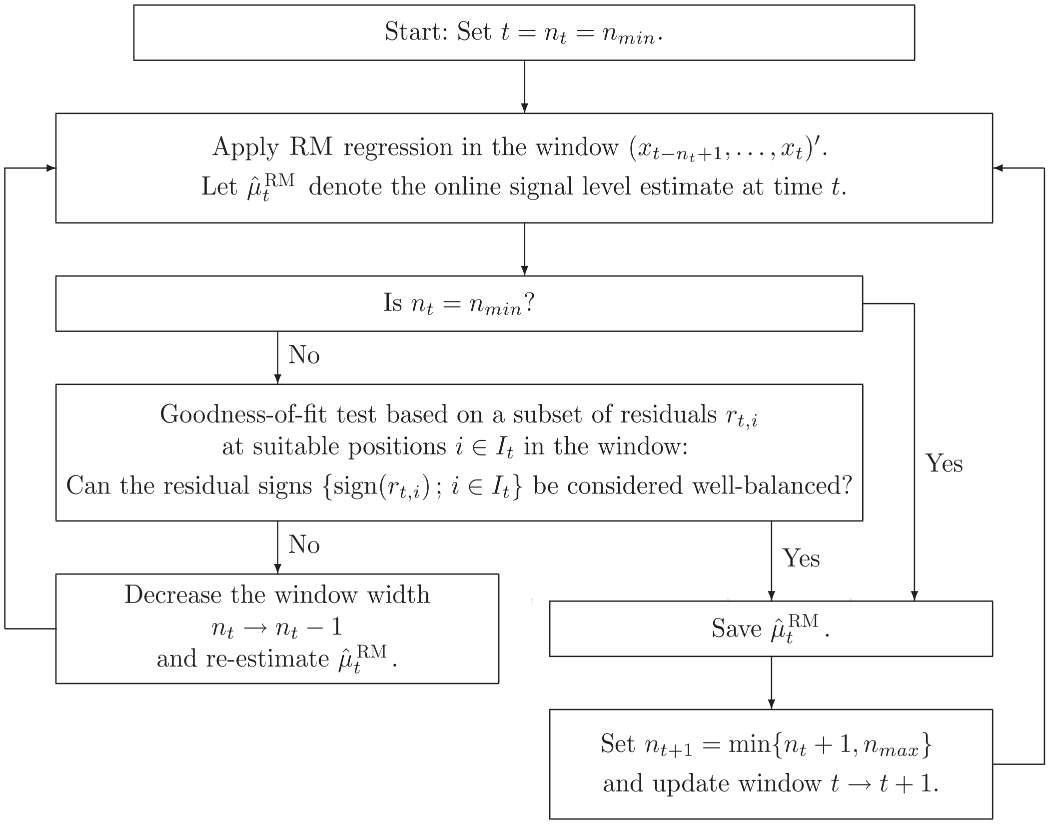

Figure 2. Flow chart of the adaptive online repeated median filter.

The first question will be addressed in the next section; the second question will be discussed thereafter, motivating the simulation study described in Section 4.

3.2. *Goodness-of-fit test*

Let $r_{t,i}$ for $i = 1, \ldots, n_t$ denote the residuals (7) in the window used for estimating the signal level at time t. Given the balance of the residual signs in the whole window (6), the proposed goodness-of-fit test is based on the idea that the balance of the signs of the residuals at certain points in time $i \in I_t \subset \{1, \ldots, n_t\}$ represents the adequacy of the fit at the last time in the window: if either negative or positive residual signs prevail, the fit at the end of the window is not considered adequate and the window width has to be adjusted.

Testing the sign balance of the residuals $\{r_{t,i};\ i \in I_t\}$ can be done using a sign test for the location of the error distribution in the selection. Here, we use $\tilde{\mu}_e^{I_t}$, the median of the distribution of the errors at points in time included in the selection I_t, as a measure of location, and we consider the test problem

$$H_0 : \tilde{\mu}_e^{I_t} = 0 \quad \text{versus} \quad H_1 : \tilde{\mu}_e^{I_t} \neq 0$$

As a test statistic, the sum of the signs of non-negative or non-positive residuals is not appropriate here, because it is non-specific about the handling of residuals with value zero. Counting the zero as non-negative or as non-positive can make a relevant difference for short windows. By construction, RM residuals can equal the value zero with positive probability. The most extreme case is given by an exact fit situation

DOI: 10.1002/acs

(5) where the number of zero residuals outweighs the number of residuals with positive or negative sign. In intensive care this situation is not a rare event since the discreteness of the measurement scale can lead to collinear data. Therefore, we choose a test statistic that takes into account that zero residuals support the null hypothesis. We define the test statistic as

$$T = \sum_{i \in I_t} \operatorname{sign}(r_{t,i}) \tag{8}$$

If $|T|$ is large, either the positive or the negative residuals in the chosen subset predominate. Hence, we reject the null hypothesis that the current signal estimation is appropriate if

$$|T| > c(n_t, I_t)$$

where $c(n_t, I_t)$ denotes a critical value depending on the window width n_t and the subset selection I_t. The critical value should be a quantile of the distribution of $|T|$ under the null. Although we can state that T can be expressed in terms of random variables with a roughly hypergeometric distribution, its true distribution is unknown because of the unknown distribution of the location of zero residuals in RM regression and the unknown dependence structure of the RM residuals and their signs.

To approximate the distribution of T for several subset selections I_t, each containing n_{I_t} values, we simulate 100 000 windows containing data from a standard normal distribution for window widths $n_t \in \{11, 12, \ldots, 121\}$. Using a distribution with heavier tails (like Cauchy) or a skewed distribution (like a lognormal) results in no difference in the distribution of T as the test statistic only considers the signs of RM residuals. The critical values we use in the following consist of the empirical α-quantiles $q_\alpha(n_t, I_t)$ of the simulated distributions, slightly modified such that the following symmetry and monotonicity conditions are fulfilled for all n_t and I_t:

$$-q_\alpha(n_t, I_t) = q_{1-\alpha}(n_t, I_t)$$

$$q_\alpha(n_t, I_t) \leqslant q_\alpha(n_s, I_t) \quad \text{for } n_t \leqslant n_s$$

$$q_\alpha(n_t, I_t) \leqslant q_\alpha(n_t, I_s) \quad \text{for } n_{I_t} \leqslant n_{I_s}$$

The simulations show that choosing the critical value as the modified 0.95-quantile of the distribution of T, depending on n_t and I_t, yields good results for online signal extraction, i.e. we get an approximate level-0.1 test at each time t. However, our procedure is meant as an exploratory tool for only judging the appropriateness of the fit based on the current window width. Supporting the null hypothesis means that larger window widths are favoured, but since this results in larger stability, smaller variance, larger robustness and smoothness of the signal estimate, it is not a disadvantage for the performance of the adaptive filter.

3.3. *Subset selection for the test statistic*

The current window width needs to be adjusted particularly if the data in the window exhibit a pattern like a level shift or trend change. Therefore, the residual signs selected to calculate the test statistic should reflect such data structures.

The left panels in Figure 3 show a window of width $n = 30$ containing a positive level shift at time $i = 21$; the right panels show a window with a change of the (linear) trend at time $i = 16$. In both cases, the online RM signal estimate deviates strongly from the true signal at the most recent time. Therefore, the test statistic or the absolute value of the sum of residual signs, respectively, should be large such that the null hypothesis (speaking for a 'good fit') can be rejected and the window width can be reduced.

Figure 3 exemplifies that the choice of the subset used for calculating the test statistic can have a large impact on the ability of the filter to adapt to certain data structures. For the level shift, choosing the most recent residual signs seems the best choice because in this subset there are mainly positive residuals resulting in a large test statistic, pointing at a bad fit at the most recent time. In the other subsets the prevalence of positive residuals might not be strong enough to indicate that the window width needs to be reduced. In the case of a trend change, the residual signs in a subset containing the most recent values are almost balanced and thus would not lead to a window width reduction. In contrast, a subset containing residuals in the centre or from both ends of the window would yield a large absolute value of T and hence would yield a detection of the misfit.

Seeing that different selections I_t are advantageous in different data situations, we want to determine the

DOI: 10.1002/acs

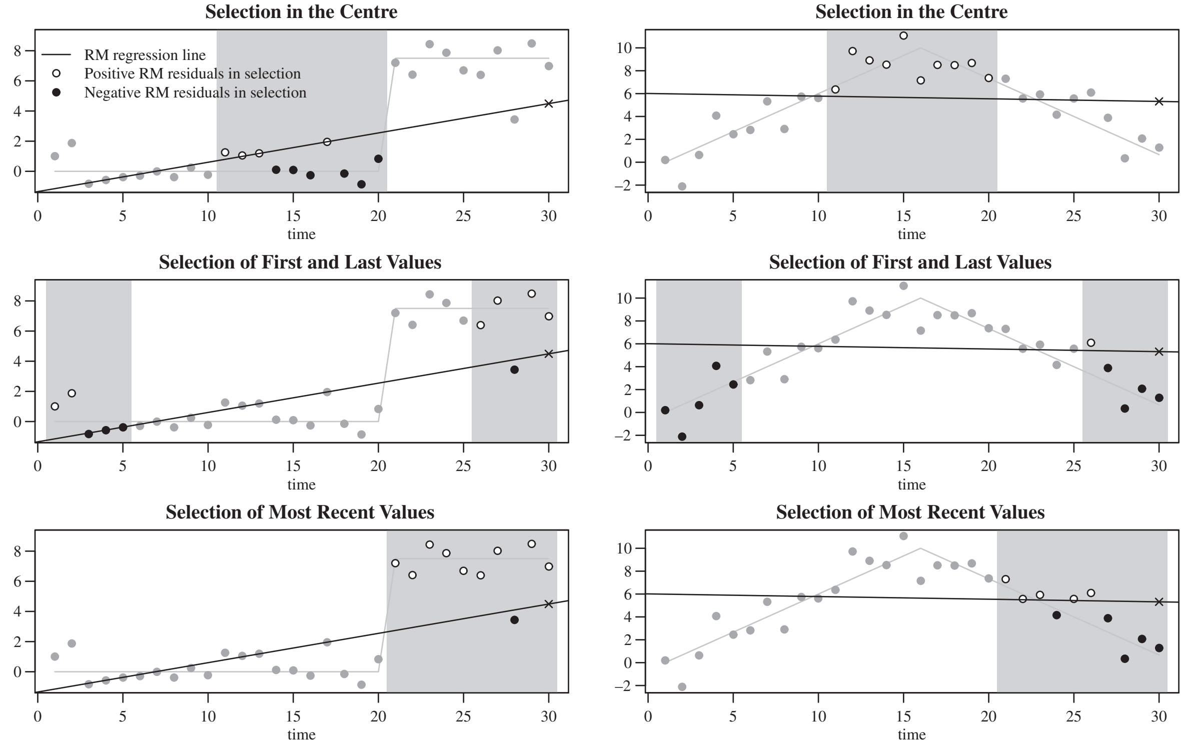

Figure 3. Repeated median regression and corresponding residuals in different subsets of size 10 in a window of width $n=30$ containing a level shift (left) and a trend change (right). The RM online estimate ($\times$) deviates strongly from the true signal (grey line).

selection that results in an adaptive online RM filter yielding the least biased signal estimates after a sudden change by a simulation study in the next section. Based on preliminary experiments, we consider those subsets that we think provide the most useful information about the fit at the most recent time.

4. SIMULATION STUDY

To identify a filter that traces an abrupt change in the data with small time delay, we investigate the performance of the adaptive RM procedure (Figure 2) using different subsets I_t with varying cardinality. Here, we consider subsets I_t such that the test statistic (8) consists either of the sum of the n_{I_t} most recent residual signs; of residual signs from the beginning and the end of the time window; or of residual signs from the centre as e.g. illustrated in Figure 3.

The subsets of times $I_t \subset \{1, \ldots, n_t\}$ used for evaluating the test statistic (8) in each window are defined by:

$$I_t^{\text{centre}} = \left\{ \left\lfloor \frac{n_t - n_{I_t} + 1}{2} \right\rfloor + 1, \ldots, \left\lfloor \frac{n_t - n_{I_t} + 1}{2} \right\rfloor + n_{I_t} \right\}$$

$$I_t^{\text{firstlast}} = \{1, \ldots, \lfloor n_{I_t}/2 \rfloor,\ n_t - \lceil n_{I_t}/2 \rceil + 1, \ldots, n_t\}$$

$$I_t^{\text{recent}} = \{n_t - n_{I_t} + 1, \ldots, n_t\}$$

each with $n_{I_t} \in \{\lfloor n_t/2 \rfloor,\ \lfloor n_t/3 \rfloor,\ \lfloor n_t/4 \rfloor\}$, i.e. with the cardinal number depending on the window width n_t. If n_{I_t} is smaller than five, we set $n_{I_t} = 5$ to guarantee a minimal number of observations in the subset I_t.

DOI: 10.1002/acs

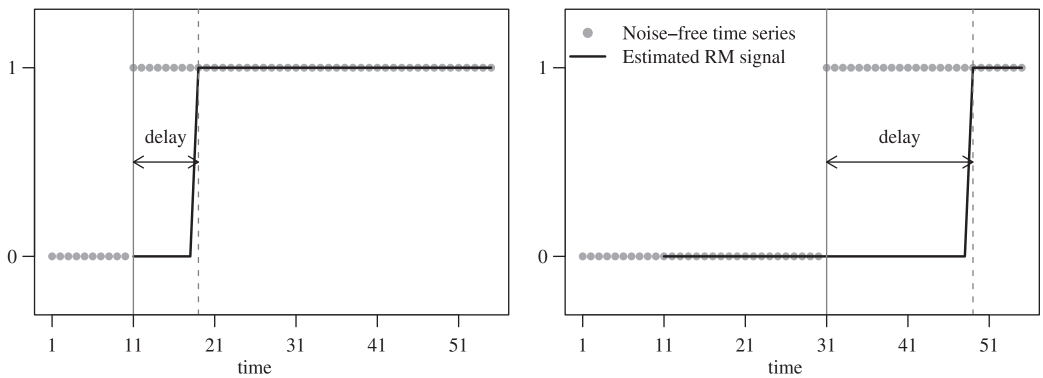

Figure 4. Noise-free time series with a level shift at $t_0 = 11$ (left panel) and at $t_0 = 31$ (right panel), and online signal estimation by an adaptive RM filter using I_t^{centre} and $n_{I_t} = \lfloor n_t/2 \rfloor$. The extracted RM signal traces the shift with shorter time delay if the constant period before the shift is shorter.

Furthermore, we consider a subset containing a fixed number n_{I_t} of most recent values, independent of the current window width n_t

$$I_t^{\text{fixed}} = \{n_t - n_{I_t} + 1, \ldots, n_t\} \quad \text{with } n_{I_t} \in \{10, 15, 30\}$$

For small window widths n_t, this subset may include a dominating part of the window, and it is even possible that $n_{I_t} > n_t$. Thus, if n_{I_t} exceeds $n_t/2$, we set $n_{I_t} = \lfloor n_t/2 \rfloor$.

We also considered subsets I_t^{thirds} that included $n_{I_t}/3$ values in the first part, $n_{I_t}/3$ in the centre and $n_{I_t}/3$ in the last part of the window. However, the filtering performance was worse than for the subsets described above in terms of time delay and bias and hence the outcomes for such a filter are not shown here.

As simulation settings we look at level shifts and trend changes as shown in Figure 3. The length of the time period before the relevant change can have an impact on the signal approximation because the window width used by the adaptive filter at the time of the change can differ. Figure 4 illustrates that this transfers to the time delay for tracing such a pattern, here, a shift. Therefore, we consider different points in time for a shift and a trend change.

Furthermore, the signal-to-noise ratio may have an effect on the accuracy of the estimations around a change point. Therefore, different jump sizes for the level shift and different slopes of linear trends for the trend change are examined. In particular, we investigate time series with the true signal

$$\mu_t = \begin{cases} \mu_1(t) & \text{for } t = 1, \ldots, t_0 - 1 \\ \mu_2(t) & \text{for } t = t_0, \ldots, 2 \cdot t_0 \end{cases} \tag{9}$$

where t_0 denotes the time of the shift or the change, respectively.

For a level shift we define

$$\mu_1(t) = 0 \quad \text{and} \quad \mu_2(t) = h \tag{10}$$

and for a trend change we let

$$\mu_1(t) = \beta \cdot t \quad \text{and} \quad \mu_2(t) = 2\beta(t_0 - 1) - \beta \cdot t \tag{11}$$

For the simulations, different settings for h, β and t_0 are considered.

4.1. Noise-free level shifts and trend changes

As extreme settings, we consider the signal μ_t (9) as time series without any additional noise or outliers. Because of the exact fit property and the regression equivariance of RM, the size of the shift h in (10) or the size of the slope β in (11), respectively, has no impact on the delay for tracing a change in the signal in noise free situations. Therefore, we examine

- a level shift of size $h = 1$ at $t_0 \in \{11, 21, \ldots, 131\}$ and
- a trend change with $\beta = 1$ at $t_0 \in \{11, 21, \ldots, 131\}$.

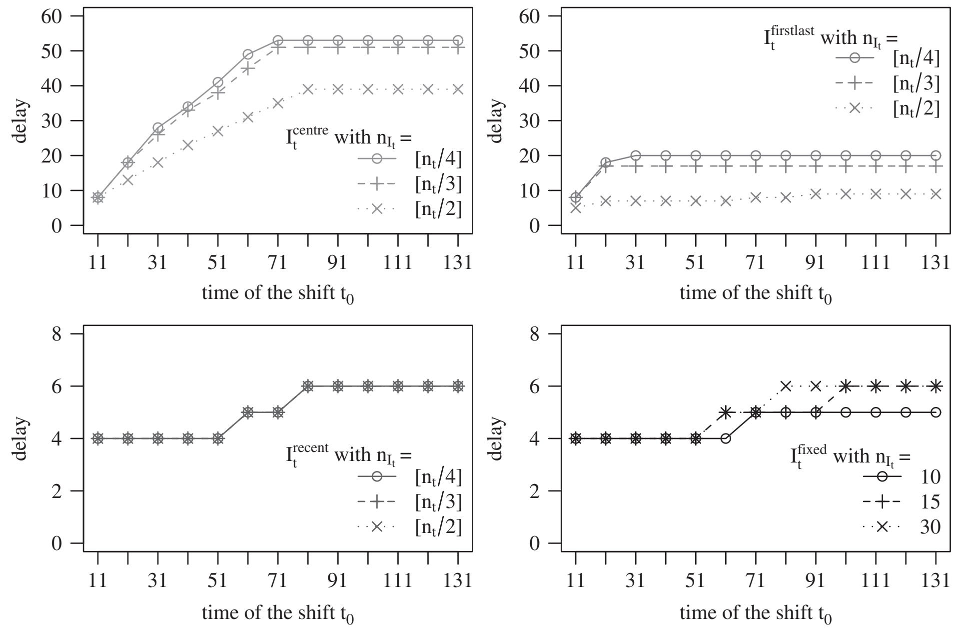

Figure 5. Time delay in tracing a noise-free level shift of size one at time $t_0 \in \{11, 21, \ldots, 131\}$ produced by adaptive RM filters with different settings.

As a measure of performance, we take the delay from the time t_0 when the true change has taken place to the time when the filter first reacts to the change, illustrated in Figure 4. Let $\hat{\mu}_t$ denote the signal estimated by an adaptive online RM filter at time t, the delay is then defined as

$$\text{delay} = \min\{t : |\hat{\mu}_t - \mu_1(t)| > 0, t > t_0\} - t_0$$

with $\mu_1(t)$ according to (10) for a shift and according to (11) for a trend change.

Using a different definition for the delay, e.g. the time difference until the estimated signal exceeds 50, 75, 90 or 95% of the difference between $\mu_1(t)$ and $\mu_2(t)$ does not change the findings presented in the following.

Figure 5 shows the delay of several adaptive RM filters needed to trace a level shift induced at different times t_0. A level shift after a large period of constancy (large t_0) induces a large window width n_{t_0} at the time of the shift and hence, Figure 5 shows, as expected, the time delay increases with increasing t_0. Filters using a test statistic based on the subsets I_t^{centre} or $I_t^{\text{firstlast}}$ show an almost linear increase for smaller t_0, while the delay using I_t^{recent} or I_t^{fixed} stays almost constant—regardless of t_0 or n_{t_0}, respectively.

Furthermore, taking a smaller portion n_{I_t} into account seems to increase the delay for filters applying a test statistic based on I_t^{centre} or $I_t^{\text{firstlast}}$, whereas the delay with I_t^{fixed} is a little lesser for some situations. When considering a certain number of residual signs depending on the current window width n_t according to I_t^{recent}, no difference in delay can be seen for the simulated settings. The outcomes for a trend change are analogous to the outcomes for a level shift and thus, are not shown here.

Concluding, adaptive RM filters based on a test statistic that takes into account the most recent residual signs by using I_t^{recent} or I_t^{fixed} result in the smallest time delay in tracing a shift or trend change if the data are noise-free. In these situations, the number of considered residual signs n_{I_t} seems to have no impact on the delay.

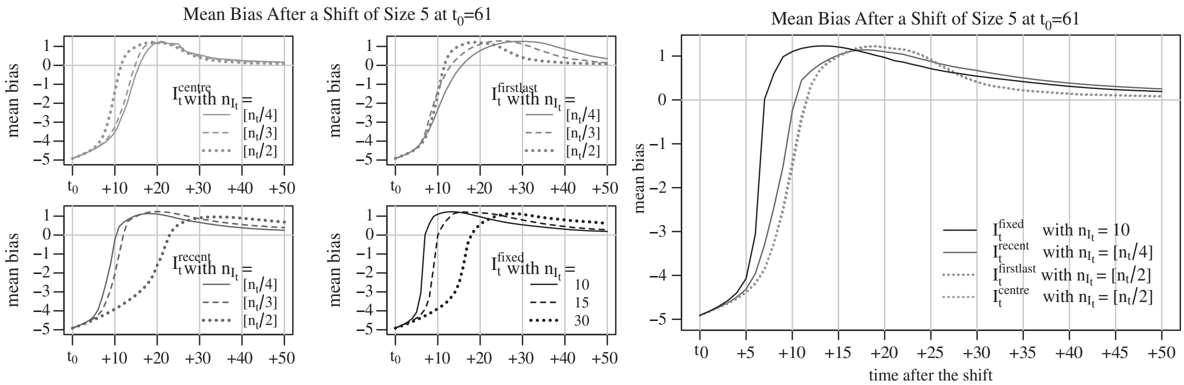

Figure 6. Mean bias after a level shift of size $h=5$ at time $t_0=61$ produced by several adaptive RM filters evaluated at 1000 simulation runs. The right panel compares the best curves from each of the left-hand side panels.

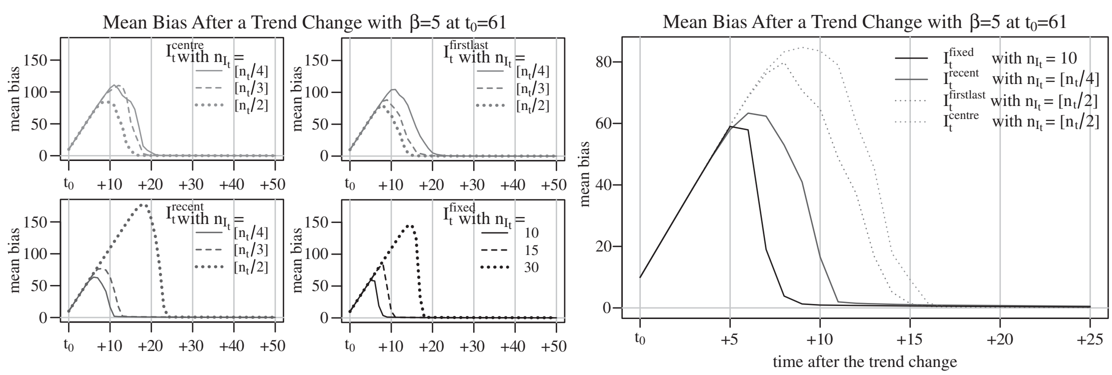

Figure 7. Mean bias after a trend change from a linear trend with $\beta=5$ to $\beta=-5$ at time $t_0=61$ produced by several adaptive RM filters evaluated at 1000 simulation runs. The right panel compares the best curves from each of the left-hand side panels.

4.2. Noisy level shifts and trend changes

The noisy data we consider are simulated from model (1), with the signal μ_t corresponding to (9) and either (10) for a level shift or (11) for a trend change. The errors are generated according to $\varepsilon_t \sim N(0,1)$ and no outliers are included, i.e. $v_t=0\ \forall t$. The situations we consider consist of

- level shifts of size $h\in\{1,2,5,10\}$ at $t_0\in\{11, 61, 121\}$ and
- trend changes with slopes $\beta\in\{1,2,5,10\}$ at $t_0\in\{11,61,121\}$.

To judge the performance of the different filter settings we use the mean and median bias right after a shift or trend change, respectively, evaluated on 1000 simulation runs. The variance and mean squared error (MSE) of the signal estimates are also investigated and commented on but will not be displayed for the sake of brevity.

Because of the robustness of the RM filters against patches of subsequent outliers, the estimated signal continues the linear trend defined by $\mu_1(t)$ for some time after t_0. For the level shift situations, this results in a bias of approximately $-h$ right after t_0, see, e.g. Figure 6. In case of a trend change, the linear trend after t_0 points into the opposite direction than the trend before, causing a linear increase of about twice the size of β for the bias as can be seen in Figure 7. A filter for which the bias curve tends more quickly towards zero after t_0 is said to be a method with shorter delay.

4.3. Level shifts

Figure 6 displays the mean bias for all considered RM filters after a level shift of size $h=5$ at $t_0=61$. The right-hand side plot compares the bias curves of the best performing filters for each of the four considered subsets I_t. It shows that filters based on a test statistic regarding a small number of the most recent residual signs outperform filters with other choices. In particular, the filter based on I_t^{fixed} with $n_{I_t}=10$ results in the smallest average time delay shown by the fact that the bias curve tends faster towards zero than for the other filters. This appears even more drastically with either, larger t_0 or larger shift size h.

For smaller shifts or shifts at an earlier time, the differences between the filters are not that obvious; for a level shift at $t_0=11$ almost no difference in mean or median bias is observed. Considering the median bias for large t_0 and large h, it is obvious that the estimated signal traces a sudden shift at some point in time, whereas the mean bias suggests a smooth and rather slow transition from the level before to the level after the shift. However, since the difference between median and mean bias is generally small, the median bias is not shown here. All the considered filters have a decreasing time delay for an increasing shift size h, but an increasing time delay for increasing t_0. However, this increase in delay is close to zero for a filter using I_t^{fixed} with small n_{I_t}, whereas it appears prominently for all other filters.

The simulation study shows, as expected, that the smaller the variance of the signal estimates, the larger the average window width n_t at a certain time t. Since the window width increases until the time of the shift t_0, the variability decreases with time. However, because each filter has a certain delay in tracing the shift which also depends on the simulated data around that time, the variability shows a peak right after t_0. This peak is almost negligible for small h and small t_0, but the peak is higher, the larger the h and t_0. Generally, this peak appears earlier for filters applying subsets I_t regarding a lesser number of recent most residual signs and also the variance curve over time returns quicker to low values while the increased variability can be present for a longer time for I_t^{centre} or $I_t^{\text{firstlast}}$ (not shown here).

Since the bias considerably dominates the MSE, this measure confirms the outcomes described above. The MSE curves for all filter settings show a large peak starting at t_0 and decrease thereafter, and the filter applying I_t^{fixed} with $n_{I_t}=10$ has the MSE curve that decreases the fastest. These curves are also excluded from the presentation.

4.4. Trend changes

Figure 7 shows the mean bias curves for all investigated filters for a trend change from $\beta=5$ to $\beta=-5$ at $t_0=61$. Again, the filters using test statistics based on the subsets I_t^{recent} and I_t^{fixed} with small n_{I_t} perform best w.r.t. bias and delay. For a trend change at $t_0=11$, a filter using I_t^{fixed} can perform worse than other filters if the trend slope is not very large ($\beta=1$ or $\beta=2$); for all other settings of β and t_0 the filter with I_t^{fixed} clearly shows the best results. The larger either β or t_0, the lesser the ‘delay’ in tracing the change.

Analogous to the outcomes for the level shifts, all signal estimates show a peak in their variance curves after t_0. Here, it is even more obvious that those filters applying I_t^{recent} or I_t^{fixed} with low n_{I_t} outperform the other filters because the variance is less and the variance curve decreases the fastest after t_0. Again, the MSE is strongly dominated by the bias and thus the MSE curves also imply using a filter based on I_t^{fixed} with $n_{I_t}=10$. Again, the outcomes for the variance and MSE are not illustrated here.

5. MODIFICATIONS OF THE ADAPTIVE RM FILTER

In the last section it could be seen that the filter output has a strong bias after a sudden level or trend change. In extreme situations this may cause the signal estimate to exceed the range of the observations. This is not really acceptable in practical applications: for example, if a physiological time series in intensive-care shows a drastic change within the alarm limits it is not the aim to trigger a threshold alarm. However, it is possible that the approximated signal may cause such an alarm because a continuation of a previous trend carries the estimation out of the alarm limits. Therefore, we restrict the estimated signal level $\hat{\mu}_t$ at time t to a value within the range of observations in the current time window, i.e.

$$\hat{\mu}_t = \begin{cases} \min(\mathbf{x}_{t,n_t}) & \text{if } \hat{\mu}_t^{\mathrm{RM}} < \min(\mathbf{x}_{t,n_t}) \\ \hat{\mu}_t^{\mathrm{RM}} & \text{if } \min(\mathbf{x}_{t,n_t}) \leqslant \hat{\mu}_t^{\mathrm{RM}} \leqslant \max(\mathbf{x}_{t,n_t}) \\ \max(\mathbf{x}_{t,n_t}) & \text{if } \max(\mathbf{x}_{t,n_t}) < \hat{\mu}_t^{\mathrm{RM}} \end{cases}$$

where $\min(\mathbf{x}_{t,n_t})$ denotes the minimum, $\max(\mathbf{x}_{t,n_t})$ denotes the maximum, and $\hat{\mu}_t^{\mathrm{RM}}$ denotes the RM estimate (4) evaluated at the window $\mathbf{x}_{t,n_t} = (x_{t-n_t+1}, \ldots, x_t)'$.

It is also possible to restrict the estimations to the interval limited by the minimum and maximum value of most recent, say n_{I_t}, observations to prevent that outliers observed at times that date back some time can have an influence on the recent estimation.

Another issue that needs to be addressed in practical applications is the treatment of missing values: intensive-care time series can contain missing values at single points in time due to short-term technical problems as well as long stretches of missings, e.g. caused by disconnection of the measurement devices. To ensure reliability of the signal estimation, a certain number of (non-missing) observations should be available for the estimation. It is conceivable to set the minimum required number of observations within one window equal to the value of the minimal window width to retain the robustness properties of the filter. If a window contains less observations, the filter returns a missing value for the estimated signal level. This ensures a continuous signal estimation in case of short-run technical failures, even if small window widths are applied.

6. EXAMPLES

This section compares the adaptive RM filters with the best settings of n_{I_t} for each of the four subsets I_t^{centre}, $I_t^{\mathrm{firstlast}}$, I_t^{recent}, and I_t^{fixed} determined in Section 4, also applying the modifications described in Section 5 to simulated and real time series. The comparisons also include filters based on fixed window widths.

First, we investigate some well-known examples from the regression context, typically used for evaluating the performance of non-parametric smoothers and filters. In particular, we take the Blocks function and the Doppler function described in [30] as a time series signal μ_t at equidistant points t according to (1). The time series observations are simulated by adding standard normal noise to the signal μ_t, which is re-scaled to achieve a signal-to-noise ratio of five, plus 5% positive, additive outliers of size five at random times. Figures 8 and 9 show the signals of these time series extracted by four adaptive RM online filters.

From the figures it can be derived that the min-/max-modification proposed in Section 5 does not avoid the over- or underestimation of the *true* signal after a level shift or trend change, but at least it prevents that the estimated signal falls outside the observational range. Such situations can for example be seen in Figure 8, e.g. after the negative level shifts around times 250 and 400, where the signal approximation is truncated by the minimum of the observations in the current window. Furthermore, filters based on a local linear fit tend to be biased in regions of curvature. If the estimate is evaluated in the centre of a certain neighbourhood, such filters tend to 'trim the hills and fill the valleys' (cf. [31, p. 171]). For an online application, Figure 9 shows that the proposed online filters, based on an estimate at the end of each window, tend to 'augment the hills and deepen the valleys'.

In the first part of the Doppler signal (Figure 9), it is difficult to distinguish signal from noise, hence, all filters yield estimated signals around zero. The filter with I_t^{fixed} and $n_{I_t} = 10$ traces the true signal closely from about time $t = 100$ on while for the other filters the estimated and the true signal are close by only from about time $t = 200$ on. Thus, the I_t^{fixed}-RM filter has a better ability to trace patterns occurring in short time intervals. The smallest delay in tracing sudden shifts can also be observed for this filter, see, Figure 8.

Figure 10 compares the performance of the adaptive RM filter using the subset I_t^{fixed} and $n_{I_t} = 10$ with two online RM filters based on fixed window widths $n = 15$ and $n = 50$ at a simulated time series. The figure illustrates that all these filters are robust against artefacts,

DOI: 10.1002/acs

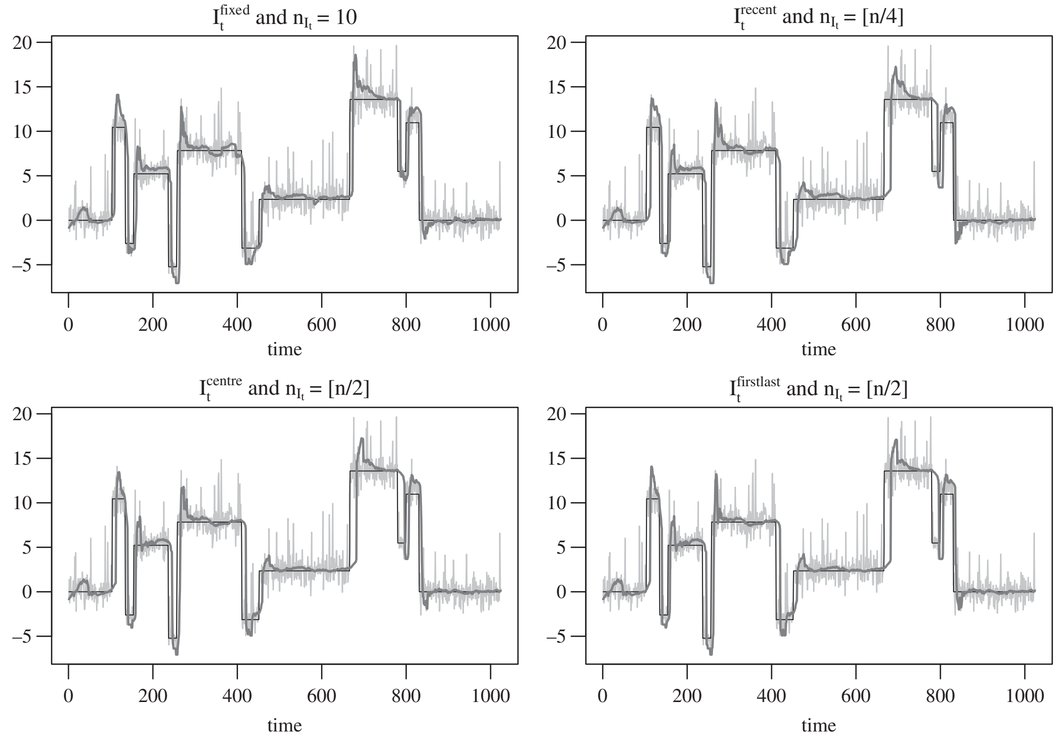

Figure 8. Signal of adaptive RM filters (dark grey) estimating the Blocks signal (black) from a time series (light grey) with a signal-to-noise ratio of 5 and 5% outliers of size 5.

e.g. around time $t=300$. The filter with large, fixed window width yields the smoothest signal estimations but also has the largest delay in tracing changes, most obvious for the shift at time $t=250$. Furthermore, the estimated signal for this filter lies outside the observational range after the trend change around time $t=75$, which could be prevented by applying the min-/max-modification proposed in Section 5. The RM filter based on $n=15$ follows changes in the data with the smallest time delay but exhibits much more variation around the true signal than the other filters in constant periods, e.g. from $t=300$ to 350. Since the underlying data structure is not known in advance for intensive-care time series, the example shown in Figure 10 illustrates the advantages of an RM filter with a data-driven choice of the window width: the signal approximation is not only smooth in times of constant or linear trend, but is also able to follow sudden changes with small delay.

As real data examples, we consider 1 h of systolic arterial blood pressure and half an hour of heart rate measurements displayed in Figure 11. Both time series are measured and stored once per second. In addition to the measurements, the graphics display the upper and lower alarm limits set by the medical staff. Furthermore, the figure shows the online signal extracted by an adaptive RM filter with I_t^{fixed}. We used $n_{I_t}=30$ for these two examples because application to several intensive-care time series showed that a filter with these settings provides online signal estimates that are sufficiently smooth and have an acceptable time delay for the detection of relevant sudden changes. Choosing a smaller value for n_{I_t} causes window width reduction for changes that may not be relevant from a medical

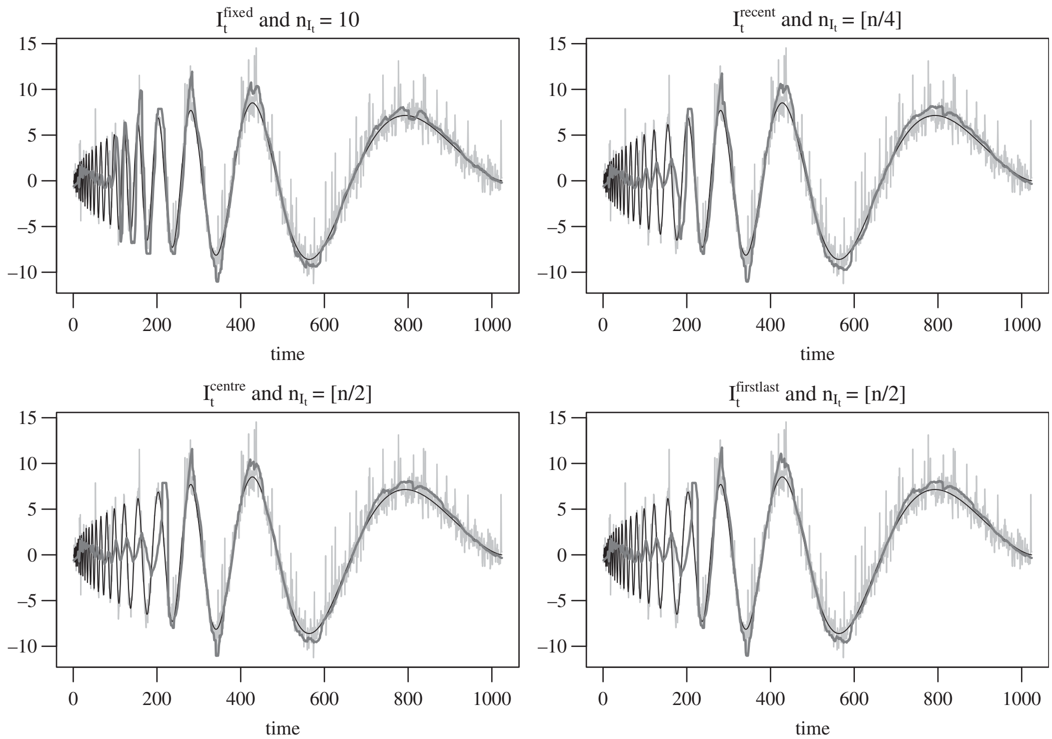

Figure 9. Signal of adaptive RM filters (dark grey) estimating the Doppler signal (black) from a time series (light grey) with a signal-to-noise ratio of 5 and 5% outliers of size 5.

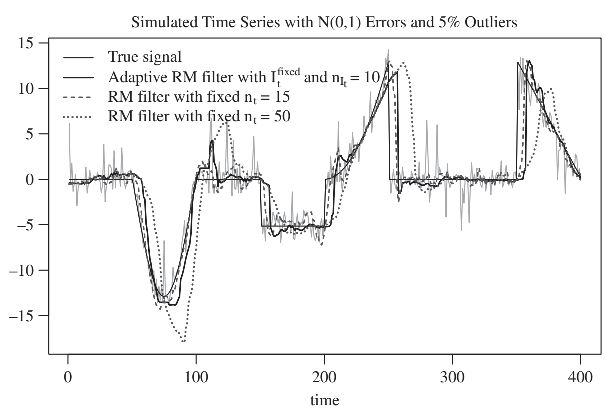

Figure 10. Comparison of two RM online filters with fixed window widths $n=15$ and $n=50$ with an adaptive RM filter with I_t^{fixed} and $n_{I_t}=10$ at a simulated time series with standard normal noise, a signal-to-noise ratio of five, and 5% additive outliers of size five with random sign at random times.

DOI: 10.1002/acs

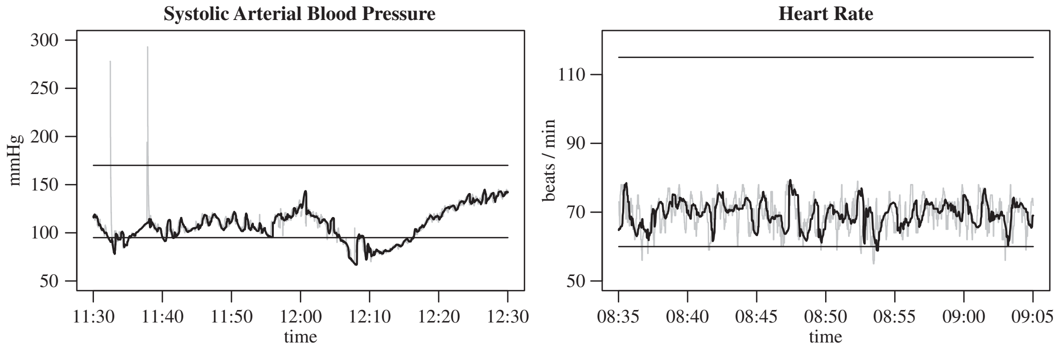

Figure 11. Physiological time series with upper and lower alarm limits and a signal estimated by an adaptive RM filter using I_t^{fixed} with $n_{I_t} = 30$.

perspective and hence results in a more variable signal estimation.

For the blood pressure, the currently used alarm system triggers two artefact alarms caused by outliers, which violate the upper alarm limit while the true blood pressure is close to the lower alarm limit. For the heart rate, a number of alarms are triggered because of lower limit violations while from a medical perspective it would rather be of interest that the heart rate is close to the lower limit with additional information about the variability.

Using the signal extracted by the adaptive RM filter with I_t^{fixed} as input to the threshold alarm system not only prevents the artefact alarms, displayed in the left panel of Figure 11, but also indicates relevant violations of the alarm limits. The right panel of Figure 11 illustrates that the total number of alarms can also be reduced because the estimated signal crosses the alarm threshold less often than the observed measurements but still transports the relevant information.

7. CONCLUSIONS

We propose an online time series filter applying repeated median (RM) regression to a moving time window with a data-driven choice of the window width. Several possibilities for adaptive window width selection are compared with simulations and applications.

The proposed adaptive filter yields a smooth signal approximation, which is robust against isolated artefacts and small patches of outlying values. It is able to trace sudden changes with small time delay without applying different tests for artefact, trend change, and level shift detection. Furthermore, the filter does not require in any parameter specifications, does not put strong assumptions on the underlying data structure, and is applicable to time series containing missing values.

We recommend a filter choosing the window width based on the sum of a fixed number of the most recent signs of RM residuals. This number of residual signs should be independent of the window width used for the signal estimation at a certain time. For adaptation to sudden data changes with minimal time delay, we suggest to use a small number of say, 10, most recent residual signs. However, this number should be chosen based on the application-oriented background, e.g. on the maximally acceptable time delay and the frequency of measurement.

For example, in applications to high-frequency data like time series from an online monitoring system in intensive-care, one may take a larger number of residual signs, e.g. at the 30 most recent points in time, to ensure that medically irrelevant sudden changes are ignored while relevant ones are traced with acceptable time delay. Using the extracted signals instead of the raw measurements as input to a threshold alarm system can thus lead to a considerable reduction of false-positive alarms.

DOI: 10.1002/acs

In [32] the proposed methodology of adaptive window width choice has been transferred to filters extracting signals from multivariate time series based on methods proposed in [33]. Applying such multivariate filters to highly correlated intensive-care time series, such as systolic, mean and diastolic blood pressures, seems a promising approach to an even stronger improvement of currently used threshold alarm systems. Combining information about the signal level with information about the local variability, e.g. extracted by online filters as proposed in [34, 35], offers a further possibility for improvement, for example, by developing new alarm rules for early warnings.

This paper introduces a new procedure and focusses on the best settings for the proposed method. Comparisons with other filtering methods still have to be carried out. Furthermore, the performance of the proposed filter in various data situations, e.g. for time series data with autocorrelated or heteroscedastic errors, is unknown and thus a topic of future research.

ACKNOWLEDGEMENTS

This work has been accomplished in collaboration with Dr M. Imhoff within a project of the collaborative research centre 475 'Reduction of Complexity in Multivariate Data Structures' financially supported by the Deutsche Forschungsgemeinschaft.

REFERENCES

1. Imhoff M, Kuhls S. Alarm algorithms in critical care monitoring. *Anesthesia and Analgesia* 2006; **102**:1525–1537.
2. Tukey JW. *Exploratory Data Analysis*. Addison-Wesley: Reading, MA, 1977.
3. Cleveland WS. Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association* 1979; **74**:829–836.
4. Davies PL, Fried R, Gather U. Robust signal extraction for on-line monitoring data. *Journal of Statistical Planning and Inference* 2004; **122**:65–78.
5. Gather U, Schettlinger K, Fried R. Online signal extraction by robust linear regression. *Computational Statistics* 2006; **21**(1):33–51.
6. Siegel AF. Robust regression using repeated medians. *Biometrika* 1982; **69**(1):242–244.
7. Fried R. Robust filtering of time series with trends. *Journal of Nonparametric Statistics* 2004; **16**(3):313–328.
8. Gather U, Fried R. Methods and algorithms for robust filtering. In *Proceedings in Computational Statistics* (*COMPSTAT 2004*), Antoch J (ed.). Physica: Heidelberg, 2004; 159–170.
9. Fried R, Bernholt T, Gather U. Repeated median and hybrid filters. *Computational Statistics and Data Analysis* 2006; **50**(9):2313–2338.
10. Bernholt T, Fried R, Gather U, Wegener I. Modified repeated median filters. *Statistics and Computing* 2006; **16**(2):177–192.
11. Fried R, Gather U. On rank tests for shift detection in time series. *Computational Statistics and Data Analysis* 2007; **52**:221–233.
12. Fried R. On the robust detection of edges in time series filtering. *Computational Statistics and Data Analysis* 2007; **52**(2):1063–1074.
13. Fried R, Einbeck J, Gather U. Weighted repeated median smoothing and filtering. *Journal of the American Statistical Association* 2007; **102**:1300–1308.
14. Brockmann M, Gasser T, Herrmann E. Locally adaptive bandwidth choice for kernel regression estimators. *Journal of the American Statistical Association* 1993; **88**:1302–1309.
15. Fan J, Hall P, Martin A, Patil P. On local smoothing of nonparametric curve estimators. *Journal of the American Statistical Association* 1996; **91**:258–266.
16. Friedman JH, Stützle W. Projection pursuit regression. *Journal of the American Statistical Association* 1981; **76**:817–823.
17. Fan J, Gijbels I. Data-driven bandwidth selection in local polynomial fitting: variable bandwidth and spatial adaptation. *Journal of the Royal Statistical Society*, *Series B* 1995; **57**(2): 371–394.
18. Fan J, Gijbels I. *Local Polynomial Modelling and Its Applications*. Chapman & Hall: London, 1996.
19. Schucany WR. Adaptive bandwidth choice for kernel regression. *Journal of the American Statistical Association* 1995; **90**:535–540.
20. Cleveland WS, Loader CL. Smoothing by local regression: principles and methods. In *Statistical Theory and Computational Aspects of Smoothing*, Härdle W, Schimek MG (eds). Springer: New York, 1996; 10–49.
21. Ruppert D. Empirical-bias bandwidths for local polynomial nonparametric regression and density estimation. *Journal of the American Statistical Association* 1997; **92**:1049–1062.
22. Goldenshluger A, Nemirovski A. On spatial adaptive estimation of nonparametric regression. *Mathematical Methods of Statistics* 1999; **6**(2):135–170.
23. Katkovnik V. A new method for varying adaptive bandwidth selection. *IEEE Transactions on Signal Processing* 1999; **47**(9):2567–2571.
24. Katkovnik V, Egiazarian K, Shmulevich I. Adaptive varying window size filtering based on intersection of confidence intervals rule. *Proceedings of the IEEE-EURASIP Workshop*

on Nonlinear Signal and Image Processing (*NSIP-01*), Baltimore, MD, U.S.A., 3–6 June 2001.

25. Chan S, Zhang Z. Robust local polynomial regression using M-estimator with adaptive bandwidth. *Proceedings of the 2004 International Symposium on Circuits and Systems* (*ISCAS 04*), Vancouver, BC, Canada, 23–26 May 2004, *Circuits and Systems* 2004; 333–336. URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=1328751&isnumber=29374.
26. Bernholt T, Fried R. Computing the update of the repeated median regression line in linear time. *Information Processing Letters* 2003; **88**(3):111–117.
27. Rousseeuw PJ, Leroy AM. *Robust Regression and Outlier Detection*. Wiley: New York, 1987.
28. Hössjer O, Rousseeuw P, Ruts I. The repeated median intercept estimator: influence function and asymptotic normality. *Journal of Multivariate Analysis* 1995; **52**(1):45–72.
29. Gather U, Fried R, Imhoff M. Online classification of states in intensive care. In *Data Analysis. Scientific Modeling and Practical Application*, Gaul W, Opitz O, Schader W (eds). Springer: Berlin, 2000; 413–428.
30. Donoho DL, Johnstone IM. Adapting to unknown smoothness via wavelet shrinkage. *Journal of the American Statistical Association* 1995; **90**(432):1200–1224.
31. Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning* (10th edn). Springer: New York, 2001.
32. Borowski M, Schettlinger K, Gather U. Multivariate real time signal extraction by a robust adaptive regression filter. *Communications in Statistics—Simulation and Computation* 2009; **38**(2):426–440.
33. Lanius V, Gather U. Robust online signal extraction from multivariate time series. *Technical Report 38/2007, SFB 475*, Technische Universität Dortmund, 2007.
34. Gelper S, Schettlinger K, Croux C, Gather U. Robust online scale estimation in time series: a model-free approach. *Journal of Statistical Planning and Inference* 2009; **139**: 335–349.
35. Nunkesser R, Fried R, Schettlinger K, Gather U. Online analysis of time series by the Q_n estimator. *Computational Statistics and Data Analysis* 2008; DOI: 10.1016/j.csda.2008.02.027.