ROBUST FILTERING OF TIME SERIES WITH TRENDS

ROLAND FRIED

Department of Statistics, University of Dortmund, Germany

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 procedure 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 [5]. Median filtering is often applied to smooth noisy time series [18], 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 about the median (MAD) 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 looses a lot of its robustness [6].

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]. Important 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 moving window techniques with a fixed window width regulated by the requirement of working online. Instead of approximating a local level by the median we fit a linear trend to the data in each time window using regression functionals with high breakdown point like the least median of squares and the repeated median. For further improving the robustness of the procedure we test and compare 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 [15]. Based on an estimate of the local variability we can detect and replace outliers online before they influence the data analysis. We compare standard strategies like trimming, winsorization and variations of these for this purpose. In order to find out the strengths and the weaknesses of the various possible combinations of these methods we perform an extensive simulation study and apply them to some time series.

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 $y_1, \ldots, y_N$ be real valued data measured at time points $t = 1, \ldots, N$, with y_t being a realization of a random variable Y_t with median μ_t and variance σ_t^2. 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]. It seems reasonable to approximate the signal within a moving time window by a straight line choosing a small to moderate window width n. The time delay admissible in the respective application further restricts the choice of n, while n should be chosen not very small adhering to these restrictions to reduce the variance and the effects of outlier patches. For our clinical application we use time windows of length $n = 2m + 1 = 31$ observations.

2.1 Robust Trend Approximation

In the following we approximate the signal μ_{t+i}, $i = -m, \ldots, m$, in the current time window centered at time point $t \geq m + 1$ by a straight line,

$$Y_{t+i} = \mu_t + i\beta_t + E_{t,i}, \quad i = -m, \ldots, m, \tag{1}$$

where μ_t is the level and β_t is the slope in the time window, while $E_{t,i}$ is an error term.

A comparative study [5] investigates the finite-sample properties of the least median of squares functional T_{LMS} [10], [14] and of the repeated median functional T_{RM} [17]. For regression against time with data measured at equally spaced time points $t + i$, $i = -m, \ldots, m$, these read

$$T_{LMS} = \text{argmin}\{(\mu, \beta) : \text{Median}(y_{t+i} - \mu - i\beta)^2\}, \tag{2}$$

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

$$\begin{aligned} \tilde{\beta}_t &= med_i \left(med_{j \neq i} \frac{y_{t+i} - y_{t+j}}{i - j} \right) \\ \tilde{\mu}_t &= med_i \left(y_{t+i} - i \cdot \tilde{\beta}_t \right) \end{aligned} \tag{3}$$

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 $\lfloor n/2 \rfloor / n$. Some important advantages of the T_{RM} are its smaller computation time [1], its smaller variance and MSE in case of a small to moderate number of outliers [5] and the instability of the T_{LMS} for small changes of the data [11]. Unlike the T_{LMS} the T_{RM} is Lipschitz-continuous in case of a fixed design and hence small changes in the data do not cause large changes of the results. On the other hand, the T_{LMS} has a much smaller bias and MSE than the 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 the T_{RM} shows the intuitive behavior that large outliers have a more serious effect. Therefore, replacing detected outliers may well improve the performance of the T_{RM}, while this is not necessarily true for the T_{LMS}.

For all these reasons, we use the T_{LMS} merely as a benchmark for global robustness and try to improve the T_{RM} to become similarly robust as the T_{LMS} against many large outliers. An obvious way for doing this is to replace outliers online based on an approximation of the local variability.

2.2 Scale Approximation

Assuming that the noise variance is locally almost constant, $\sigma_{t+i} \approx \sigma_t, i = -m, \ldots, m$, we can approximate σ_t applying a scale estimator to the residuals in the current time window. Let $r_i = y_{t+i} - \tilde{\mu}_t - \tilde{\beta}_t i$, $i = -m, \ldots, m$, be these residuals, where $(\tilde{\mu}_t, \tilde{\beta}_t)$ is an estimate of location and slope.

The classical robust scale estimator is the MAD

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

where $c_{1,n}$ is a finite-sample correction factor depending on the window width $n = 2m + 1$. Many other robust scale estimators have been suggested. In [7] the finite-sample properties of robust scale estimators which can be calculated in $O(n \log n)$ time are inspected in the regression setting. It turns out that the length of the shortest half [9], [16]

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

where $r_{(1)}, \ldots, r_{(n)}$ are the ordered residuals, and Rousseeuw and Croux's [15] suggestion

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

are particularly interesting. An algorithm for computation of $\tilde{\sigma}_{QN}$ in $O(n \log n)$ time is presented in [3]. Moreover, we consider the nested scale statistic

$$\tilde{\sigma}_{SN} = c_{4,n} \cdot med_i \, med_{j \neq i} |r_i - r_j| \ ,$$

which has also been proposed in [15]. The $\tilde{\sigma}_{LSH}$ shows extremely good resistance against a large percentage of outliers (see also [13]). On the other hand, $\tilde{\sigma}_{QN}$ and $\tilde{\sigma}_{MAD}$ perform better for inliers, e.g. for identical measurements due to a small variability relatively to the measurement scale. The $\tilde{\sigma}_{QN}$ works very well also in case of a level shift, where we sample from a mixture distribution. In conclusion, the finite-sample efficiency of $\tilde{\sigma}_{QN}$, $\tilde{\sigma}_{SN}$ and $\tilde{\sigma}_{LSH}$ (in this ordering) is higher than that of other explicit high breakdown point scale estimators.

2.3 Outlier Detection

Applying any of the previous scale estimators we can check whether the incoming observation y_{t+m+1} is an outlier by comparing the residual $r_{m+1} = y_{t+m+1} - \tilde{\mu}_t - \tilde{\beta}_t(m+1)$ for the extrapolation of the next time point to the estimate $\tilde{\sigma}_t$ of the current standard deviation σ_t. A general strategy for online cleaning of the data is to replace y_{t+m+1} by

$$\tilde{y}_{t+m+1} = \tilde{\mu}_t + \tilde{\beta}_t(m+1) + d_1 sgn(r_{m+1})\tilde{\sigma}_t \quad \text{if} \quad |r_{m+1}| > d_0 \tilde{\sigma}_t \tag{4}$$

where $0 \leq d_1 \leq d_0$ are a-priori specified 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. A constant $d_1 > 0$ is reasonable if we do not regard outliers as measurement artifacts but as disturbed values. Sometimes the sign of an outlier may be informative although its absolute value is too large. A choice $d_1 = d_0$ means that we winsorize the residuals. Intuitively, in the short run we expect $d_0 > d_1 = 0$ to result in more stable results w.r.t. the approximation of the level and the slope, and $d_0 = d_1 > 0$ to result in more stable results w.r.t. the scale. In the long run this is not as clear as these goals interact. A choice $d_0 > d_1 > 0$ might thus mean a compromise.

We will treat two outlier generating mechanisms which correspond to outliers that are completely non-informative (substitutive outliers) and somewhat informative (additive outliers) later on. In case of a substitutive outlier, instead of the value y_{t+i} of the underlying process we observe a value ω_O which does not depend on y_{t+i}. Measurement artifacts can typically be described by substitutive outliers. On the other hand, an additive outlier is generated by adding ω_O to y_{t+i}. Thus, an additive outlier describes a kind of shock. An additive outlier is more likely to be detected if ω_O and the error at that time point have the same sign, and vice versa the sign of an additive outlier provides some information on the undisturbed observation.

Finding an overall optimal choice of d_0 and d_1 does not seem possible since we do not want to use large sample asymptotics for a moderate window width, and since we look at several criteria (approximation of μ_t, β_t, σ_t) and outlier generating mechanisms. Therefore, we restrict to the following heuristic choices:

T $\quad d_0 = 3$, $d_1 = 0$ (trimming)

L $\quad d_0 = 3$, $d_1 = 1$ (downsizing large values)

M $\quad d_0 = 2, d_1 = 1$ (downsizing moderate values)

W $\quad d_0 = 2, d_1 = 2$ (winsorization)

Trimming is commonly applied and means to treat outliers as non-informative. We use a standard 3σ rule for outlier detection and simple predictions then. A preliminary study showed that for the scale approximation detected outliers should be treated as missing values and the finite-sample correction factor be adjusted for the reduced number of observations then. Otherwise we possibly underestimate the scale largely. In the other strategies, we use the adjusted time series for the approximations without correcting the sample size. In order to simplify notation we add a letter to the scale functional in the following to denote the outlier replacement strategy. E.g., $T\tilde{\sigma}_{MAD}$ stands for trimming based on the MAD, while $L\tilde{\sigma}_{LSH}$ is downsizing large values using the LSH.

Outlier detection cannot be performed online in the first time window. Instead, we approximate the level, the slope and the scale for this initial period and check the observations retrospectively for outlyingness using analogous rules as stated above. If we find outlying observations we reanalyze the first time window with detected outliers being replaced. Although outward procedures, where the initial estimates are calculated from all observations, are known to be prone to masking effects, we nevertheless prefer them to inward procedures since they are easy to implement and computationally fast, and since we use functionals with high breakdown points.

2.4 The Procedure

Now we formulate a basic algorithm based on the components mentioned above. 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_t, $t \in \mathbb{N}$, observed subsequently in time. We set $t = m + 1$.

1. Set $\tilde{y}_{t+j} = y_{t+j}$ and $out(t + j) = 0$, $j = -m, \ldots, m$.
2. Estimate $\tilde{\mu}_t$, $\tilde{\beta}_t$ and $\tilde{\sigma}_t$ from $\tilde{y}_{t+j}$, $j = -m, \ldots, m$, using T_{RM} and $\tilde{\sigma}$.
3. 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 $\tilde{\mu}_t + j\tilde{\beta}_t + sgn(r_j)d_1\tilde{\sigma}_t$ and set $out(t + j) = sgn(r_j)$ for these j.
4. If $\#\{j = -m, \ldots, m : out(t + j) = 1\} > m$ re-replace these $\tilde{y}_{t+j}$ and $out(t + j)$ by y_{t+j} and 0 respectively. Act in the same way if $\#\{j = -m, \ldots, m : out(t + j) = -1\} > m$.

5. If $\#\{j = -m, \ldots, m : \ |out(t+j)| = 0\} < \max\{\lfloor m/3 \rfloor, 5\}$ reset all $\tilde{y}_{t+j}$ to y_{t+j} and $out(t+j)$ to zero, $j = -m, \ldots, m,$.

6. Estimate $\tilde{\mu}_t$, $\tilde{\beta}_t$ and $\tilde{\sigma}_t$ from $\tilde{y}_{t+j}$, $j = -m, \ldots, m$, using T_{RM} and $\tilde{\sigma}$.

7. Set $\tilde{y}_{t+m+1}$ to y_{t+m+1} if $|r_{m+1}| = |y_{t+m+1} - \tilde{\mu}_t - \tilde{\beta}_t(m+1)| < d_0\tilde{\sigma}_t$ and to $\tilde{\mu}_t + \tilde{\beta}_t(m+1) + sgn(r_{m+1})d_1\tilde{\sigma}_t$ otherwise. Set $out(t+m+1) = 0$ in the former and $out(t+m+1) = sgn(r_{m+1})$ in the latter case.

8. Set t to $t+1$ and go to 4.

Steps 4 and 5 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 may approach zero because of a temporary bad regression fit or strong underestimation of the variance. Both these reasons may result in many of the incoming observations being regarded as outliers and may cause a vicious circle when replacing them by extrapolations. We found steps 4 and 5 to be very helpful to overcome these problems: If more than half of the observations within the current time window have been regarded as positive (negative) outliers and replaced by predictions, the regression line might underestimate (overestimate) the true levels. Using the original observations which have been predicted to small (large) while still using the replacements for the detected negative (positive) outliers might then improve the results. If very few observations within the current time window have not been replaced by extrapolations at all we should use all the original observations for the regression fit since possibly the variability is strongly underestimated.

In order to simplify the simulation study in Section 3 we prove that the results do not depend on an underlying constant trend. The following lemma states some invariance properties of a filtering procedure as described above. For its formulation we use the operators $f : \mathbb{R}^N \mapsto \mathbb{R}^{2N-4m}$, $(y_1, \ldots, y_N)' \mapsto (\tilde{\mu}_{m+1}, \tilde{\beta}_{m+1}, \ldots, \tilde{\mu}_{N-m}, \tilde{\beta}_{N-m})'$ and $\hat{\sigma} : \mathbb{R}^N \mapsto \mathbb{R}^{N-2m}$, $(y_1, \ldots, y_N)' \mapsto (\tilde{\sigma}_{m+1}, \ldots, \tilde{\sigma}_{N-m})$ which map a time series to its decomposition into local level and slope and into 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:*

$$\begin{aligned} f\left(\alpha \mathbf{y} + \mathbf{X}\boldsymbol{\theta}\right) &= \alpha f(\mathbf{y}) + \mathbf{1} \otimes \boldsymbol{\theta}, \\ \hat{\sigma}\left(\alpha \mathbf{y} + \mathbf{X}\boldsymbol{\theta}\right) &= |\alpha| \hat{\sigma}(\mathbf{y}), \end{aligned}$$

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, $\mathbf{1}$ 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 for the possibly replaced observations still holds $\mathbf{z}_t = \alpha\mathbf{y}_t + \mathbf{X}_t\boldsymbol{\theta}$ 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 the 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

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

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 for all time points 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 again as well as 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

$$\alpha \tilde{z}_{t+m} \;=\; \mathbf{x}_{t+m}' (\alpha T(\mathbf{y}_{t-1}) + \boldsymbol{\theta}) + d_1 |\alpha| \tilde{\sigma}(\mathbf{y}_{t-1}) sgn \left(\alpha [\tilde{y}_{t+m} - \mathbf{x}_{t+m}' T(\mathbf{y}_{t-1})] \right) = \alpha \tilde{y}_{t+m} + \mathbf{x}_{t+m}' \boldsymbol{\theta},$$

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 $|\alpha|\tilde{\sigma}_t$. □

3 SIMULATION STUDY

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

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

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 Gaussian white noise with zero mean and unit variance. We use a window width of $n = 31$ observations.

3.1 Finite-sample Correction Factors

In order 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 100000 samples for each of sizes $k = 5, \ldots, 31$, and calculate the corrections such that the estimators become unbiased, see Figure 1. Obviously, the corrections for $\tilde{\sigma}_{LSH}$ and $\tilde{\sigma}_{QN}$ strongly depend on whether the sample size is odd or even.

Since we replace extreme observations in the course of the procedure these corrections may not be sufficient after the first time window. Therefore we analyse the temporal behaviour of all combinations of scale estimators and outlier replacement strategies generating 20000 time series of length $N = 300$ each. From this we derive temporal corrections $c_{i,t}^{(S)}$ for each time point t, scale

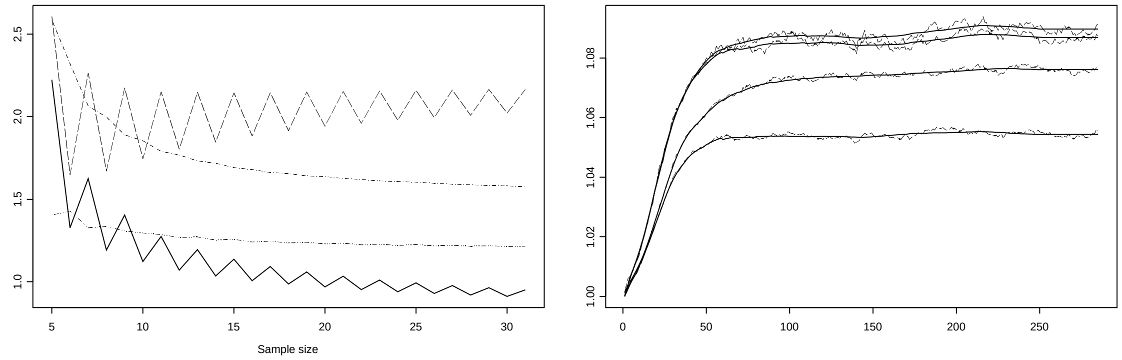

Figure 1: Left: Finite-sample correction factors $c_{i,k}$: $\tilde{\sigma}_{MAD}$ dash-dot, $\tilde{\sigma}_{LSH}$ solid, $\tilde{\sigma}_{QN}$ dashed, $\tilde{\sigma}_{SN}$ dash-dot-dot-dot. Right: Temporal correction factors $c_{i,t}^{(T)}$ for trimming, top down: $\tilde{\sigma}_{MAD}$, $\tilde{\sigma}_{LSH}$, $\tilde{\sigma}_{SN}$, $\tilde{\sigma}_{QN}$. The smooth curves are derived by local linear smoothing of the simulated wiggled curves using an adaptive bandwidth.

estimator i and strategy $S \in \{T, L, M, W\}$ such that the total correction is $c_{i,t,k}^{(S)} = c_{i,t}^{(S)} \cdot c_{i,k}$. For trimming k is n less the number of outliers detected in the current time window, while for the other strategies we always have $k = n = 31$. Figure 1 also exemplifies the temporal corrections $c_{i,t}^{(T)}$ for trimming. All curves are increasing and stabilize after between 30 and 70 time points. For reducing the variability we smooth the curves using a locally linear fit with an adaptive bandwidth.

In the following we report all results for the time window $\{70, \ldots, 100\}$ centered at $t = 85$ if not stated otherwise, i.e. we use the first 69 time windows for burn-in.

3.2 Efficiency

For calculation of the finite-sample efficiencies as measured by the mean square error MSE we generate 10000 time series of length 150 and compare the various methods for each time window. Using the finite-sample corrections derived above we find all methods to be unbiased. As the results stabilize rather soon after about 30 time points we concentrate on the MSE in the time window centered at $t = 85$, cf. Table 1.

Table 1: N(0,1) errors: Finite-sample efficiencies relatively to least squares measured by the simulated MSE (in percent), and percentage of replaced outliers.

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

Obviously, the choice of the scale estimator influences the efficiency for the level and the slope only slightly. Trimming is less efficient than the other strategies, which are even slightly more efficient than using no outlier replacement, but it is still much more efficient than the T_{LMS}. W.r.t. the scale approximation, winsorization is more efficient than using no outlier replacement, while downsizing L is slightly worse and the others are considerably worse than the latter. Furthermore, the efficiencies of $\tilde{\sigma}_{LSH}$ and $\tilde{\sigma}_{MAD}$ do not depend a lot on whether we apply them to the residuals obtained from T_{LMS} or from T_{RM}. For the scale approximation we always find the ordering $\tilde{\sigma}_{QN} >> \tilde{\sigma}_{SN} >> \tilde{\sigma}_{LSH} > \tilde{\sigma}_{MAD}$, with $>$ denoting "better" in the sense of smaller MSE. This ordering 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 leads to zero estimates ("implosion"). Therefore we investigate the effect of identical observations replacing an increasing number $0, \dots, 15$ 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 10000 times and the squared bias, variance and MSE are calculated.

Since the variances of all scale estimators are slightly decreasing with increasing number of zero measurements with minor differences only 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 the T_{LMS} with either the $\tilde{\sigma}_{MAD}$ or the $\tilde{\sigma}_{LSH}$. Within the strategies, we always find $\tilde{\sigma}_{QN} > \tilde{\sigma}_{SN} \approx \tilde{\sigma}_{MAD} > \tilde{\sigma}_{LSH}$.

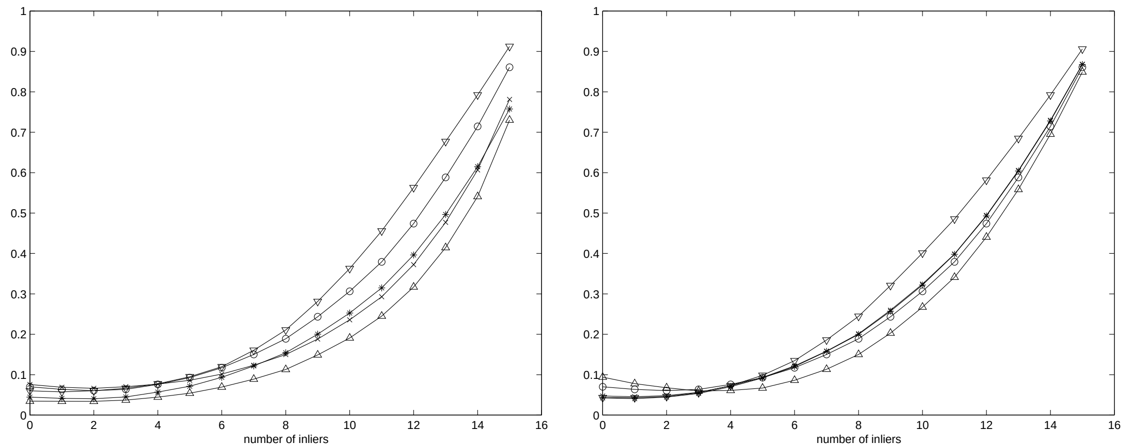

Figure 2: Increasing number of inliers, MSE for the scale: Left: Trimming with $\tilde{\sigma}_{MAD}$ x, $\tilde{\sigma}_{LSH}$ $\circ$, $\tilde{\sigma}_{QN}$ $\triangle$, $\tilde{\sigma}_{SN}$ $*$, and T_{LMS} & $\tilde{\sigma}_{MAD}$ $\bigtriangledown$. Right: $T\tilde{\sigma}_{LSH}$ $\circ$, $L\tilde{\sigma}_{LSH}$ x, $M\tilde{\sigma}_{LSH}$ $\triangle$, $W\tilde{\sigma}_{LSH}$ $*$, and T_{LMS} & $\tilde{\sigma}_{LSH}$ $\bigtriangledown$.

3.4 Small Percentage of Contamination

Next we examine the influence of a small to moderate fraction of outliers. We replace an increasing number $0, \dots, 8$ of observations by outliers of increasing size $\omega_O \in \{2, 4 \dots, 10\}$ 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 $t = 54$. Each of the 25 cases is simulated 1000 times and the squared bias, variance and MSE are calculated for the level, the slope and the scale.

Figure 3 illustrates the results for additive outliers with random sign. We restrict to the MSE as outliers with random sign do not cause the level and the slope approximation to be biased, and we only depict the results for $\tilde{\sigma}_{MAD}$ 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 $\tilde{\sigma}_{LSH}$ or $\tilde{\sigma}_{MAD}$ provides the best protection. Trimming and downsizing L based on $\tilde{\sigma}_{QN}$ or $\tilde{\sigma}_{SN}$ perform well for large outliers but have weaknesses in case of many small outliers as they become biased then, similarly as $\tilde{\sigma}_{LSH}$ applied to the T_{LMS} residuals. Closer inspection reveals that for trimming the variance increases with the number of outliers according to the reduced number of observations.

When inserting one-sided positive outliers the results are essentially the same as when we additionally insert a level shift as discussed in the next subsection.

3.5 Level Shift and Outliers

Now we additionally insert a level shift of size $\omega_S \in \{-10, -5, -3, 3, 5, 10\}$ into the time window as the occurrence of large shifts is an important problem. In intensive care, changes that last five or more minutes are often clinically relevant [12]. Therefore, we generate a level shift by adding ω_S to the last five observations at the end of the window centered at $t = 85$. Additionally, we replace an increasing number $0, 2, \ldots, 8$ of observations by outliers of increasing size $\omega_O \in \{2, 4, \ldots, 10\}$ at random time points in the window as described before. Again, we generate 1000 samples for each case and calculate the squared bias, variance and MSE.

Figure 4 depicts the results for a moderate shift of size $\omega_S = 3$ and positive additive outliers. The results for the other shift sizes are very similar, also for negative shifts. Trimming gives the best results w.r.t. the level and the slope, while downsizing L seems better than trimming for the scale with $\tilde{\sigma}_{QN}$ being best. The T_{LMS} shows better performance than the T_{RM}-based methods only in case of a substantial number of large outliers and only for the level. Closer examination shows that the advantages of trimming for the level and the slope are due to bias as the variances of all methods are rather stable. The differences w.r.t. the scale are mainly due to variance, although all methods except $T\tilde{\sigma}_{SN}$ show an increasing bias for more than five outliers.

When inserting positive substitutive outliers (not shown here), the results are similar to those for positive additive outliers. Trimming shows the best overall performance, while the T_{LMS} offers advantages in comparison to the T_{RM}-based procedures for the level in case of many large outliers. Downsizing L is again the closest competitor to trimming, particularly for the scale. The $\tilde{\sigma}_{LSH}$ and even more $\tilde{\sigma}_{QN}$ perform very well here, especially for the slope.

When inserting a level shift and replacing an increasing number of observations by additive outliers with random sign the results look very much like those reported in the previous subsection for the case without a shift.

3.6 Explosion

So far we have seen that the high breakdown point methods cope rather well with 25% or less outliers. The differences depend mainly on the strategy for outlier replacement then. 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 [2] find the maximum asymptotic bias of $\tilde{\sigma}_{LSH}$ for almost 50% contamination in a location-scale model to go considerably slower to infinity than that of the other scale functionals applied here.

Figure 5 depicts the results when replacing an increasing number $7, \ldots, 15$ of observations by positive outliers of size $\omega_O \in \{2, 4, \ldots, 10\}$. Trimming seems better than the other strategies in

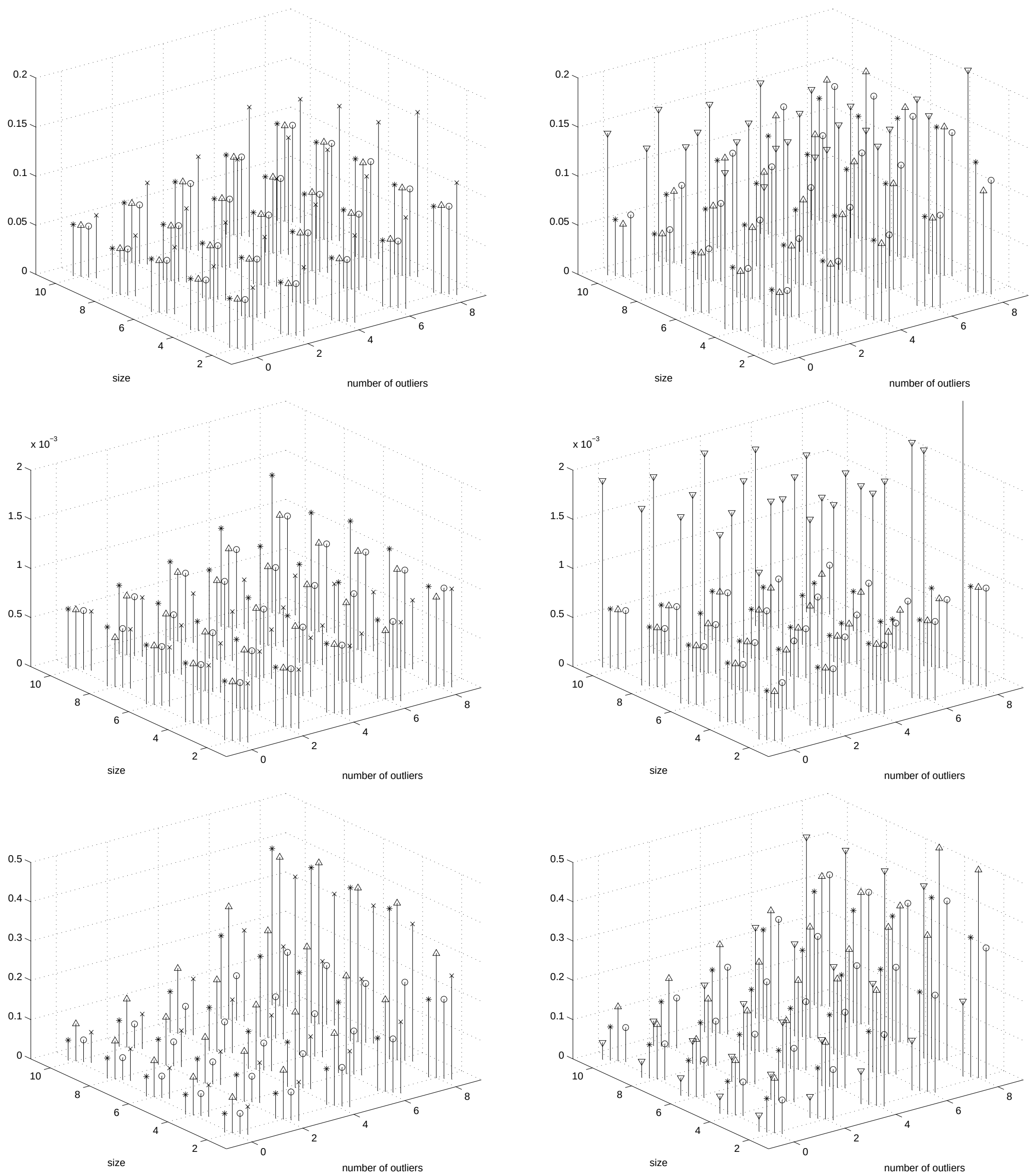

Figure 3: Small to moderate number of additive outliers with random sign, MSE for the level (top), for the slope (middle) and for the scale (bottom): $T\tilde{\sigma}_{MAD}$ x, $L\tilde{\sigma}_{MAD}$ $\circ$, $M\tilde{\sigma}_{MAD}$ $\triangle$, $W\tilde{\sigma}_{MAD}$ $*$ (left), and $T\tilde{\sigma}_{LSH}$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

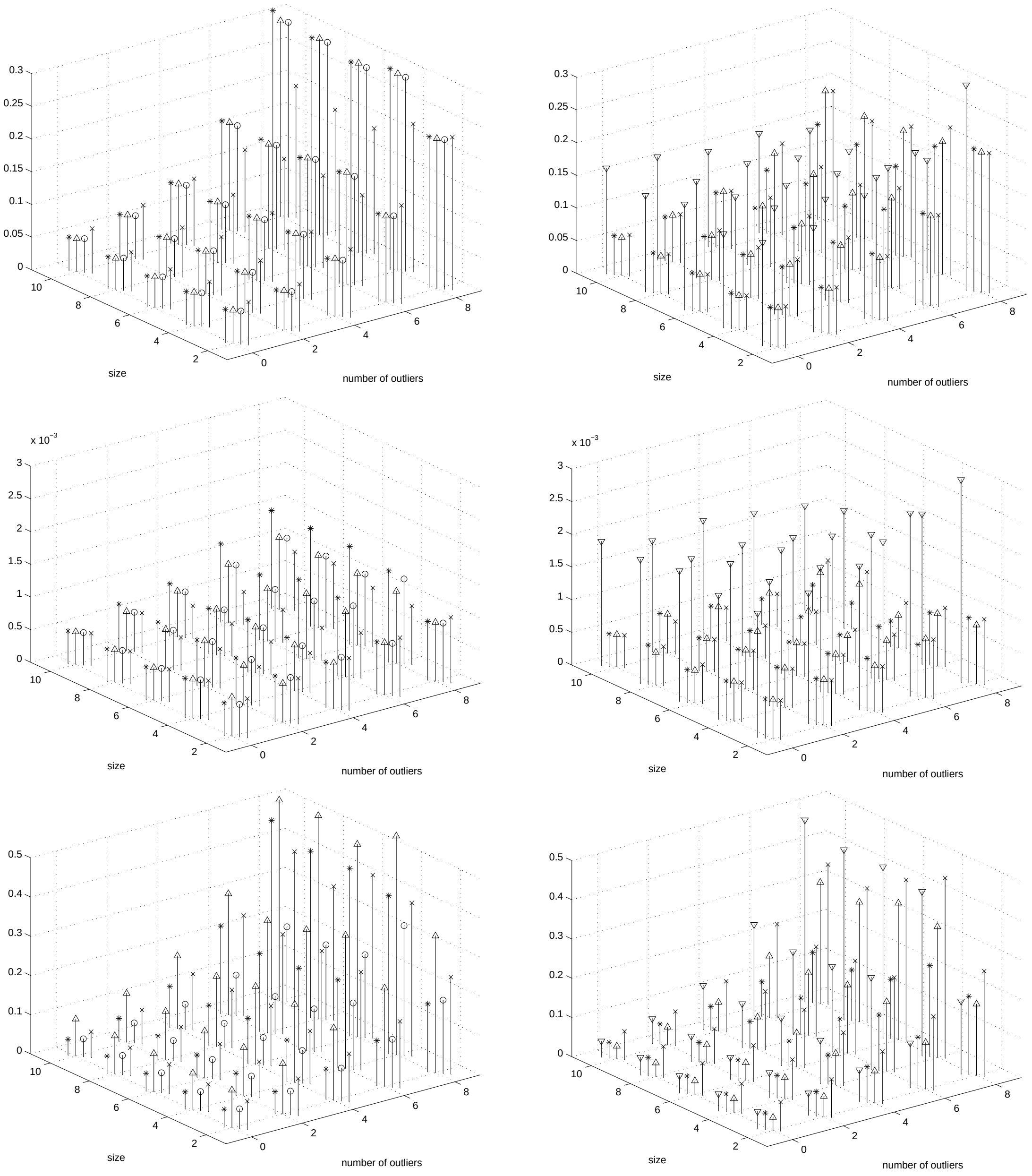

Figure 4: Small to moderate number of positive additive outliers and positive shift of size $\omega_S = 3$, MSE for the level (top), for the slope (middle) and for the scale (bottom): $T\tilde{\sigma}_{LSH}$ x, $L\tilde{\sigma}_{LSH}$ $\circ$, $M\tilde{\sigma}_{LSH}$ $\triangle$, $W\tilde{\sigma}_{LSH}$ $*$ (left), and $T\tilde{\sigma}_{MAD}$ x, $T\tilde{\sigma}_{QN}$ $\triangle$, $T\tilde{\sigma}_{SN}$ $*$, $T_{LMS}\,\&\,\tilde{\sigma}_{LSH}$ $\bigtriangledown$ (right).

case of many outliers. Downsizing L is the only serious competitor. Among the scale functionals, $\tilde{\sigma}_{QN}$ and $\tilde{\sigma}_{LSH}$ are superior for these two strategies, particularly w.r.t. the level and the scale. The T_{LMS} is best among all methods if there are more than ten large outliers. Closer examination shows that the differences are mostly due to bias as the variances increase much slower.

The results for substitutive outliers are slightly better than those for the same number of additive outliers presented here, but the ordering of the methods is essentially the same. The results for outliers with random sign also not presented here are much better than those for outliers with the same sign. Like in the case of 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 except for trimming, for which $\tilde{\sigma}_{SN}$ performs worse than the others.

4 APPLICATION

In the following we apply the methods to simulated and real time series for further comparison. Because of the previous results we restrict to procedures based on trimming or downsizing L and compare them to the T_{LMS}.

4.1 Tracking Shifts

In online monitoring data sometimes systematic sudden shifts occur. If applied without modifications for outlier and shift detection, the T_{RM} tends to smooth such level shifts as it becomes increasingly biased when more than, say, 30% of the observations in the window are affected by the shift, while the T_{LMS} resists a shift much better [5]. Therefore we add a simple rule for shift detection to the T_{RM}-based procedure as we want to track sudden shifts well.

A medical rule of thumb states that five subsequent observations which are of about the same size and differ substantially from the proceeding observations are often clinically relevant. This suggests to use rules based on runs of outlying observations for shift detection [12]. 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 the shift. Therefore, we use another simple rule for shift detection based on the residuals $r_1, \ldots, r_m$ to the right of the center of the current time window. More precisely, we decide that a positive level shift may have happened if

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

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.

In the following we add this rule for shift detection to the procedure between steps 6 and 7. 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 estimate up to time point $t + j_1 - 1$, move the center of the time window to $t + m + 1$ and restart the algorithm with step 1 as there are at most $\lfloor m/2 \rfloor$ observations in the time window which have been measured before the shift. As noted before 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 $t + m + 1$.

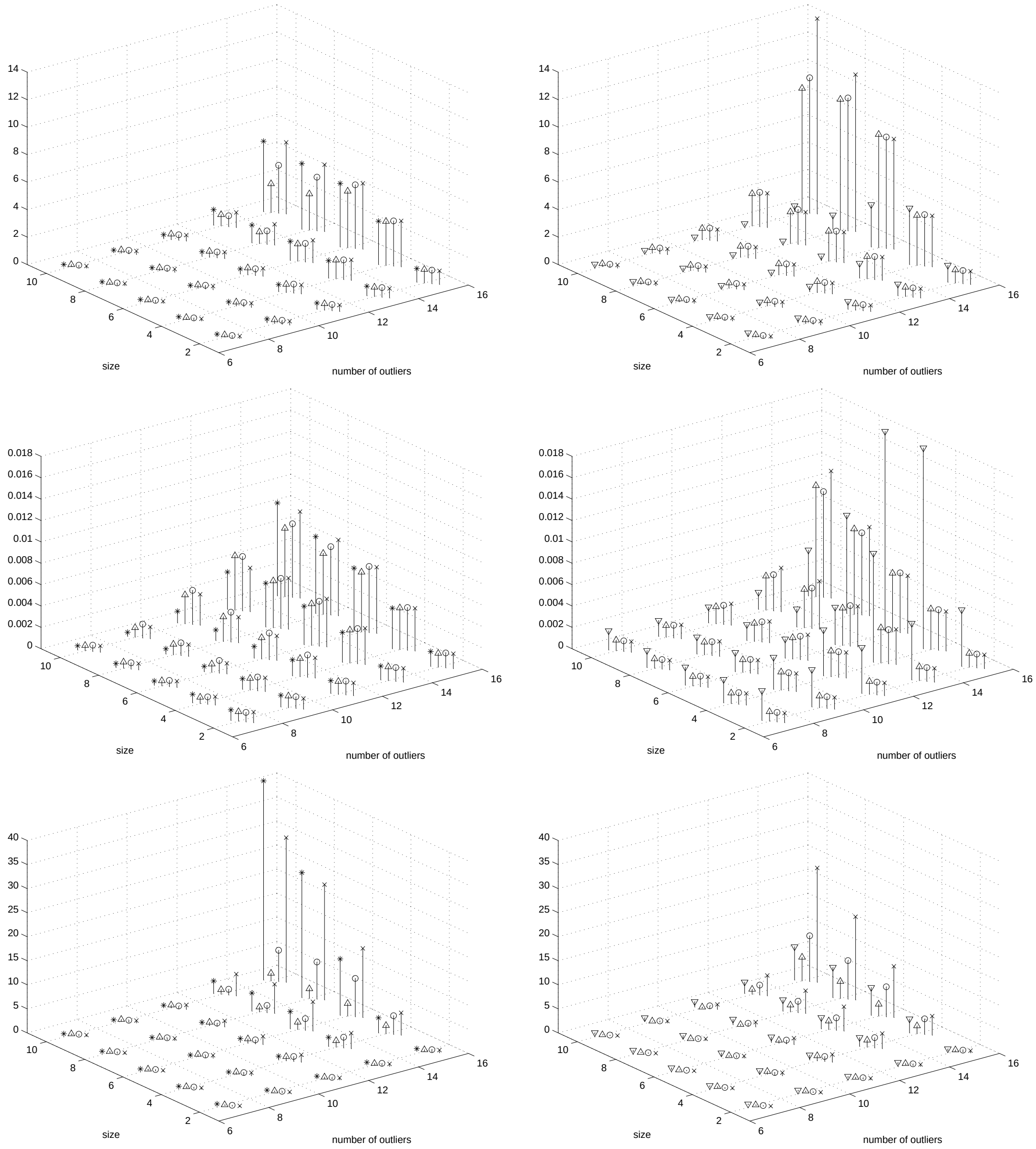

Figure 5: Large number of positive additive outliers, MSE for the level (top), for the slope (middle) and for the scale (bottom): $T\tilde{\sigma}_{MAD}$ x, $T\tilde{\sigma}_{LSH}$ $\circ$, $T\tilde{\sigma}_{QN}$ $\triangle$, $T\tilde{\sigma}_{SN}$ $*$ (left), and $L\tilde{\sigma}_{MAD}$ x, $L\tilde{\sigma}_{LSH}$ $\circ$, $L\tilde{\sigma}_{QN}$ $\triangle$, T_{LMS} & $\tilde{\sigma}_{LSH}$ $\triangledown$ (right).

There is a lot of space for improvement of shift detection as the minimal time delay of the previous rule is $\lfloor m/2 \rfloor + 1$. Complimentary run rules could be added for instance to speed up shift detection. Note, however, that it takes some time to distinguish outlier patches from level shifts anyway.

4.2 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 6. The time series is generated from Gaussian white noise with unit variance, and a deterministic trend period as well as two level shifts of size 4 and 6 respectively have been inserted. Moreover, patches of 4 (2×), 3 (4×), 2 (9×) and 1 (12×) subsequent observations have been replaced by positive additive outliers of size 6, i.e. there are 10% outliers altogether.

Downsizing L with any of the scale functionals overestimates the signal right after the end of the trend period because of an outlier patch. Every version of the procedure detects the level shifts and times them correctly at $t = 300$ and $t = 400$, while the T_{LMS} slightly increases before the first downward shift and then drops down to early. The slope approximation based on the T_{LMS} is more volatile than that based on the T_{RM}, and it is much more affected by the shifts and the outlier patches. Generally, trimming performs superior here. We got similar results for other time series simulated from similar models.

4.3 Real Time Series

The second example is a real physiologic time series representing heart rate (Figure 7). 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 [5] the positive outliers do not cause the T_{RM} in combination with trimming based on $\tilde{\sigma}_{LSH}$, $\tilde{\sigma}_{QN}$ or $\tilde{\sigma}_{SN}$ to overestimate the signal, while using $\tilde{\sigma}_{MAD}$ 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 the T_{RM} with trimming. Only for $\tilde{\sigma}_{MAD}$ we get a large negative slope at about $t = 180$. Again the T_{RM} outperforms the T_{LMS}, which is more volatile and strongly influenced by some patterns in the data.

4.4 Time Series with Non-Linear Trend

Finally, we apply the methods to 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. 10% of the observations are disturbed by additive $N(0,9)$ outliers organized in patches of 4 (3×), 3 (6×), 2 (10×) and 1 (10×) 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 Section 3 trimming showed its main weaknesses in the case of two-sided outliers considered here.

Figure 8 depicts the results for trimming with $\tilde{\sigma}_{QN}$. The results for $\tilde{\sigma}_{LSH}$ or downsizing L are very similar. Application of the T_{RM} with either of these combinations reproduces the underlying non-linear trend well, there are only some small problems with the minimum and the maximum of the signal. The T_{LMS} is again much more volatile with some superfluous bumps.

5 CONCLUSION

The extraction of an underlying signal from noisy data is a basic task for automatic online monitoring. We have found that the repeated median suggested in [5] can be further improved by

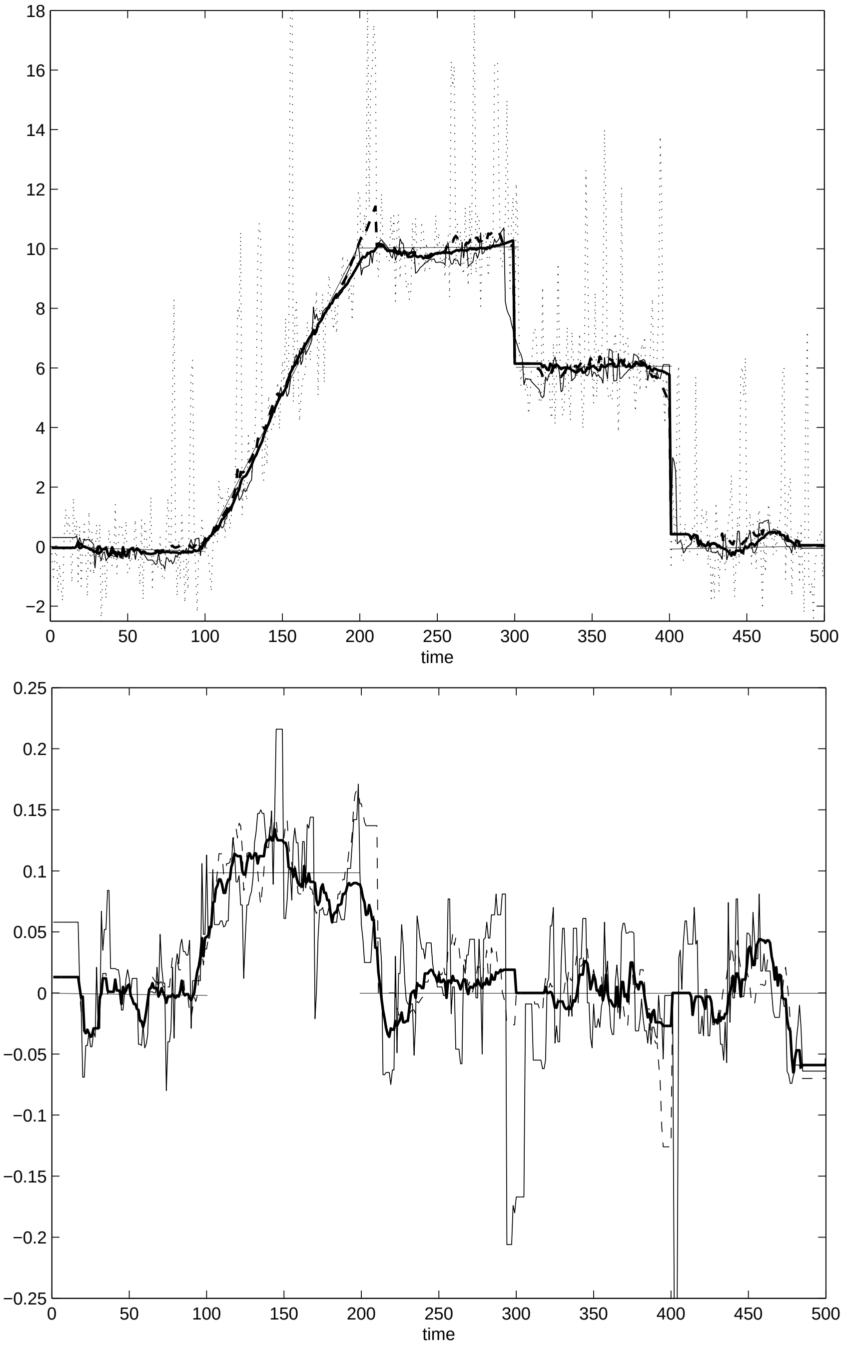

Figure 6: Top: Simulated time series (dotted), underlying level (thin solid) and level approximates: T_{LMS} (solid), $\tilde{\mu}_{RM}$ with $T\tilde{\sigma}_{QN}$ (bold solid), and $\tilde{\mu}_{RM}$ with $L\tilde{\sigma}_{QN}$ (dashed). Bottom: Underlying slope and slope approximates, same styles. Using $\tilde{\sigma}_{LSH}$ gives almost the same results.

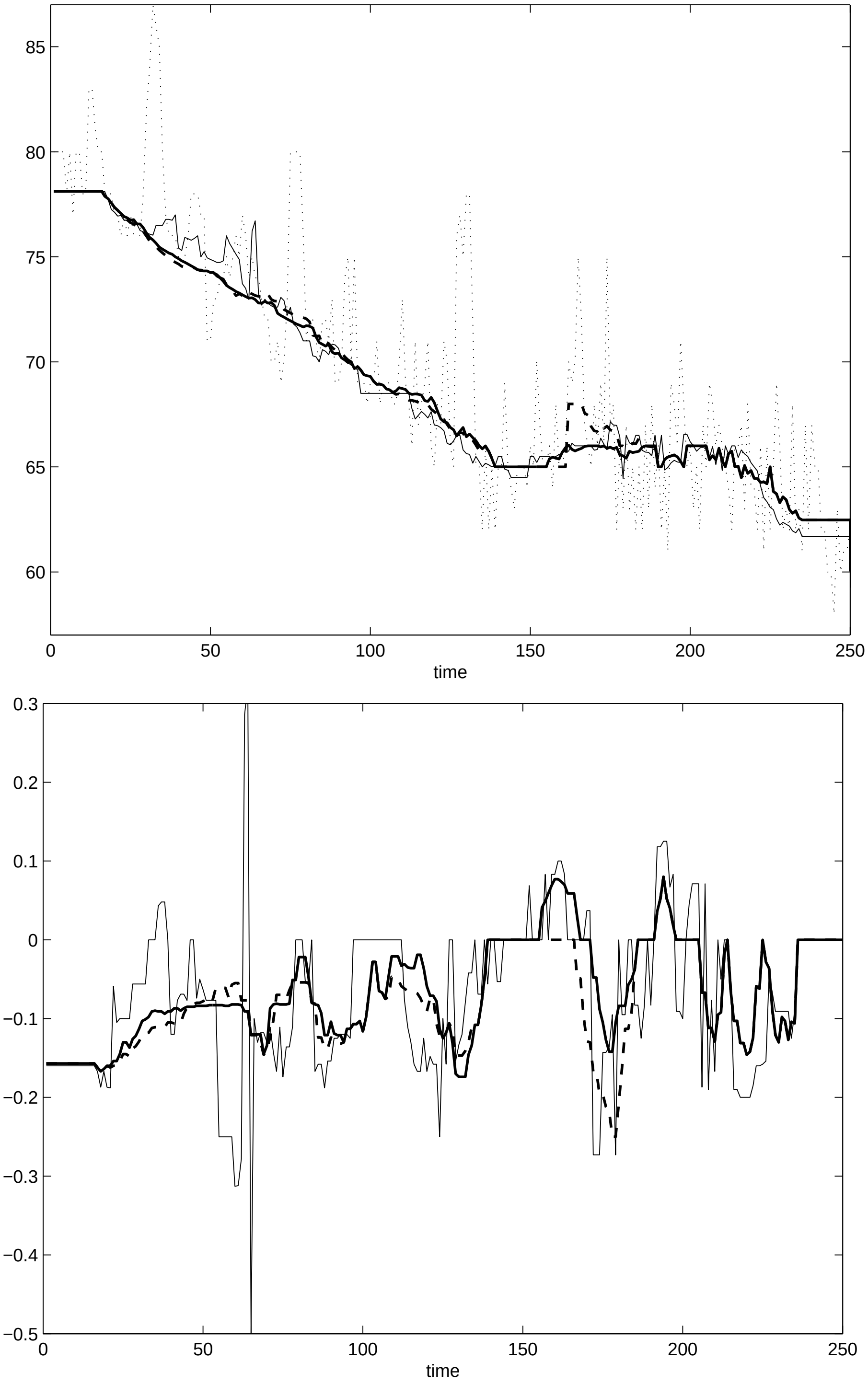

Figure 7: Top: Heart rate (dotted) and level approximates: T_{LMS} (solid), $\tilde{\mu}_{RM}$ with $T\tilde{\sigma}_{QN}$ (bold solid), and $\tilde{\mu}_{RM}$ with $T\tilde{\sigma}_{MAD}$ (dashed). Bottom: Slope approximates, same styles.

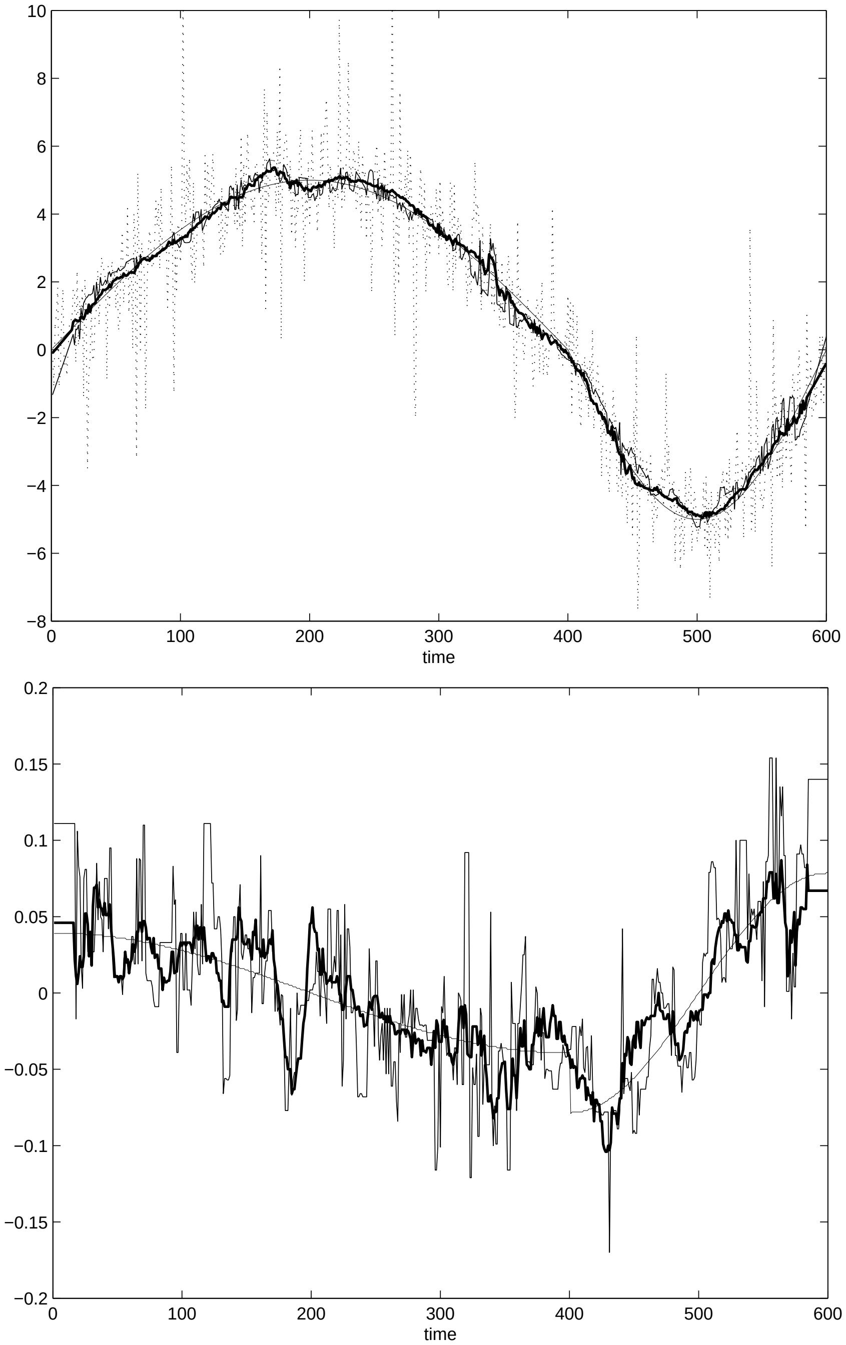

Figure 8: Top: Simulated time series (dotted), underlying level (thin solid) and level approximates: T_{LMS} (solid) and $\tilde{\mu}_{RM}$ with $T\tilde{\sigma}_{QN}$ (bold solid). Bottom: Underlying slope and slope approximates, same styles.

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 can almost achieve the high robustness of the least median of squares even in extreme outlier situations. W.r.t. the choice of the scale estimator there apparently are better choices than the classical MAD. The $\tilde{\sigma}_{QN}$ shows excellent performance for outliers of similar size, e.g. when a level shift occurs. The reason might be that $\tilde{\sigma}_{QN}$ seems well-adapted to mixtures of shifted distributions. If there are many large outliers of different sizes, however, using $\tilde{\sigma}_{LSH}$ is better as it provides stronger worst-case protection. Such combined procedures seem preferable to the LMS because of the better performance in 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.

Identical measurements due to small variability may cause problems for automatic outlier replacement. We suggest using $\tilde{\sigma}_{QN}$ then and have proposed some modifications which work very well at least in case of normal errors. An automatic procedure can still fail e.g. 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 time delay. In some applications we have an idea about a minimal variability in the data or a minimal relevant outlier and shift size which can be incorporated in the algorithm. Otherwise, one might add uniform noise according to the measurement scale both to the observations and the replacements.

Acknowledgment

The author thanks Laurie Davies and Ursula Gather for the interesting discussions and Karen Schettlinger for complementary data analysis. The financial support of the Deutsche Forschungsgemeinschaft (SFB 475, "Reduction of complexity in multivariate data structures") is gratefully acknowledged.

References

[1] Bernholt, T. and Fried, R., Computing the update of the repeated median regression line in linear time, Technical Report 43/2002, SFB 475, Department of Statistics, University of Dortmund, Germany, to appear in *Information Processing Letters.*

[2] Berrendero, J.R. and Zamar, R.H., Global robustness of location and dispersion estimates, *Statist. Probab. Lett.* **44** (1999), 63–72

[3] Croux, C. and Rousseeuw, P.J., Time-efficient algorithms for two highly robust estimators of scale, in Y. Dodge and J. Whittaker (eds.): *Computational Statistics*, Vol. **1**, Physica-Verlag, Heidelberg, 1992, 411–428.

[4] Davies, P.L., Aspects of robust linear regression, *Ann. Statist.* **21** (1993), 1843–1899.

[5] Davies, P.L., Fried, R. and Gather, U., Robust signal extraction from on-line monitoring data. Technical Report 2/2002, SFB 475, University of Dortmund, Germany, to appear in *J. Statist. Plann. and Inference.*

[6] Fried, R. and Gather, U., Fast and robust filtering of time series with trends, in W. Härdle and B. Rönz (eds.): *Proceedings in Computational Statistics COMPSTAT 2002*, Physica-Verlag, Heidelberg, 2002, 367–372.

[7] Gather, U. and Fried, R., Robust estimation of scale for local linear temporal trends, in F. Stulajter (ed.) *Proceedings of the Fourth International Conference on Mathematical Statistics PROBASTAT 2002*, to appear.

[8] Gather, U., Fried, R., Lanius, V. and Imhoff, M., Online-monitoring of high-dimensional physiological time series - a case-study, *Estadistica* **53** (2001), 259–298.

[9] Grübel, R., The length of the shorth, *Ann. Statist.* **16** (1988), 619–628.

[10] Hampel, F.R., Beyond location parameters: Robust concepts and methods, *Bulletin of the Int. Statist. Inst.* **46** (1975), 375–382.

[11] Hettmansperger, T.P. and Sheather, S.J., A cautionary note on the method of least median squares, *Amer. Statist.* **46** (1992), 79–83.

[12] Imhoff, M., Bauer, M., Gather, U. and Löhlein, D., Statistical pattern detection in univariate time series of intensive care on-line monitoring data, *Intensive Care Medicine* **24** (1998), 1305–1314.

[13] Martin, R.D. and Zamar, R.H., Bias robust estimation of scale, *Ann. Statist.* **21** (1993), 991–1017.

[14] Rousseeuw, P.J., Least median of squares regression, J. Amer. Statist. Assoc. **79** (1984), 871–880.

[15] Rousseeuw, P.J. and Croux, C., Alternatives to the median absolute deviation, *J. Amer. Statist. Assoc.* **88** (1993), 1273–1283.

[16] Rousseeuw, P.J. and Leroy, A.M., A robust scale estimator based on the shortest half, *Statistica Neerlandica* **42** (1988), 103–116. *J. Nonparametr. Statist.* **14** (2002), 155–168.

[17] Siegel, A.F., Robust regression using repeated medians, *Biometrika* **68** (1982), 242–244.

[18] Tukey, J.W., *Exploratory Data Analysis*, Addison-Wesley, Reading, Mass., 1977 (preliminary ed. 1971).