Nonparametric Statistics
Vol. 16(3–4), June–August 2004, pp. 313–328

ROBUST FILTERING OF TIME SERIES WITH TRENDS

ROLAND FRIED*

Department of Statistics, University of Dortmund, 44221 Dortmund, Germany

(Received 17 October 2002; Revised 6 November 2003; In final form 8 November 2003)

We develop and test a robust procedure for extracting an underlying signal in form of a time-varying trend from very noisy time series. The application we have in mind is online monitoring data measured in intensive care, where we find periods of relative constancy, slow monotonic trends, level shifts and many measurement artifacts. A method is needed which allows a fast and reliable denoising of the data and which distinguishes artifacts from clinically relevant changes in the patient's condition. We use robust regression functionals for local approximation of the trend in a moving time window. For further improving the robustness of the procedure, we investigate online outlier replacement by *e.g.* trimming or winsorization based on robust scale estimators. The performance of several versions of the procedure is compared in important data situations and applications to real and simulated data are given.

Keywords: Online monitoring; Signal extraction; Level shift; Trend; Outlier; Bias curve

1 INTRODUCTION

In intensive care, physiological variables like the heart rate are recorded at least every minute. Methods working in real time are needed which extract the underlying clinically relevant signal from the observed time series while resisting the frequent irrelevant measurement artifacts, which often even emerge as patches of several subsequent outliers [1]. Median filtering is often applied to smooth noisy time series [2], and simple rules for detection of atypical observations (outliers) and sudden changes (level shifts) in the data generating mechanism based on the median absolute deviation (MAD) about the median are sometimes used in addition. However, in view of high sampling frequencies most changes occur gradually. In trend periods, most scale estimators like the MAD are strongly biased, and a running median loses a lot of its robustness [3].

A procedure for extraction of a time-varying deterministic trend from the noisy time series observed in intensive care needs to work automatically and online. It must behave well, or at least not disastrously, in many different aspects and situations as any flaw might be life-threatening. Therefore, we prefer construction of a procedure with specific properties rather than achieving optimality in a single sense [4]. Some basic criteria are the existence of a unique solution, low computation time, high robustness against outliers and satisfactory finite-sample efficiency.

Similar to a running median, we use a moving time window with a fixed width regulated by the requirement of working online. Instead of approximating a local level by the median

* Corresponding author. E-mail: fried@statistick.uni-dortmund.de

ISSN 1048-5252 print; ISSN 1029-0311 online © 2004 Taylor & Francis Ltd
DOI: 10.1080/10485250410001656444

we fit a linear trend to the data in each time window using regression functionals with high breakdown points like the least median of squares and the repeated median. For increasing the robustness of the procedure we add automatic rules for outlier detection and replacement based on robust scale estimators. Besides the classical MAD, we investigate other high breakdown point methods, namely the length of the shortest half (LSH) and Rousseeuw and Croux's Q_α and SN [5]. On the basis of an estimate of the local variability we can detect and replace outliers online before they influence the data analysis. We test standard strategies like trimming, winsorization and variations of these for this purpose. To find out the strengths and the weaknesses of the various possible combinations, we perform an extensive simulation study and apply them to some time series.

There is a lot of further work on signal extraction from noisy data using linear or non-linear filters like a running mean or median. The advantages of a local linear instead of a local constant fit have been emphasized in Ref. [6]. In Ref. [7] a procedure is developed based on local linear M-smoothing, which preserves sudden shifts in the data, but it does not aim at distinguishing short term outlier patches from long term changes, which is the goal here.

We proceed as follows. In Section 2, we develop a procedure which combines robust functionals for regression and scale estimation and we describe some modifications for an automatic application. In Section 3, we report the results of a simulation study. In Section 4, applications to real and simulated time series are presented before we draw some conclusions.

2 METHODS

Let Y_t be a real-valued random variable with median μ_t and variance σ_t^2 at time $t \in \mathbb{N}$, and $y_1, \ldots, Y_N$ be a realization of $Y_1, \ldots, Y_N$. The aim in the following is to extract the signal formed by the sequence μ_t, $t = 1, \ldots, N$, from the data.

In intensive care monitoring data, we often find systematic drifts corresponding to slow monotonic changes in μ_t, and the variance σ_t^2 may vary over time as well [8]. The idea underlying the following methods for signal extraction is that the signal varies smoothly with a few abrupt changes, while the noise is rough, and we assume that the observation times are sufficiently dense for approximating the signal in a moving time window of fixed width n by a straight line. This means a straightforward modification of a running median replacing a location by a linear trend estimate within each window [1]. The time delay admissible further restricts the choice of n, while n should be chosen large adhering to these restrictions to reduce the variance and the effects of outlier patches. For our clinical application, we use windows of length $n = 2m + 1 = 31$ observations.

2.1 Model and Assumptions

A referee recommended formulation of a global model for making the previous terms precise. It is hard to specify realistic assumptions for physiological data measured in intensive care, but the following might help the reader understand the basic problems and ideas. We aim at a method which is able to cope with data from a model like

$$Y_t = \mu_t + E_t + \sum_{k=1}^{K} \sum_{i=0}^{\delta_k - 1} \omega_{k,i} 1_{\left\{\sum_{j=1}^{k} T_j + i\right\}}(t) \tag{1}$$

$$\mu_t = \eta_t + \sum_{l=1}^{L} \nu_l 1_{\left[\sum_{j=1}^{l} S_j, \infty\right]}(t). \tag{2}$$

Here, $1_I(t)$ indicates whether t is in a set I or not, and E_t represents a random error at time point t generated from a distribution with median zero and variance σ_t^2. K and L denote the possibly infinite numbers of outlier patches and shifts respectively, T_j, S_j, $j \in \mathbb{N}$, are the times between subsequent outliers and level shifts, and δ_k is the duration of the kth outlier patch with $\delta_k = 1$ meaning an isolated outlier. Furthermore, $\omega_{k,i}$ denotes the size of the ith outlier in the kth patch, while ν_l is the size of the lth level shift. For simplicity, we treat the errors as white noise neglecting correlations. Similar models have been used in a Bayesian framework regarding all variables as random [9] although proper specification of the dependencies is hardly possible. The observation equation (1) describes additive outliers only as we simply add a constant to the undisturbed value $\mu_t + E_t$. For incorporating substitutive outliers for instance, we need to replace Y_t by a constant at the corresponding time points. Substitutive outliers are useful to describe pure measurement artifacts, while additive outliers describe a kind of shock.

For separating signal and noise, we assume η_t and σ_t to vary smoothly in time. The data are on a fixed grid, so assuming that the η_t stem from a continuous or differentiable function η would allow arbitrarily large changes between subsequent time points. Instead we might assume that for a sufficiently small choice of $n = 2m + 1$, there is a small $\varepsilon > 0$ and for every $t \in \mathbb{N}$ a $\beta_t \in \mathbb{R}$ such that the relative change $|\sigma_{t+j} - \sigma_t|/\sigma_t \leq \varepsilon$ and the scaled distance $|\eta_{t+j} - \eta_t - j\beta_t|/\sigma_t \leq \varepsilon$ for $j \in \{-m, \dots, m\}$. These assumptions, *e.g.*, hold if σ_t is constant, $\sigma_t = \sigma$, and the η_t stem from a real function η with second derivative $|\eta''| \leq 2\sigma\varepsilon/m^2$, but this is not necessary. Note that if σ_t is bounded above, this approximation cannot hold for every $\varepsilon > 0$ unless the sequence (η_t) forms a straight line, which is restrictive, or we set m to zero, which is not reasonable for noise reduction. Moreover, we distinguish outlier patches and shifts by their duration assuming that the δ_k are bounded to be less than an upper bound d_u, and that there is a lower bound $d_l \geq d_u$ for the time $S_{l+1} - S_t$ between subsequent shifts. In our clinical application, a medical rule of thumb states that five subsequent observations (measured once per minute) which are about the same size and differ substantially from the proceeding observations are often clinically relevant, *i.e.* we set $d_u = 5$.

If possible, we should choose the window width n such that $d_u \leq \lfloor n/2 \rfloor$ and $n \leq d_l$. Then a single outlier patch always affects less than 50% of the observations in a time window and two shifts never happen in the same window. Extraction of μ_t is difficult if outlier patches occur in short time lags or close to a level shift. These are common problems in intensive care monitoring. A method should better point at a shift and not ignore patterns which might be caused by either a level shift or subsequent outlier patches to attract the attention of the health care professional.

2.2 Robust Trend Approximation

As we apply high breakdown point methods, we neglect outliers and level shifts at first and approximate $\mu_{t+i}, i = -m, \dots, m$, in the current window centered at $t \geq m + 1$ by a straight line,

$$Y_{t+i} = \mu_t + i\beta_t + E_{t+i} + r_{t+i}, \quad i = -m, \dots, m, \tag{3}$$

where μ_t and β_t represent level and slope in the window, E_{t+i} is random noise with median zero and variance σ^2, and r_{t+i} is the approximation error, $|r_{t+i}| \leq \varepsilon\sigma_t$. Obviously, ε and thus the choice of m regulates the model bias.

A comparative study [1] investigates the finite-sample properties of the least median of squares functional T_{LMS} [10, 11] and of the repeated median functional T_{RM} [12]. For regression against time with data measured at equally spaced time points $t + i, i = -m, \dots, m$, these read

$$T_{\mathrm{LMS}} = \operatorname{argmin}\{(\mu, \beta)\colon \operatorname{median}(y_{t+i} - \mu - i\beta)^2\}, \tag{4}$$

and $T_{\mathrm{RM}} = (\tilde{\mu}_t, \tilde{\beta}_t)$ with

$$\tilde{\beta}_t = \mathrm{med}_i \left(\mathrm{med}_{j \neq i} \frac{y_{t+i} - y_{t+j}}{i - j} \right) \tag{5}$$

$$\tilde{\mu}_t = \mathrm{med}_i (y_{t+i} - i \tilde{\beta}_t),$$

respectively. Application of T_{RM} or T_{LMS} to the observations in a time window allows approximation of the level and the slope in the center of the window. For approximation of the level and the slope at the first and the last m time points of the series, we can use the level $\tilde{\mu}_t + i\tilde{\beta}_t$ and the slope approximates $\tilde{\beta}_t$ fitted in the first and the last time window, respectively.

The T_{RM} and the T_{LMS} both have the optimal breakdown point for a regression-equivariant line estimator, that is $\lfloor n/2 \rfloor / n$. Important advantages of T_{RM} are its smaller computation time [13], its smaller variance and MSE in case of a small to moderate number of outliers [6] and the instability of T_{LMS} in case of small changes in the data [14]. Unlike the T_{LMS}, the T_{RM} is Lipschitz-continuous for a fixed design and hence small changes in the data do not cause large changes of the results. On the other hand, T_{LMS} has a much smaller bias and mean square error (MSE) than T_{RM} when there is a large percentage of 30% or more outliers in a single time window. The T_{LMS} is typically even less influenced by large than by small outliers as it may ignore the former completely, while T_{RM} is more affected by large outliers. Therefore, replacing outliers may well improve the performance of T_{RM}, while this is not necessarily true for T_{LMS}. For all these reasons, we use T_{LMS} merely as a benchmark for robustness and try to improve T_{RM} to become similarly robust as T_{LMS}. An obvious way for doing this is to replace outliers online based on an approximation of the local variability.

2.3 Scale Approximation

As we assume σ_t to vary smoothly we can approximate it applying a scale estimator to the residuals $r_i = y_{t+i} - \tilde{\mu}_t - \tilde{\beta}_t i$, $i = -m, \ldots, m$, in the window. The classical robust method is the MAD

$$\tilde{\sigma}_{\mathrm{MAD}} = c_{1,n} \mathrm{med}\{|r_{-m}|, \ldots, |r_m|\},$$

where $c_{1,n}$ is a finite-sample correction depending on the window width $n = 2m + 1$. Many other robust scale estimators have been suggested. In Ref. [15] the properties of robust scale estimators are inspected in the regression setting. It turns out that LSH [16, 17]

$$\tilde{\sigma}_{\mathrm{LSH}} = c_{2,n} \min\{|r_{(i+m)} - r_{(i)}|; i = 1, \ldots, n - m\},$$

where $r_{(1)} \leq r_{(1)} \leq \cdots \leq r_{(n)}$ are the ordered residuals, and Rousseeuw and Croux's [5] suggestion

$$\tilde{\sigma}_{\mathrm{QN}} = c_{3,n} \{|r_i - r_j| \colon -m \leq i < j \leq m\}_{(h)}, \quad h = \binom{m+1}{2}$$

are particularly interesting. An algorithm for computation of $\tilde{\sigma}_{\mathrm{QN}}$ in $\mathrm{O}(n \log n)$ time is presented in Ref. [18]. Moreover, we consider the nested scale statistic

$$\tilde{\sigma}_{\mathrm{SN}} = c_{4,n} \mathrm{med}_i \, \mathrm{med}_{j \neq i} |r_i - r_j|,$$

proposed in Ref. [5]. The $\tilde{\sigma}_{\mathrm{LSH}}$ shows extremely good resistance against a large percentage of outliers [see also Ref. 19]. The $\tilde{\sigma}_{\mathrm{QN}}$ and $\tilde{\sigma}_{\mathrm{MAD}}$ perform better for inliers like identical measurements, and $\tilde{\sigma}_{\mathrm{QN}}$ works very well in case of a level shift, where we sample from shifted distributions. In conclusion, $\tilde{\sigma}_{\mathrm{QN}}, \tilde{\sigma}_{\mathrm{SN}}$ and $\tilde{\sigma}_{\mathrm{LSH}}$ (in this ordering) have large finite-sample efficiencies.

2.4 Outlier Detection

Application of a scale estimator allows to check whether the incoming observation y_{t+m+1} is an outlier by comparing the extrapolation residual $r_{m+1} = y_{t+m+1} - \tilde{\mu}_t - \tilde{\beta}_t(m+1)$ to an estimate $\tilde{\sigma}_t$ of σ_t. A general strategy for online cleaning is to replace y_{t+m+1} by

$$\tilde{y}_{t+m+1} = \tilde{\mu}_t + \tilde{\beta}_t(m+1) + d_1 \mathrm{sgn}\,(r_{m+1})\tilde{\sigma}_t \quad \text{if } |r_{m+1}| > d_0\tilde{\sigma}_t, \tag{6}$$

where $0 \leq d_1 \leq d_0$ are constants and sgn is the signum function. The idea underlying $d_1 = 0$ is that outliers do not provide relevant information and should be set to a prediction, while $d_1 > 0$ is reasonable if we do not regard outliers as artifacts but as disturbed values. $d_1 = d_0 > 0$ means winsorization. Additive outliers may provide some relevant information for signal extraction while substitutive outliers do not. Intuitively, in the short run we expect $d_0 > d_1 = 0$ to give more stable results for the level and the slope, and $d_0 = d_1 > 0$ for the scale. In the long run this is not as clear as these goals interact. A choice $d_0 > d_1 > 0$ may be a compromise. Finding an overall optimal choice of d_0 and d_1 does not seem possible since there are several criteria (approximation of μ_t, β_t, σ_t) and outlier generating mechanisms. Therefore, we restrict to the following heuristic choices:

T	$d_0 = 3,$	$d_1 = 0$	(trimming)
L	$d_0 = 3,$	$d_1 = 1$	(downsizing large values)
M	$d_0 = 2,$	$d_1 = 1$	(downsizing moderate values)
W	$d_0 = 2,$	$d_1 = 2$	(winsorization)

A preliminary study showed that for trimming detected outliers should be treated as missing values in the scale approximation and the finite-sample correction adjusted for the reduced number of observations. Otherwise, we possibly underestimate the scale largely. In the other strategies, we use the adjusted time series without correcting the sample size.

Outlier detection cannot be performed online in the first time window. Instead, we check the observations retrospectively for outlyingness using an initial fit and analogous rules as stated above. If we find outliers we reanalyze the first time window with detected outliers being replaced.

2.5 The Procedure

Now we formulate a basic algorithm based on the previous components. Let $\tilde{\sigma}$ be any of the scale estimators, and $m \in \mathbb{N}$ as well as $d_0 \geq d_1 \geq 0$ be given constants. The input of the algorithm is a time series $y_j,\ j \in \mathbb{N}$. We set $\tilde{y}_j = y_j$, $\mathrm{out}(j) = 0$, $j \in \mathbb{N}$, and $t = m + 1$.

1. Estimate $\tilde{\mu}_t$, $\tilde{\beta}_t$ and $\tilde{\sigma}_t$ from $\tilde{y}_{t+j},\ j = -m, \ldots, m$, using T_{RM} and $\tilde{\sigma}$.
2. Replace all $\tilde{y}_{t+j},\ j = -m, \ldots, m$ for which $|r_j| = |\tilde{y}_{t+j} - \tilde{\mu}_t - j\tilde{\beta}_t| > d_0\tilde{\sigma}_t$ by $\tilde{\mu}_t + j\tilde{\beta}_t + \mathrm{sgn}(r_j)d_1\tilde{\sigma}_t$ and set $\mathrm{out}(t+j) = \mathrm{sgn}(r_j)$ for these j.
3. If $\#\{j = -m, \ldots, m\colon \mathrm{out}(t+j) = 1\} > m$ re-replace these $\tilde{y}_{t+j}$ and $\mathrm{out}(t+j)$ by y_{t+j} and 0 respectively. Act in the same way if $\#\{j = -m, \ldots, m\colon \mathrm{out}(t+j) = -1\} > m$.
4. If $\#\{j = -m, \ldots, m\colon |\mathrm{out}(t+j)| = 0\} < \max\{\lfloor m/3 \rfloor, 5\}$ reset all $\tilde{y}_{t+j}$ to y_{t+j} and $\mathrm{out}(t+j)$ to zero, $j = -m, \ldots, m$.

5. Estimate $\tilde{\mu}_t$, $\tilde{\beta}_t$ and $\tilde{\sigma}_t$ from $\tilde{y}_{t+j}$, $j = -m, \ldots, m$, using T_{RM} and $\tilde{\sigma}$.
6. Set $\tilde{y}_{t+m+1}$ to $\tilde{\mu}_t + \tilde{\beta}_t(m+1) + \mathrm{sgn}(r_{m+1})d_1\tilde{\sigma}_t$ and $\mathrm{out}(t+m+1)$ to $\mathrm{sgn}(r_{m+1})$ if $|r_{m+1}| = |y_{t+m+1} - \tilde{\mu}_t - \tilde{\beta}_t(m+1)| > d_0\tilde{\sigma}_t$.
7. Increase t to $t+1$ and go to 3.

Steps 3 and 4 have been added because of some infrequent, but severe problems with automatic outlier replacement. It may happen that many observations are replaced within a short time period and then the scale estimate can approach zero because of a bad regression fit or strong underestimation of the variance. Both reasons can result in many incoming observations being regarded as outliers causing a vicious circle. We found steps 3 and 4 to be very helpful to overcome these problems: if more than half of the observations within a window have been regarded as positive (negative) outliers and replaced, the regression line might underestimate (overestimate) the true levels. Using the original observations which have been predicted too small (large) while still using the replacements for the detected negative (positive) outliers can then improve the results. If very few observations within a time window have not been replaced at all we should use all the original observations in the regression since possibly the variability is strongly underestimated.

To simplify the simulation study in Section 3, we prove that the outcome of the procedure does not depend on a constant trend. The following lemma states some invariance properties of a filtering procedure as described above. For its formulation we use the operators $f\colon \mathbb{R}^N \mapsto \mathbb{R}^{2N-4m}$, $(Y_1, \ldots, Y_N)' \mapsto (\tilde{\mu}_{m+1}, \tilde{\beta}_{m+1}, \ldots, \tilde{\mu}_{N-m}, \tilde{\beta}_{N-m})'$ and $\hat{\sigma}\colon \mathbb{R}^N \mapsto \mathbb{R}^{N-2m}$, $(y_1, \ldots, y_N)' \mapsto (\tilde{\sigma}_{m+1}, \ldots, \tilde{\sigma}_{N-m})$ which map a time series to its decomposition into local level and slope and local scale, respectively, obtained by application of a filtering procedure as described above.

LEMMA *If a regression- and affine-equivariant regression functional T and an affine-equivariant scale estimator $\tilde{\sigma}$ are used in the filtering procedure described above then it fulfills the following equivariance properties, in which $\boldsymbol{\theta} \in \mathbb{R}^2$ and $\alpha \in \mathbb{R}$ are arbitrary constants:*

$$f(\alpha\mathbf{y} + \mathbf{X}\boldsymbol{\theta}) = \alpha f(\mathbf{y}) + 1 \otimes \boldsymbol{\theta},$$

$$\hat{\sigma}(\alpha\mathbf{y} + \mathbf{X}\boldsymbol{\theta}) = |\alpha|\hat{\sigma}(\mathbf{y}),$$

where $\mathbf{y} = (y_1, \ldots, y_N)'$ is the vector of all observations, $\mathbf{X}$ is a $(N \times 2)$-design matrix with the first column consisting of ones and the second column denoting the time points, 1 is an $(N-2)$-dim. vector of ones, and $\otimes$ denotes the Kronecker product.

Proof For the ease of notation let $\mathbf{y}_t = (\tilde{y}_{t-m}, \ldots, \tilde{y}_{t+m})'$ be the observations and $\mathbf{X}_t$ be the $(n \times 2)$-design matrix used for the regression fit in the time window centered at time point t. Further let z_j, $1 = 1, \ldots, N$, be the jth component of $\alpha\mathbf{y} + \mathbf{X}\boldsymbol{\theta}$, and $\tilde{z}_j$ be the corresponding value (observed or replaced) with which we work in the algorithm, and define $\mathbf{z_t} = (\tilde{z}_{t-m}, \ldots, \tilde{z}_{t+m})'$.

In every step, if the possibly replaced observation still holds $\mathbf{z}_t = \alpha\mathbf{y}_t + \mathbf{X}_t\theta$ then we have $T(\mathbf{z}_t) = \alpha T(\mathbf{y}_t) + \boldsymbol{\theta}$ as the regression functional is regression- and affine-equivariant. Hence, $\mathbf{z}_t - \mathbf{X}_t T(\mathbf{z}_t) = \alpha[\mathbf{Y}_t - \mathbf{X}_t T(\mathbf{y}_t)]$, *i.e.* the residuals obtained from applying T to $\mathbf{z}_t$ are α-times those for $\mathbf{y}_t$, and thus we have $\tilde{\sigma}(\mathbf{z}_t) = |\alpha|\tilde{\sigma}(\mathbf{y}_t)$ because of the affine-equivariance of $\tilde{\sigma}$.

Using these remarks, we verify the proposition by induction on t. For the initial fit in the first time window, $t = m+1$, we find that each observation in $\mathbf{z}_t$ is detected as positive (negative) outlier if this is true for the corresponding observation in $\mathbf{y}_t$. An outlier at time point i is replaced by

$$\begin{aligned}
\tilde{z}_i &= \mathbf{x}_i' T(\mathbf{z}t) + d_1\tilde{\sigma}(\mathbf{z}t)\mathrm{sgn}(z_i - \mathbf{x}_i' T(\mathbf{z}t)) \\
&= \mathbf{x}_i'(\alpha T(\mathbf{y}_t) + \boldsymbol{\theta}) + d_1|\alpha|\tilde{\sigma}(\mathbf{y}_t)\mathrm{sgn}(\alpha[y_i - \mathbf{x}_i' T(\mathbf{y}_t)]) \\
&= \alpha\mathbf{x}_i' T(\mathbf{y}_t) + \mathbf{x}_i'\boldsymbol{\theta} + d_1|\alpha|\tilde{\sigma}(\mathbf{y}_t)\mathrm{sgn}(\alpha)\mathrm{sgn}(y_i - \mathbf{x}_i' T(\mathbf{y}_t)) = \alpha\tilde{y}_i + \mathbf{x}_i'\boldsymbol{\theta},
\end{aligned}$$

with $\mathbf{x}_i'$ denoting the ith row of $\mathbf{X}$. Hence, after the replacements we find the same relation for the observations as before. As we detect exactly the same time points as positive (negative) outliers for $\mathbf{z}_t$ as for $\mathbf{y}_t$, we need to do some resetting (steps 4 and 5) in exactly the same situations and again find the same basic relation afterwards. Applying the preliminary remarks to the observations in the time window with replacement again we find that the level, the slope and the scale approximate at time point $t = m + 1$ obtained from $\mathbf{z}$ are $\alpha\tilde{\mu}_t + \theta_1$, $\alpha\tilde{\beta}_t + \theta_2$ and $|\alpha|\tilde{\sigma}_t$ respectively, where $\tilde{\mu}_t$, $\tilde{\beta}_t$ and $\tilde{\sigma}_t$ are the corresponding approximates for $\mathbf{y}$.

Now consider $t > m + 1$ and assume that the assertion is proved up to $t - 1$. When moving from $t - 1$ to t we first check the residual at the new time point $t + m$. Using the preliminary remarks and the induction assumption we see that the prediction residual for $z_{t+m} = \alpha y_{t+m} + \mathbf{x}_{t+m}'\boldsymbol{\theta}$ is

$$z_{t+m} - \mathbf{x}_{t+m}'T(\mathbf{z}_{t-1}) = \alpha y_{t+m} + \mathbf{x}_{t+m}'\boldsymbol{\theta} - \mathbf{x}_{t+m}'[\alpha T(\mathbf{y}_{t-1}) + \boldsymbol{\theta}] = \alpha[y_{t+m} - \mathbf{x}_{t+m}'T(\mathbf{y}_{t-1})],$$

i.e. α-times that for y_{t+m}, and we further have $\tilde{\sigma}(\mathbf{z}_{t-1}) = |\alpha|\tilde{\sigma}(\mathbf{y}_{t-1})$. Hence, z_{t+m} is regarded as positive (negative) outlier iff this is true for y_{t+m}. For the replacement we then have

$$\begin{aligned}\alpha\tilde{z}_{t+m} &= \mathbf{x}_{t+m}'(\alpha T(\mathbf{y}_{t-1}) + \boldsymbol{\theta}) + d_1|\alpha|\tilde{\sigma}(\mathbf{y}_{t-1})\mathrm{sgn}(\alpha[\tilde{\mathbf{y}}_{t+m} - x_{t+m}'T(y_{t-1})])\\ &= \alpha\tilde{y}_{t+m} + \mathbf{x}_{t+m}'\boldsymbol{\theta},\end{aligned}$$

i.e. in any case we find the same basic relation as before. Moreover, we find positive and negative outliers in the new, updated time window at the same time points when observing $\mathbf{z}_t$ as when observing $\mathbf{y}_t$. Thus, we can reason in the same way as for $t = m + 1$ to see that the estimates obtained from $\mathbf{z}$ at time point t are $\alpha\tilde{\mu}_t + \theta_1$, $\alpha\tilde{\beta}_t + \theta_2$ and $|\alpha|\tilde{\sigma}_t$. ■

2.6 Tracking Shifts

If applied without modifications, T_{RM} tends to smooth level shifts as it becomes increasingly biased when more than, say, 30% of the observations in the window are affected by the shift, while T_{LMS} resists a shift much better [1]. A rule for shift detection [20] may be added to the T_{RM}-based procedures for better tracking of sudden shifts.

The medical rule of thumb stated in Section 2.1 suggests to use rules based on runs of outliers for shift detection [21]. However, this rule of thumb does not apply in any case, and run rules are not 'robust' as they may fail because of single outliers immediately after a shift. Therefore, we base our rule on the residuals $r_1, \ldots, r_m$ to the right of the center of the current time window. We decide that a positive level shift may have happened if

$$\sum_{j=1}^{m} I_{\{r_j > d_2\tilde{\sigma}\}} > \sum_{j=1}^{m} I_{\{r_j \le d_2\tilde{\sigma}\}},$$

i.e. if more than half of these residuals are large positive, and we use an analogous rule for negative level shifts. The constant d_2 has to be chosen as a relevant threshold. We use $d_2 = 2$ in the following as small shifts are usually irrelevant and influence the subsequent outcomes of the filtering procedure less than large shifts. Using such a rule, the breakdown point of the regression functional drops down to $\lfloor m/2\rfloor/n \approx 1/4$, but a shift can still be detected if $\lfloor m/2\rfloor$ of the first m observations after the shift are outlying.

For better tracking of level shifts we may add this rule to the procedure after step 5. As we need to restart the algorithm when we detect a shift, denote in case of a positive (negative) shift the smallest $j \in \{1, \ldots, m\}$ with $r_j > d_2\tilde{\sigma}$ $(r_j < -d_2\tilde{\sigma})$ by j_1. This is the time point where we assume the shift to have happened. We extrapolate the current trend up to time point $t + j_1 - 1$, set t to $t + m + 1$ and restart the algorithm with step 1 as there are at most $\lfloor m/2 \rfloor$ observations in the new window from the time before the shift. High breakdown point methods can cope with a percentage of contamination less than 25%. For approximating the signal at time points $t + j_1, \ldots, t + m$ we extrapolate the trend estimate derived at $t + m + 1$. Note that the invariance properties stated in the previous lemma still hold except for the extrapolated time points.

Using this modification, we can detect shifts separated by at least $t + m + 2 - (t + \lfloor (m + 1)/2 \rfloor) = \lfloor m/2 \rfloor + 2$ time points, and the minimal delay for shift detection is $\lfloor m/2 \rfloor + 1$. This may further guide the choice of m. Complementary run rules could be added to speed up shift detection. Note, however, that it takes some time to distinguish outlier patches from level shifts anyway.

3 SIMULATION STUDY

In the following, we compare the finite-sample performance of the various versions of the T_{RM}-based procedure to the T_{LMS} without outlier replacement. We simulate data from the model

$$Y_t = \mu + t\beta + E_t, \quad t = 1, \ldots, N,$$

with $\mu = \beta = 0$ as both T_{RM} and T_{LMS} are affine- and regression-equivariant and all scale estimators are affine-equivariant. The errors E_t are always standard normal white noise. We use a window width of $n = 31$ observations. The reported results correspond to the window $\{70, \ldots, 100\}$ centered at $t = 85$ if not stated otherwise, *i.e.* we use the first 69 time windows for burn-in. A more detailed analysis of the results can be found in the homonymous discussion paper to be found at http://www.sfb475.uni-dortmund.de/dienst/en/content/veroeff-e/veroeff-e.html.

3.1 Finite-Sample Correction Factors

To obtain finite-sample correction factors $c_{i,k}$ for the scale estimators $i = 1, \ldots, 4$ when applied to the regression residuals we generate 100,000 samples for each of sizes $k = 5, \ldots, 31$, and calculate corrections for the estimators to become unbiased, see Figure 1.

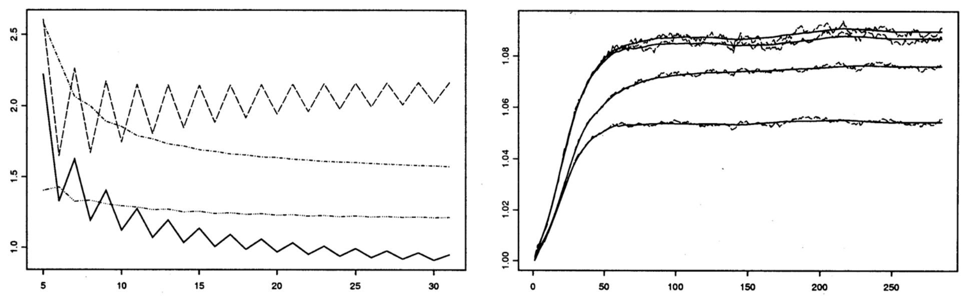

FIGURE 1 Left: finite-sample correction factors $c_{i,k}$: $\tilde{\sigma}_{\mathrm{MAD}}$ (−·−·−·), $\tilde{\sigma}_{\mathrm{LSH}}$ (———), $\tilde{\sigma}_{\mathrm{QN}}$ (**———**), $\tilde{\sigma}_{\mathrm{SN}}$ [−·−·−·]. Right: temporal correction factors $c_{i,t}^{(T)}$ for trimming, top down: $\tilde{\sigma}_{\mathrm{MAD}}$, $\tilde{\sigma}_{\mathrm{LSH}}$, $\tilde{\sigma}_{\mathrm{SN}}$, $\tilde{\sigma}_{\mathrm{QN}}$. The smooth curves are derived by local linear smoothing of the simulated wiggled curves using an adaptive bandwidth.

Since we replace extreme observations in the procedure these corrections may not be sufficient after the first time window. Therefore we generate 20,000 time series of length $N = 300$ each for deriving time (t) corrections $c_{i,t}^{(S)}$ for each scale estimator i and strategy $S \in \{T, L, M, W\}$. The total correction is $c_{i,t,k}^{(S)} = c_{i,t}^{(S)} c_{i,k}$. For trimming k is n less the number of outliers detected in the current time window, while for the other strategies we always have $k = n = 31$. $c_{i,t}^{(S)}$ is always increasing in t and stabilizes after between 30 and 70 time points, see Figure 1.

3.2 Efficiency

For calculation of the finite-sample efficiencies as measured by the MSE we generate 10,000 time series of length 150. Using the finite-sample corrections we find all methods to be unbiased. As the results stabilize soon at about $t = 30$ we concentrate on the MSE at $t = 85$.

Table I shows that the choice of the scale estimator influences the efficiency for the level and the slope only slightly. All strategies but trimming are even slightly more efficient than using

TABLE I N(0,1) Errors: Finite-Sample Efficiencies Relative to the MLE Measured by the Simulated MSE (in percent), and Percentage of Replaced Outliers.

	$\tilde{\sigma}_{MAD}$	$\tilde{\sigma}_{LSH}$	$\tilde{\sigma}_{QN}$	$\tilde{\sigma}_{SN}$
No outlier replacement				
Level	64.3	64.3	64.3	64.3
Slope	71.4	71.4	71.4	71.4
Scale	35.0	39.5	66.4	54.4
Outliers	0	0	0	0
Trimming				
Level	51.4	51.5	55.4	55.3
Slope	70.6	70.7	70.9	71.0
Scale	23.7	25.0	50.5	38.9
Outliers	2.8	2.9	1.6	1.6
Downsizing L				
Level	64.3	64.4	64.6	64.5
Slope	72.3	72.3	72.5	72.4
Scale	33.0	36.8	62.5	51.1
Outliers	2.0	1.9	1.2	1.4
Downsizing M				
Level	64.9	64.9	64.7	64.7
Slope	73.2	72.9	73.0	72.9
Scale	27.4	30.1	40.0	38.1
Outliers	8.5	8.0	6.5	7.4
Winsorization				
Level	65.0	65.0	64.8	64.8
Slope	73.2	73.2	73.1	73.2
Scale	36.7	41.7	68.7	58.7
Outliers	8.8	8.5	7.4	7.9
T_{LMS}				
Level	20.6	20.6		
Slope	21.3	21.3		
Scale	30.1	40.7		
Outliers	0	0		

no outlier replacement, and trimming is much more efficient than T_{LMS}. For the MSE of the scale approximation we always find the ordering $\tilde{\sigma}_{\text{QN}}$, $\tilde{\sigma}_{\text{SN}}$, $\tilde{\sigma}_{\text{LSH}}$, $\tilde{\sigma}_{\text{MAD}}$, which is well-known in the standard location-scale situation without outlier replacement.

3.3 Inliers

Some variables measured in intensive care may have low variability in comparison to the measurement scale. This can result in identical measurements causing scale estimators to become negatively biased and possibly even leading to zero estimates (implosion). Therefore, we investigate the effect of identical observations replacing an increasing number $0, \ldots, 15$ of observations by zero values (inliers) at time points chosen at random in the window centered at $t = 85$. Each of the 16 cases is simulated 10,000 times and the squared bias, variance and MSE are calculated.

Since the variances of all scale estimators are slightly decreasing for an increasing number of zero measurements with minor differences we restrict the comparison to the MSE, see Figure 2. Downsizing M is best for all scale estimators, while the other strategies are close to each other and better than T_{LMS}. Within the strategies, $\tilde{\sigma}_{\text{QN}}$ seems best and $\tilde{\sigma}_{\text{LSH}}$ worst.

3.4 Moderate Contamination

Next we examine the influence of a small to moderate fraction of outliers. We replace an increasing number $0, \ldots, 8$ of observations by outliers of increasing size $\omega \in \{2, 4, \ldots, 10\}$ at random time points in the window centered at $t = 85$. This corresponds to between 0% and 25.8% contaminated observations. Since outliers in the previous time windows may affect the results, we also replace the same number of observations in the preceding non-overlapping window centered at $t = 54$. Again 10,000 simulation runs are used for each of the 25 cases.

Figure 3 illustrates the results for additive outliers with random sign. We restrict to the MSE as the level and the slope approximation is unbiased here, and we only depict the results for $\tilde{\sigma}_{\text{MAD}}$ and for trimming in any combination. The T_{RM} with outlier replacement outperforms the T_{LMS} here. Trimming results in a slightly larger MSE for the level than the other strategies, but it performs best for the slope. For the level and the slope, the differences among the scale functionals are small. For the scale, downsizing L based on $\tilde{\sigma}_{\text{LSH}}$ or $\tilde{\sigma}_{\text{MAD}}$ provides the best protection.

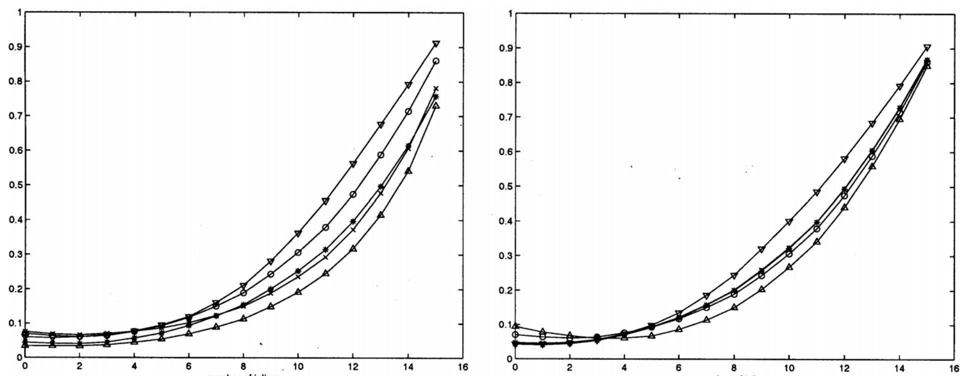

FIGURE 2 Increasing number of inliers, MSE for the scale: Left: trimming with $\tilde{\sigma}_{\text{MAD}}$ ($\times$), $\tilde{\sigma}_{\text{LSH}}$ ($\circ$), $\tilde{\sigma}_{\text{QN}}$ (Δ), $\tilde{\sigma}_{\text{SN}}$ ($*$), and T_{LMS} and $\tilde{\sigma}_{\text{MAD}}$($\nabla$). Right: $T\tilde{\sigma}_{\text{LSH}}$ ($\circ$), $L\tilde{\sigma}_{\text{LSH}}$ ($\times$), $M\tilde{\sigma}_{\text{LSH}}$ (Δ), $W\tilde{\sigma}_{\text{LSH}}$ ($*$), and T_{LMS} and $\tilde{\sigma}_{\text{LSH}}$ (∇).

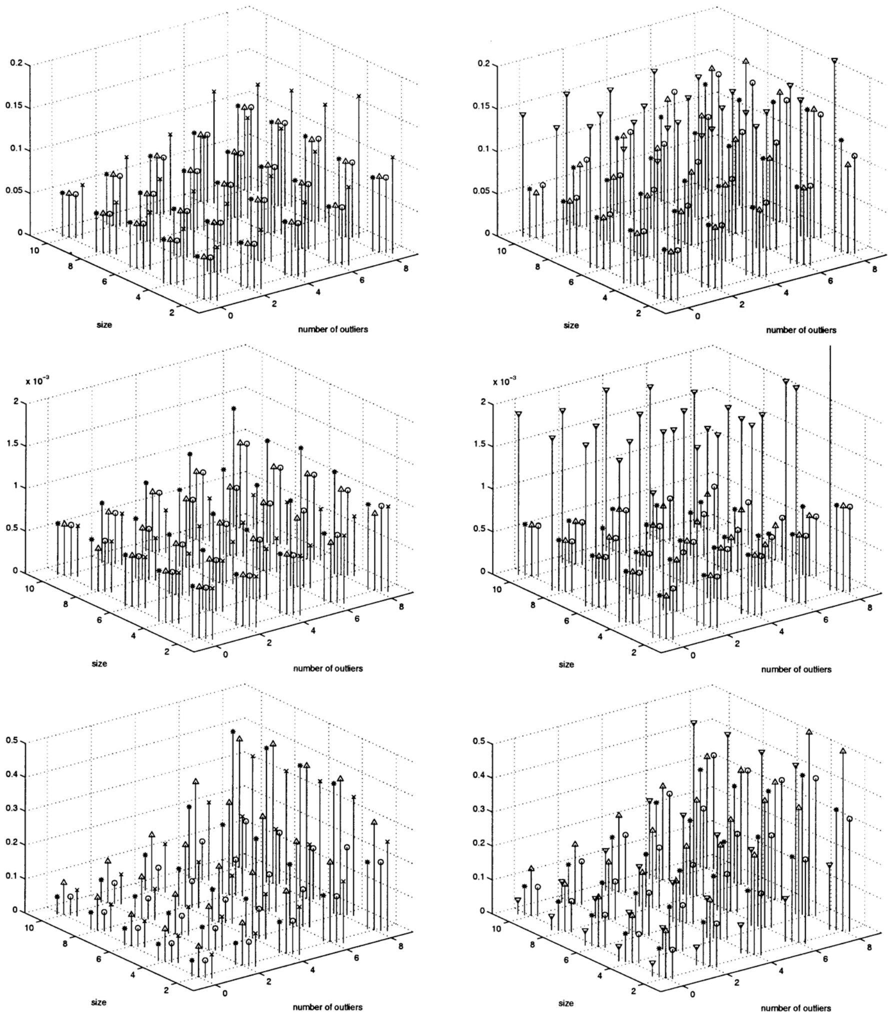

FIGURE 3 Small to moderate number of additive outliers with random sign, MSE for the level (top), for the slope (middle) and for the scale (bottom): $T\tilde{\sigma}_{\mathrm{MAD}}$ ($\times$), $L\tilde{\sigma}_{\mathrm{MAD}}$ ($\circ$), $M\tilde{\sigma}_{\mathrm{MAD}}$ (Δ), $W\tilde{\sigma}_{\mathrm{MAD}}$ ($*$) (left), and $T\tilde{\sigma}_{\mathrm{LSH}}$ ($\circ$), $T\tilde{\sigma}_{\mathrm{QN}}$ (Δ), $T\tilde{\sigma}_{\mathrm{SN}}$ ($*$), T_{LMS} and $\tilde{\sigma}_{\mathrm{LSH}}$ (∇)(right).

When inserting positive additive or substitutive outliers (not depicted here), trimming gives the best results for the level and the slope, while downsizing L seems better for the scale. $\tilde{\sigma}_{\mathrm{QN}}$ and $\tilde{\sigma}_{\mathrm{LSH}}$ perform best then. The T_{LMS} shows advantages for the level in case of many large outliers. We also study the effects of additional level shifts of sizes $\nu \in \{-10, -5, -3, 3, 5, 10\}$ in the time window. According to the rule of thumb stated in Section 2.6 we generate a level shift by adding ν to the last five observations at the end of the window centered at $t = 85$. We find the results to be close to those without level shift reported before, *i.e.* the procedure obviously resists a level shift very well even if already 5 out of 31 observations in the time window are affected by the shift.

3.5 Explosion

So far, we have seen that the high breakdown point methods cope well with 25% or less outliers. However, estimators with the same breakdown point can be very differently affected by a fraction of outliers which is close to the breakdown point. Berrendero and Zamar [22] find the maximum asymptotic bias of $\tilde{\sigma}_{\mathrm{LSH}}$ for almost 50% contamination in a location-scale model to go considerably slower to infinity than that of the other scale estimators applied here.

Figure 4 depicts the results when replacing $7, \ldots, 15$ observations by positive outliers of size $\omega \in \{2, 4, \ldots, 10\}$. Trimming seems better than the other strategies, and downsizing L is the only serious competitor. Among the scale functionals, $\tilde{\sigma}_{\mathrm{QN}}$ and $\tilde{\sigma}_{\mathrm{LSH}}$ perform superior. The T_{LMS} is best if there are more than 10 large outliers. The differences are mostly due to bias.

The results for additive outliers with random sign look much better. As with a moderate number of outliers, trimming is worse than the other strategies for the level, but better for the slope and the scale, while there are no big differences among the scale estimators.

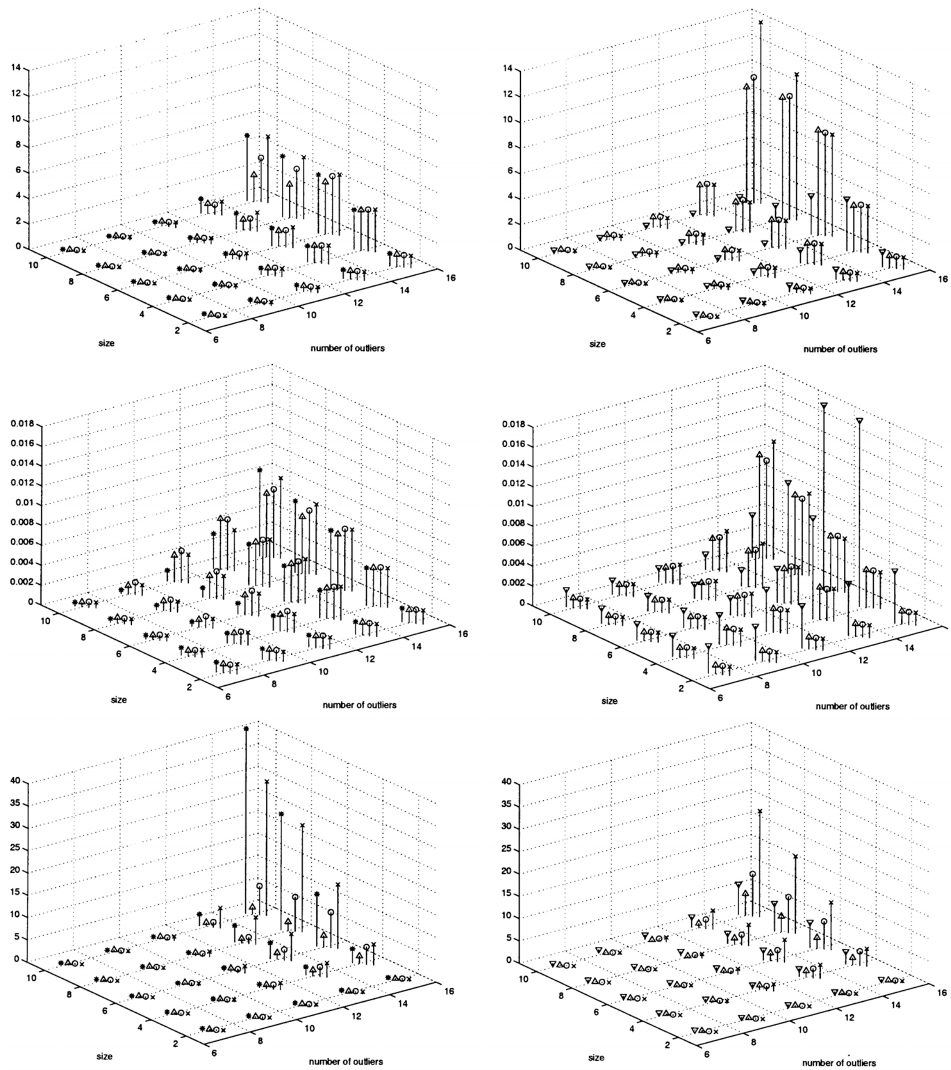

FIGURE 4 Large number of positive additive outliers, MSE for the level (top), for the slope (middle) and for the scale (bottom): $T\tilde{\sigma}_{\mathrm{MAD}}$ ($\times$), $T\tilde{\sigma}_{\mathrm{LSH}}$ ($\circ$), $T\tilde{\sigma}_{\mathrm{QN}}$ (Δ), $T\tilde{\sigma}_{\mathrm{SN}}$ ($*$)(left), and $L\tilde{\sigma}_{\mathrm{MAD}}$ $\times$, $L\tilde{\sigma}_{\mathrm{LSH}}$ ($\circ$), $L\tilde{\sigma}_{\mathrm{QN}}$ (Δ), $T\tilde{\sigma}_{\mathrm{LMS}}$ and $\tilde{\sigma}_{\mathrm{LSH}}$ (∇)(right).

4 APPLICATION

In the following, we apply the methods to simulated and real time series for further comparison. In view of the previous results, we restrict ourselves to trimming, downsizing L and T_{LMS}.

4.1 Time Series with Shifts

First we discuss a simulated time series of length 500 comparing the outcomes of the filtering procedures to the 'true' values, see Figure 5. The time series is generated from Gaussian white noise with unit variance, and a deterministic trend as well as two level shifts of sizes 4 and 6 have been inserted. Moreover, patches of $4(2\times)$, $3(4\times)$, $2(9\times)$ and $1(12\times)$ subsequent observations have been replaced by positive additive outliers of size 6 summing up to 10% outliers.

Every version of the procedure detects the level shifts and times them correctly at $t = 300$ and $t = 400$, while T_{LMS} slightly increases before the first downward shift and then drops down to early. The slope approximation (not depicted here) based on T_{LMS} is more volatile than that based on T_{RM}, and it is much more affected by the shifts and the outlier patches. Trimming performs superior here as downsizing L with any of the scale functionals overestimates the signal right after the end of the trend period because of an outlier patch.

4.2 Real Time Series

The second example is a real physiologic time series representing heart rate (Fig. 5) analyzed in Ref. [1]. An experienced physician found a clinically relevant downward trend and some irrelevant outlier patches in this time series. Here, we compare the performance of the trimming based methods. As opposed to applying T_{RM} without replacement [1] the positive outliers do not cause T_{RM} in combination with trimming based on $\tilde{\sigma}_{\text{LSH}}$, $\tilde{\sigma}_{\text{QN}}$ or $\tilde{\sigma}_{\text{SN}}$ to overestimate the signal, while using $\tilde{\sigma}_{\text{MAD}}$ results in a spurious increase at $t = 170$. The T_{LMS} exhibits a large spike at $t = 63$ due to a special pattern in the data. The slope approximates almost constantly signal a monotonic decrease up to $t = 140$, while they vary about zero thereafter when using TRM with trimming. Only when using $\tilde{\sigma}_{\text{MAD}}$, we get a large negative slope at about $t = 180$. Again T_{RM} outperforms T_{LMS}, which is more volatile and strongly influenced by some patterns in the data.

4.3 Time Series with Non-Linear Trend

Finally, we analyze a simulated time series of length 600 with an underlying sinusoidal trend $\mu_t = 5\sin(\pi/400)1_{t\leq 400} - 5\sin(\pi/200)1_{t>400}$, that is overlayed by $N(0, 1)$ noise. Ten percent of the observations are disturbed by additive $N(0, 9)$ outliers organized in patches of $4(3\times)$, $3(6\times)$, $2(10\times)$ and $1(10\times)$ subsequent outliers. All outlier sizes are generated independently, also within the patches, for getting a very distinct scenario to those considered before. We note that in Section 3 trimming showed its main weaknesses in the case of two-sided outliers considered here.

Figure 5 depicts the results for trimming with $\tilde{\sigma}_{\text{QN}}$. The results for $\tilde{\sigma}_{\text{LSH}}$ or downsizing L are very similar. Application of T_{RM} with either of these combinations reproduces the underlying non-linear trend well, and there are only some small problems with the minimum and the maximum of the signal. T_{LMS} is again much more volatile with some superfluous bumps.

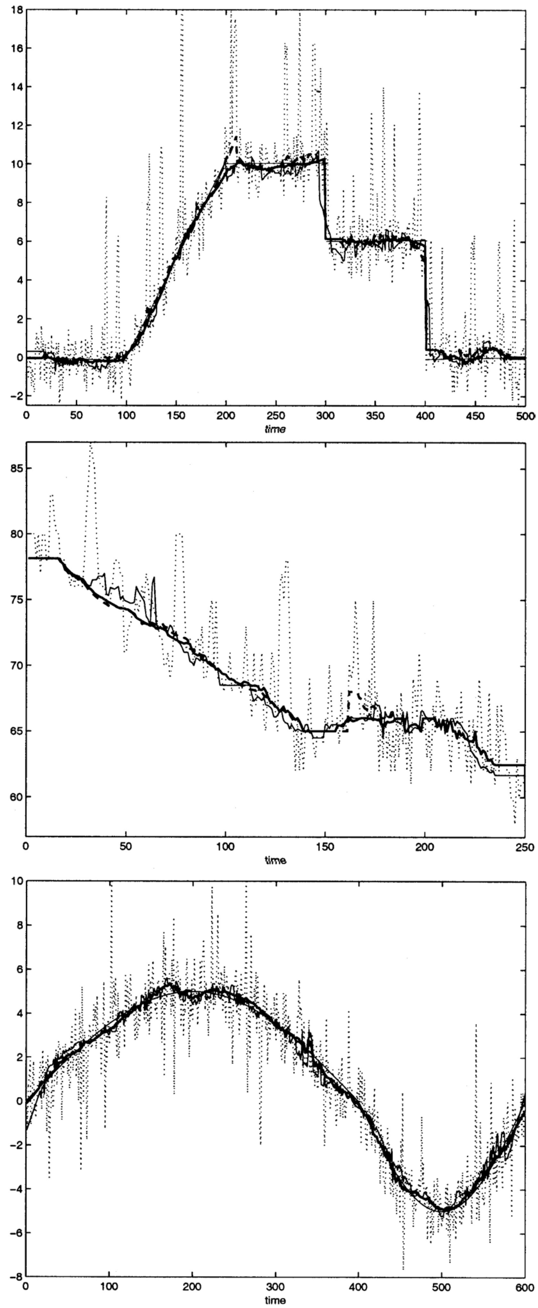

FIGURE 5 Top: Simulated time series (······), underlying level (———) and level approximates: T_{LMS}(———), $\tilde{\mu}_{\mathrm{RM}}$ with $T\tilde{\sigma}_{\mathrm{QN}}$ (**———**) and $\tilde{\mu}_{\mathrm{RM}}$ with $L\tilde{\sigma}\mathrm{QN}$ (– – – –). Center: Heart rate (······) and level approximates: T_{LMS}(———), $\tilde{\mu}_{\mathrm{RM}}$ with $T\tilde{\sigma}_{\mathrm{QN}}$ (**———**) and $\tilde{\mu}_{\mathrm{RM}}$ with $T\tilde{\sigma}_{\mathrm{MAD}}$ (– – – –). Bottom: Simulated time series (······), underlying level (———) and level approximates: T_{LMS} (———) and $\tilde{\mu}_{\mathrm{RM}}$ with $T\tilde{\sigma}_{\mathrm{QN}}$(**———**).

5 CONCLUSION

The extraction of an underlying signal from noisy data is a basic task for automatic online monitoring. The repeated median suggested in Ref. [1] can be further improved by online outlier replacement using high breakdown point scale estimators. Although downsizing can be better if there is a moderate number of positive and negative outliers, trimming seems generally superior as it helps to almost achieve the high robustness of the LMS even in extreme situations. There seem to be better choices than the classical MAD for the scale estimation. The $\tilde{\sigma}_{\mathrm{QN}}$ seems well-adapted to mixtures of shifted distributions as it shows excellent performance for outliers of similar size and when a level shift occurs. If there are many large outliers of different sizes $\tilde{\sigma}_{\mathrm{LSH}}$ provides stronger worst-case protection. Such combined procedures look preferable to the LMS because of the better performance in the case of a moderate number of outliers and the smaller computational costs. Reliable rules for shift detection can be based on the regression residuals, and trends might be detected from the sequence of slope approximates.

Many identical measurements may cause problems for automatic outlier replacement. We suggest using $\tilde{\sigma}_{\mathrm{QN}}$ then and some modifications which work very well at least in the case of normal errors. An automatic procedure can still fail when all measurements in the time window are identical. Increasing the window width may sometimes help but is not always possible as it increases the delay. In some applications, we have an idea about a minimal variability in the data or minimal relevant outlier and shift sizes, which can be incorporated in the algorithm. Otherwise, one can add uniform noise according to the measurement scale both to the observations and the replacements.

Acknowledgements

The author thanks Karen Schettlinger for complementary data analysis. The remarks of a referee, which were helpful to improve the presentation, and the financial support of the Deutsche Forschungsgemeinschaft (SFB 475, ‘Reduction of complexity in multivariate data structures’) are gratefully acknowledged.

References

[1] Davies, P. L., Fried, R. and Gather, U. Robust signal extraction from on-line monitoring data. *J. Statist. Plann. and Inference*, **122**, 65–78.
[2] Tukey, J. W. (1977). *Exploratory Data Analysis*. Addison-Wesley, Reading, Mass. (preliminary ed. 1971).
[3] Fried, R. and Gather, U. (2002). Fast and robust filtering of time series with trends. In: Härdle, W. and Rönz, B. (Eds.), *Proceedings in Computational Statistics COMPSTAT 2002*. Physica-Verlag, Heidelberg, pp. 367–372.
[4] Davies, P. L. (1993). Aspects of robust linear regression. *Ann. Statist.*, **21**, 1843–1899.
[5] Rousseeuw, P. J. and Croux, C. (1993). Alternatives to the median absolute deviation. *J. Amer. Statist. Assoc.*, **88**, 1273–1283.
[6] Fan, J., Hu, T.-Ch. and Truong, Y. K. (1994). Robust non-parametric function estimation. *Scand. J. Statist.*, **21**, 433–446.
[7] Rue, H., Chu, Ch.-K., Godtliebsen, F. and Marron, J. S. (2002). M-Smoother with local linear fit. *J. Nonparametr. Statist.*, **14**, 155–168.
[8] Gather, U., Fried, R., Lanius, V. and Imhoff, M. (2001). Online-monitoring of high-dimensional physiological time series – a case-study. *Estadistica*, **53**, 259–298.
[9] Smith, A. F. M. and West, M. (1983). Monitoring renal transplants: An application of the multiprocess Kalman filter. *Biometrics*, **39**, 867–878.
[10] Hampel, F. R. (1975). Beyond location parameters: Robust concepts and methods. *Bulletin of the Int. Statist. Inst.*, **46**, 375–382.
[11] Rousseeuw, P. J. (1984). Least median of squares regression. *J. Amer. Statist. Assoc.*, **79**, 871–880.
[12] Siegel, A. F. (1982). Robust regression using repeated medians. *Biometrika*, **68**, 242–244.
[13] Bernholt, T. and Fried, R. Computing the update of the repeated median regression line in linear time. *Information Processing Letters*, **88**, 111–117.
[14] Hettmansperger, T. P. and Sheather, S. J. (1992). A cautionary note on the method of least median squares. *Amer. Statist.*, **46**, 79–83.
[15] Gather, U. and Fried, R. Robust estimation of scale for local linear temporal trends. In: Stulajter, F. (Ed.), *Proceedings of the Fourth International Conference on Mathematical Statistics PROBASTAT 2002*, Tatra Mountain Mathematical Publications, **26**, 87–101.
[16] Grübel, R. (1988). The length of the shorth. *Ann. Statist.*, **16**, 619–628.

[17] Rousseeuw, P. J. and Leroy, A. M. (1988). A robust scale estimator based on the shortest half. *Statistica Neerlandica*, **42**, 103–116.
[18] Croux, C. and Rousseeuw, P. J. (1992). Time-efficient algorithms for two highly robust estimators of scale. In: Dodge, Y. and Whittaker, J. (Eds.), *Computational Statistics*, Vol. 1. Physica-Verlag, Heidelberg, pp. 411–428.
[19] Martin, R. D. and Zamar, R. H. (1993). Bias robust estimation of scale. *Ann. Statist.*, **21**, 991–1017.
[20] Basseville, M. and Nikiforov, I. V. (1993). *Detection of Abrupt Changes – Theory and Application*. Prentice-Hall, Englewood Cliffs, NJ.
[21] Imhoff, M., Bauer, M., Gather, U. and Löhlein, D. (1998). Statistical pattern detection in univariate time series of intensive care on-line monitoring data. *Intensive Care Medicine*, **24**, 1305–1314.
[22] Berrendero, J. R. and Zamar, R. H. (1999). Global robustness of location and dispersion estimates. *Statist. Probab. Lett.*, **44**, 63–72.