Weighted Repeated Median Smoothing and Filtering

Roland Fried
University Carlos III de Madrid
Department of Statistics
28903 Getafe, Spain

Jochen Einbeck
National University of Ireland
Department of Mathematics
Galway, Ireland

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

We propose weighted repeated median filters and smoothers for robust non-parametric regression in general and for robust signal extraction from time series in particular. The proposed methods allow to remove outlying sequences and to preserve discontinuities (shifts) in the underlying regression function (the signal) in the presence of local linear trends. Suitable weighting of the observations according to their distances in the design space reduces the bias arising from non-linearities. It also allows to improve the efficiency of (unweighted) repeated median filters using larger bandwidths, keeping their properties for distinguishing between outlier sequences and long-term shifts. Robust smoothers based on weighted L_1-regression are included for the reason of comparison.

KEY WORDS: Signal extraction; Robust regression; Outliers; Breakdown point.

1 Introduction

When extracting a time-varying level (the signal) from noisy time series, we commonly want to preserve relevant details such as monotonic trends and abrupt shifts, while eliminating irrelevant spikes due to measurement errors. Robust filtering procedures for detail-preserving signal extraction should also be fast and simple. Time series filtering is a special case of non-parametric smoothing with a fixed design.

Standard median filters suggested by Tukey (1977) remove spikes and preserve shifts. However, as reported e.g. by Davies, Fried and Gather (2004), they have difficulties if the implicit assumption that the signal is constant within each window is not fulfilled. These problems may be coped with by weighting the observations according to their temporal distances to the current target point. While the median of observations $y_1, \ldots, y_n$ minimizes the L_1-distance (or least absolute deviation, LAD), the weighted median $\hat{\mu}$ (hereafter: WM) of $y_1, \dots, y_n$ with positive weights $w_1, \dots, w_n$, which dates back to Edgeworth (1887), minimizes the weighted L_1-distance

$$\hat{\mu} = \underset{\mu}{\operatorname{argmin}} \sum_{i=1}^{n} w_i \cdot |y_i - \mu|. \tag{1}$$

In time series filtering with data $y_1, \dots, y_n$ measured at fixed design points $x_1, \dots, x_n$, we choose the w_i depending on the distances between the x_i and the target point x, $w_i = w(x - x_i)$. Here, w is a weight function decreasing monotonically at both sides from zero. Generally, locally weighted median smoothing, studied firstly by Härdle and Gasser (1984), is an effective robust nonparametric method for estimating the conditional median $\mu = g(x)$ of a response Y given a covariate x. The design variables can represent something else than time, as e.g. in image restoration.

Weighted median filters are popular because of their flexibility. For a given minimal length $\ell + 1$ of signal details to be preserved one can select a WM filter with window width larger than the $2\ell + 1$ necessary for a standard median filter. This allows more efficient noise suppression (e.g. Yang, Yin, Gabbouj, Astola and Neuvo 1995).

Local linear fits are usually preferable to local constant fits (Fan, Hu and Truong 1994). Davies et al. (2004) suggest the repeated median (RM, Siegel 1982) for the extraction of monotonic trends from time series. The repeated median estimate $(\tilde{\mu}^{RM}(x), \tilde{\beta}^{RM}(x))$ of the median and the slope at a target point x is

$$\begin{aligned} \tilde{\beta}^{RM}(x) &= \operatorname{med}_{j=1,\dots,n} \left(\operatorname{med}_{i \neq j} \frac{y_i - y_j}{x_i - x_j} \right), \\ \tilde{\mu}^{RM}(x) &= \operatorname{med} \left(y_1 - (x_1 - x)\tilde{\beta}^{RM}(x), \dots, y_n - (x_n - x)\tilde{\beta}^{RM}(x) \right). \end{aligned} \tag{2}$$

The repeated median inherits the optimal asymptotic 50% breakdown point of the standard median. Instead of a constant level, it relies on a constant slope.

In this paper we combine the concepts of weighted and repeated medians, developing robust nonparametric smoothers which adapt to monotonic trends. The resulting weighted repeated median (WRM) filters allow for application of longer time windows than 'standard' repeated median filters, without being severely biased when the signal slope varies over time. We consider two basic situations: In retrospective analysis we approximate the signal at the window center by applying a symmetric weight function putting more weight on central observations. In online analysis we approximate the signal at the current time point without time delay, using half-sided monotonic kernels giving largest weight to the most recent observations.

The paper is organized as follows. Section 2 reviews weighted medians and introduces weighted repeated medians and weighted L_1-regression. Section 3 derives analytical properties of these methods. Section 4 reports results from simulations. Section 5 exemplifies the methods on some time series, followed by some conclusions.

2 Robust smoothing and filtering

We start with alternative derivations of weighted medians. Weighted median filters give less weight to remote observations, but do not explicitly consider trends. This reduces problems due to trends, but does not overcome them completely. For further improvement we apply regression techniques with weighting according to the temporal distances. The advantages of local linear smoothers resulting from L_2-regression as compared to their local constant counterparts are well-known (Fan 1992, Hastie and Loader 1993). However, robust methods are needed in the presence of outliers. We review weighted L_1-regression before introducing weighted repeated medians.

2.1 Alternative derivations of weighted medians

For non-negative integer valued weights $w_1, \ldots, w_n$, a simple representation of the weighted median of real numbers $y_1, \ldots, y_n$ is given by

$$\hat{\mu} = \text{med}\{w_1 \diamond y_1, \ldots, w_n \diamond y_n\} \tag{3}$$

where $w \diamond y$ denotes replication of y to obtain w identical copies of it.

Notation (3) can be used in an extended way also for positive real weights: Let $y_{(1)} \leq \ldots \leq y_{(n)}$ denote the ordered observations and $w_{(1)}, \ldots, w_{(n)}$ the corresponding positive weights. Then the weighted median of $y_1, \ldots, y_n$ is $\hat{\mu} = y_{(k)}$, where

$$k = \max\left\{h : \sum_{i=h}^{n} w_{(i)} \geq \frac{1}{2}\sum_{i=1}^{n} w_i\right\}. \tag{4}$$

For example, the WM of 1, 2, 3, 7 with weights 0.1, 1.6, 1.4 and 0.5 is $y_{(3)} = 3$, since $0.5 + 1.4 \geq 3.6/2$. Generally, (4) and (1) yield the same results. However, the whole interval $[y_{(k-1)}, \ldots, y_{(k)}]$ solves (1) whenever $\sum_{i=k}^{n} w_{(i)} = \frac{1}{2}\sum_{i=1}^{n} w_i$. The solution $y_{(k-1)}$ would be obtained in (4) by summing from the bottom instead of from the top. This ambiguity can be solved as usual by choosing the midpoint of the interval.

Two weighted medians with respective weights $w_1, \ldots, w_n$ and $w'_1, \ldots, w'_n$ are called equivalent iff they give the same result for every sample. This is the case iff for every subset of indexes $I \subset \{1, \ldots, n\}$ we have

$$\sum_{i \in I} w_i \geq 0.5 \sum_{i=1}^{n} w_i \iff \sum_{i \in I} w'_i \geq 0.5 \sum_{i=1}^{n} w'_i\,.$$

For $n = 3$, the WM with weights $(w_1, w_2, w_3) = (2, 4, 3)$ is equivalent to the standard median: crucial for this is that the weights are balanced, such that no subset of less than $\lfloor (n+1)/2 \rfloor$ weights sums up to at least half the total mass. The WM is an order statistic with its rank depending on the observations and the weights.

2.2 Weighted median smoothing and filtering

Let $y_1, \ldots, y_N$ be observed at fixed design points $x_1 \leq \ldots \leq x_N$ under the model

$$Y_i = g(x_i) + u_i + v_i, \quad i = 1, \ldots, N, \tag{5}$$

where u_i is symmetric observational noise with mean zero and finite variance σ^2, and v_i is impulsive noise from an outlier generating mechanism. The goal is to approximate the signal $g(x)$ for $x \in [x_1, x_N]$, representing the level of Y as a function of x. To distinguish signal and noise we assume $\mu = g(x)$ to be smooth with infrequent shifts. The observational noise is assumed to be rough and the number of subsequent spikes to be small as compared to the durations between the shifts.

Fan and Hall (1994) and Wang and Scott (1994) propose local constant weighted L_1-estimates $\hat{g}(x)$ based on the minimization (1),

$$\hat{g}(x) = \underset{\mu}{\operatorname{argmin}} \sum_{i=1}^{N} w_i(x)|y_i - \mu| = \operatorname{med}\{w_1(x) \diamond y_1, \ldots, w_N(x) \diamond y_N\}. \tag{6}$$

In the context of nonparametric smoothing, the term *weighted* typically refers to *locally weighted*, i.e. weighting is performed by means of a kernel $K(\cdot)$, which is a continuous symmetric probability density. A common choice of the weights is

$$w_i(x) = \frac{1}{Nh} K\left(\frac{x_i - x}{h}\right). \tag{7}$$

In time series filtering, the design is usually equidistant, $x_i = i$, $i = 1, \ldots, N$. In retrospective applications, when some delay is possible, we usually approximate the level in the window center. We then apply bell-shaped weights which are symmetric to the center and monotonically decreasing to both sides of it. Symmetric bell-shaped weights can be obtained by means of symmetric unimodal kernels as in (7).

In online analysis, the target point x where we estimate the signal is at the end of the window. Then we apply monotonically increasing weights, which can be derived using half-sided bell-shaped kernels, see e.g. Einbeck and Kauermann (2003).

When using a kernel K with bounded support, say $[-1, 1]$ for a symmetric and $[-1, 0]$ for a half-sided kernel, the WM smoother (6) with weights as in (7) becomes

$$\hat{g}(x) = \underset{\mu}{\operatorname{argmin}} \sum_{i=1}^{N} w_i(x) \cdot 1_{\{-m \leq x_i - x \leq \tilde{m}\}} |y_i - \mu|, \tag{8}$$

where m is the bandwidth. The symmetric kernels in retrospective analysis result in $\tilde{m} = m > 0$, while the half-sided kernels in online analysis give $0 = \tilde{m} < m$. For every target point $x \in [m + 1; N - \tilde{m}]$ the window $\{x_i : -m \leq x_i - x \leq \tilde{m}\}$ corresponding to non-zero weights contains the same number of elements $n = m + \tilde{m} + 1$. We hence obtain weighted median filters as special cases from weighted median smoothers.

2.3 Weighted L_1-regression

Fan et al. (1994) treat a robust nonparametric median smoother based on local linear L_1-regression. They show that the theoretical properties of local linear mean estimators carry over to local linear median estimators. The local linear median at point x is given by $\hat{\mu}$, where $\hat{\mu}$ and $\hat{\beta}$ are the solutions of the weighted LAD problem

$$\min! \sum_{i=1}^{N} w_i(x)|y_i - \mu - \beta(x_i - x)| \tag{9}$$

which means fitting a straight line to the data using an additional weight function.

Like for the median, the solution of weighted L_1-regression is not unique in general. In case of a fixed design, the weights $w_1(x), \ldots, w_N(x)$ are fixed and weighted L_1-regression minimizes the regression residuals w.r.t. a norm. Thus, if the solution is not unique, the set of minimizing values is at least convex.

Several algorithms have been developed for L_1-regression in particular and quantile regression in general (Portnoy and Koenker 1997, Koenker 2005), which can be adapted to weighted L_1-regression since the ordinary L_1-solution of the modified problem

$$\min! \sum_{i=1}^{N} |w_i(x) \cdot y_i - w_i(x) \cdot \mu - \beta \cdot w_i(x) \cdot (x_i - x)| \tag{10}$$

with data $(w_i(x), w_i(x) \cdot x_i, w_i(x) \cdot y_i)$ is the same as the weighted L_1-solution of the original problem. We use an approximative L_1-procedure, which offers simplicity and increased robustness. Starting from the standard repeated median, the algorithm iterates a finite number of steps between maximization of the objective function w.r.t. μ given the current solution for β and vice versa.

2.4 Weighted repeated medians

Davies et al. (2004) investigate robust regression techniques like the standard repeated median and L_1-regression for delayed signal extraction from time series. Online versions of such procedures are compared by Gather et al. (2006). The repeated median smoother (2) is found to be preferable to the inspected alternatives in both situations. In time series filtering, the setting $n = 2m + 1$, $x_1 = t - m, \ldots, x_n = t + m$, and $x = t$ corresponds to the retrospective, symmetric situation, while $n = m + 1$, $x_1 = t - m, \ldots, x_n = t$, and $x = t$ corresponds to the online, no delay version.

The resulting (standard) repeated median filters fit a linear trend $\mu_{t+j} = \mu_t + j\beta_t$, $j = -m, \ldots, \tilde{m}$, to the data in each time window, where $\tilde{m} = m$ or $\tilde{m} = 0$ depending on the situation. The assumption of a locally constant signal underlying the standard median is replaced by a locally linear trend with constant slope. This motivates us to generalize the repeated median, permitting localization by weighting.

Consider a window of width n with observations $(x_1, y_1), \ldots, (x_n, y_n)$, where w.l.o.g. the predictors are ordered such that $x_1 < \ldots < x_n$. The weighted repeated median (WRM) with two possibly different sets of weights w_i, $\tilde{w}_i$, $i = 1, \ldots, n$, is given by

$$\tilde{\beta}^{WRM}(x) = \text{med}_{j=1,\ldots,n} \tilde{w}_j \diamond \left(\text{med}_{i \neq j} \tilde{w}_i \diamond \frac{y_i - y_j}{x_i - x_j} \right), \tag{11}$$

$$\begin{aligned} \tilde{\mu}^{WRM}(x) &= \text{med}\left(w_1 \diamond \left(y_1 - (x_1 - x)\tilde{\beta}^{WRM}(x) \right), \ldots, \right. \\ &\quad \left. w_n \diamond \left(y_n - (x_n - x)\tilde{\beta}^{WRM}(x) \right) \right), \end{aligned} \tag{12}$$

i.e. we weight the pairwise slopes in the inner median by a weight depending on the position of x_i, and in the outer median on the position of x_j when estimating the slope $\beta(x)$. The set of weights $w_1, \ldots, w_n$ used for the level $\mu(x)$ can be identical to $\tilde{w}_1, \ldots, \tilde{w}_n$. Anyway, we choose both sets of weights w_i and $\tilde{w}_i$ to be symmetric bell-shaped in retrospective and to be monotonic in online applications.

We call two WRMs with weights $w_1, \ldots, w_n, \tilde{w}_1, \ldots, \tilde{w}_n$ and $w'_1, \ldots, w'_n, \tilde{w}'_1, \ldots, \tilde{w}'_n$ equivalent if the slope and the level estimate are always identical. A necessary condition for this is the equivalence of the WMs corresponding to $w_1, \ldots, w_n$ and $w'_1, \ldots, w'_n$: if the slope estimates are identical, there are samples such that the WMs of the slope-corrected observations are different otherwise. The following additional condition for $\tilde{w}_1, \ldots, \tilde{w}_n$ and $\tilde{w}'_1, \ldots, \tilde{w}'_n$ guarantees the equivalence of WRMs:

The weighted medians corresponding to $\tilde{w}_1, \ldots, \tilde{w}_n$ and to $\tilde{w}'_1, \ldots, \tilde{w}'_n$ are equivalent, and for each $i \in \{1, \ldots, n\}$ the weighted medians corresponding to $\tilde{w}_1, \ldots, \tilde{w}_{i-1}, \tilde{w}_{i+1}, \ldots, \tilde{w}_n$ and to $\tilde{w}'_1, \ldots, \tilde{w}'_{i-1}, \tilde{w}'_{i+1}, \ldots, \tilde{w}'_n$ are also equivalent.

This condition is sufficient, not necessary. It is stricter than the equivalence of the WMs corresponding to $\tilde{w}_1, \ldots, \tilde{w}_n$ and $\tilde{w}'_1, \ldots, \tilde{w}'_n$: For $n = 3$, the WRM with $(\tilde{w}_1, \tilde{w}_2, \tilde{w}_3) = (2, 4, 3)$ is not equivalent to the standard RM, although the WM is equivalent to the standard median. For the sample $(1, 0), (2, 1), (3, 5)$, the standard RM slope is $\text{med}(1.75, 2.5, 3.25) = 2.5$, while for the WRM it is $\text{med}(1, 4, 4) = 4$.

In nonparametric regression, when approximating the regression function g at x given N data points, it is natural to employ kernel weights $w_i = w_i(x)$, $\tilde{w}_i = \tilde{w}_i(x)$ as defined in (7). The estimated regression function is then given by $\tilde{g}(x) = \tilde{\mu}^{WRM}(x)$.

2.5 Alternative Approaches

There are locally weighted versions of more robust regression techniques: Equal weighting results in the highest efficiency of weighted Theil-Sen estimators and the highest asymptotic breakdown point of 29.3% among all efficiency-optimal weighting schemes in the case of an equally spaced fixed design (Scholz 1978). Simpson and Yohai (1998) discuss the stability of one-step GM estimators (including weighted L_1-regression) in approximately linear regression with a random design.

3 Analytical properties

We analyze properties of the smoothers described above. Applying a kernel K with bounded support and weights as in (7), for every target point x the subset $W(x)$ of design points with non-zero weights forms a window of subsequent points. We discuss a single window of width n. Let $y_1, \ldots, y_n$ be the corresponding values of a response observed at fixed $x_1 < \ldots < x_n$. Denote the corresponding sets of strictly positive weights by $w_1, \ldots, w_n$ and $\tilde{w}_1, \ldots, \tilde{w}_n$, suppressing the dependence on x.

3.1 Equivariances

Equivariances guarantee that an estimate reacts as expected to systematic changes in the data. Location equivariance means that adding a constant c changes the estimate by c. Scale equivariance means that multiplying all of $y_1, \ldots, y_n$ by c changes the estimate by the same factor. The level estimates obtained from weighted medians and weighted repeated medians possess both these properties.

We also require that the quality of the smoothing does not depend on linear trends. This can be guaranteed by applying regression equivariant estimators. When regressing a variable y on a variate $z \in \mathbb{R}^d$, regression equivariance means that adding a vector multiple $c'z$ of z to y for a $c \in \mathbb{R}^d$ changes the estimate by this vector c.

(Weighted) repeated medians for simple linear regression as defined here are equivariant w.r.t. adding a vector multiple $(a, b)z_i = a + bx_i$ of $z_i = (1, x_i)'$ to y_i, $i = 1, \ldots, n$. A procedure for (weighted) L_1-regression fulfills this equivariance if the initial estimator, e.g. the repeated median, fulfills it since we just act on the residuals thereafter. The performance of weighted medians depends on trends since they do not make use of the covariate values $x_1, \ldots, x_n$. They regress on a constant level only, i.e. $z = 1$, so that regression and location equivariance coincide.

3.2 Removal of spiky noise

The removal of impulsive noise (spikes, outliers) and the preservation of relevant signal details, in particular of long-term shifts, are essential properties of robust smoothers. The performance of a regression method w.r.t. outliers can be measured by two related quantities, the breakdown point and the exact fit point.

The asymptotic breakdown point of the standard median and repeated median is 50%. This asymptotic breakdown point is the limit of the finite sample replacement breakdown point, which measures the minimal fraction of data which can drive an estimate beyond all bounds when being set to arbitrary values (Donoho and Huber 1983). In the context of nonparametric smoothing by moving window techniques, this corresponds to the minimal fraction of contamination within a window which can cause an arbitrarily large spike in the output. It is well known that for local fits based on (weighted) least squares a single outlier can cause an arbitrarily large spike. See Davies and Gather (2005) for a discussion of breakdown points.

Another popular quantity in signal extraction is the number of spikes a procedure can remove completely from a prototype signal in noise-free conditions, where the variance σ^2 of the observational noise equals zero. When applying a regression functional to a moving window assuming a locally linear signal trend, this number of spikes corresponds to the exact fit point of the functional. The exact fit point is the smallest fraction of observations which can cause an estimated regression hyperplane to deviate from another hyperplane although all the remaining data points lie on that hyperplane (Rousseeuw and Leroy 1987, Section 3.4). For regression and scale equivariant functionals the exact fit point is not smaller than the finite sample breakdown point. Let $\lfloor a \rfloor$ be the largest integer not larger than a. The standard median fits a constant exactly if less than $\lfloor (n+1)/2 \rfloor$ out of n observations are distinct from it, which equals its breakdown point. Up to $\lfloor (n-1)/2 \rfloor$ subsequent spikes are removed completely from a constant signal. In retrospective application, a shift from one constant to another one is preserved exactly when applying an odd $n = 2m+1$. In online application, the shift gets delayed by m time points then.

However, within a trend period a standard median cannot preserve exactly a shift into the opposite direction, and a single spike causes smearing (e.g. Fried, Bernholt and Gather 2006). This is an advantage of regression techniques: The removal of outliers and the preservation of shifts does not depend on linear trends since the WRM and weighted L_1-regression are equivariant to them. The breakdown and the exact fit point of the standard RM for fitting a straight line both equal $\lfloor n/2 \rfloor / n$. Thus, the standard RM can remove $\lfloor n/2 \rfloor - 1$ subsequent spikes from a linear trend, which is only slightly less than for the standard median when the signal is constant.

For the derivation of breakdown and exact fit points of robust weighted regression methods, let $z_i \in \mathbb{R}^d$ be fixed regressors, $\gamma \in \mathbb{R}^d$ the parameter to be estimated, and

$$y_i = z_i' \cdot \gamma + u_i, \; i = 1, \dots, n.$$

Weighted L_1-regression can be analysed using results for standard L_1-regression considering the modified problem (10). From He, Jureckova, Koenker and Portnoy (1990, Theorem 5.3), Ellis and Morgenthaler (1992, Theorem 2.3) and Mizera and Müller (1999, Theorem 2) we can conclude that the breakdown point and the exact fit point of weighted L_1-regression are identical and equal to k/n, where $k = \min |I|$, $I \subset \{1, \dots, n\}$, for which $0 \neq \tilde{\gamma} \in \mathbb{R}^d$ exists such that

$$\sum_{i \in I} w_i \cdot |z_i' \cdot \tilde{\gamma}| \geq \sum_{i \notin I} w_i \cdot |z_i' \cdot \tilde{\gamma}| \, . \tag{13}$$

Since a WM regresses on a constant, $z_i \equiv 1$, its breakdown and exact fit point is the minimal fraction of weights which sum up to at least $0.5 \sum_{i=1}^n w_i$. It is straightforward to show that a WM which is not equivalent to the standard median has breakdown point smaller than the optimal value $\lfloor (n+1)/2 \rfloor / n$ of the latter. The loss in robustness due to weighting will be the larger, the more the weights vary.

Calculating the numerical value of the breakdown and exact fit point of (weighted) L_1-regression is more difficult in case of $d \geq 2$ since more directions need to be considered then. An algorithmic solution is given by Giloni and Padberg (2004).

In the case of simple linear regression, $y_i = \mu + \beta(x_i - x)$, we can derive simple upper bounds, choosing the coordinate axis as directions $\tilde{\gamma}$ in (13): The breakdown point of weighted L_1-regression with weights $w_1, \ldots, w_n$ is not larger than $\min\{k_l, k_s\}/n$, where k_l is the minimal cardinality of $I \subset \{1, \ldots, n\}$ such that

$$\sum_{i \in I} w_i \geq \sum_{i \notin I} w_i$$

and k_s is the minimal cardinality of $I \subset \{1, \ldots, n\}$ such that

$$\sum_{i \in I} w_i |x_i - x| \geq \sum_{i \notin I} w_i |x_i - x| \,.$$

This upper bound is generally not strict as it only considers two directions: For standard L_1-regression and an equidistant, centered design the upper bound is $1 - 1/\sqrt{2} = 29.3\%$ asymptotically. However, from Ellis and Morgenthaler (1992, Proposition 4.1) we derive that the asymptotic breakdown point of standard L_1-regression is smaller, namely at most 25%. Nevertheless, the simple upper bound is attained by the approximative weighted L_1-algorithm outlined in Section 2.3.

In case of $n = 7$ and an equidistant centered design, weighting allows to increase the breakdown point of approximative L_1-regression from 2/7 to 3/7 by choosing $w_i = 1/\sqrt{1 + |x_i - x|}$. The terms w_i and $w_i|x_i - x|$ in the two restricting inequalities are identical for these weights, except for the center, where $x_i = x$.

Next we address breakdown and exact fit of weighted repeated medians.

Proposition 1 *Let $(\tilde{\mu}, \tilde{\beta})$ be a weighted repeated median of n observations with weights $w_1, \ldots, w_n$ and $\tilde{w}_1, \ldots, \tilde{w}_n$.*

a) A lower bound for the breakdown and the exact fit point of $(\tilde{\mu}, \tilde{\beta})$ is given by $\min\{k_s, k_l\}/n$, where k_s is the minimal number of weights for which $\sum_{i=1}^{k_s} \tilde{w}_{[i]} \geq \sum_{i=k_s+2}^{n} \tilde{w}_{[i]}$, with $\tilde{w}_{[1]} \geq \tilde{w}_{[2]} \geq \ldots \geq \tilde{w}_{[n]}$ denoting the ordered sequence of weights, and k_l is defined as the minimal number of weights $w_{[1]} \geq w_{[2]} \geq \ldots \geq w_{[n]}$ for which $\sum_{i=1}^{k_l} w_{[i]} \geq \sum_{i=k_l+1}^{n} w_{[i]}$.

b) An upper bound for the exact fit and the breakdown point of $(\tilde{\mu}, \tilde{\beta})$ is given by $\min\{k'_s - 1, k_l\}/n$, where k_l is as in a) and k'_s is the minimal number of weights for which $\sum_{i=2}^{k'_s} \tilde{w}_{[i]} \geq \sum_{i=k'_s+1}^{n} \tilde{w}_{[i]}$.

c) *The breakdown point and the exact fit point of* $(\tilde{\mu}, \tilde{\beta})$ *do not exceed the* $\lfloor n/2 \rfloor /n$ *value of the standard repeated median.*

Proof of Proposition 1. Since for regression and scale equivariant functionals like WRMs the exact fit point (EFP) is at least as large as the finite-sample breakdown point (BP), it suffices to prove a) for the BP and b) for the EFP.
a) Less than $k = \min\{k_s, k_l\}$ modifications have bounded effect on the level and the slope: When excluding an unmodified, 'clean' observation y_j, the sum of the weights is still larger for the clean than for the modified observations. Hence, for every clean y_j the inner median in the slope corresponds to a clean pair and is bounded. The WRM slope is bounded by the same quantity. The weighted majority of the slope corrected y_j and thus the WRM level is then also bounded.
b) Because of regression equivariance we may assume that all observations are zero, and need to find $k = \min\{k'_s - 1, k_l\}$ substitutions causing the fit to deviate from the horizontal axis. If $k = k'_s - 1$, let the positions $I = \{i_1, \ldots, i_{k+1}\}$ correspond to the largest weights $\tilde{w}_{[1]} \geq \ldots \geq \tilde{w}_{[k+1]}$. Set the rightmost k of these observations, i.e. with largest x, on an increasing line with slope $b > 0$ through the leftmost of them. For each observation in I the total weight of the other observations in I is at least the total weight of the unmodified zero observations. The corresponding inner medians and the WRM slope is hence at least $b/2$. If $k = k_l$, set the k observations with largest w_i to an arbitrary value M, obtaining a WRM level of at least $M/2$.
c) The standard RM has maximal BP among regression equivariant methods including WRMs. Its EFP is maximal as it equals its upper bound, $k_s = k'_s - 1$. □

The lower and the upper bound given in a) and b) are not always identical, consider $n = 5$ and $(w_1, \ldots, w_5) = (\tilde{w}_1, \ldots, \tilde{w}_5) = (1, 1, 1, 3, 2)$, for which $k_s = 1$, but $k'_s = 3$. The next result shows that the lower bound is attained in the most relevant cases.

Proposition 2 *The breakdown and the exact fit point of a weighted repeated median with symmetric bell-shaped or monotonic weights equal* $\min\{k_s, k_l\}/n$.

Proof of Proposition 2. It is sufficient to prove that the EFP equals its lower bound. We assume w.l.o.g. that all n observations equal zero and show that $k = \min\{k_l, k_s\}$ modifications can make the WRM line deviate from the horizontal axis.
Symmetric bell-shaped weights and monotonic weights can be treated in the same way. The k largest weights $\tilde{w}_j$ are at subsequent positions $x_{i-k+1} < \ldots < x_i$.
If $k_s \leq k_l$, proceed as follows: If $\tilde{w}_1 + \ldots + \tilde{w}_{i-k} \geq \tilde{w}_{i+1} + \ldots + \tilde{w}_n$, set the k observations at $x_{i-k+1}, \ldots, x_i$ to an increasing line with slope 1 through $(x_{i-k}, 0)$. $\tilde{w}_{i-k}$ is the $(k+1)$th largest $\tilde{w}_j$ then. The pairwise slope is 1 if both design points are selected from $x_{i-k}, \ldots, x_i$, it is strictly positive if one is from $x_1, \ldots, x_{i-k-1}$ and the other from $x_{i-k+1}, \ldots, x_i$, and it is zero if both are from $x_1, \ldots, x_{i-k}, x_{i+1}, \ldots, x_n$. The inner median corresponding to x_{i-k} is strictly positive since the total weight of

the modified is at least that of the unmodified observations. This also holds for those at $x_{i-k+1}, \ldots, x_i$ since the pairwise slopes through $x_1, \ldots, x_i$ are larger than zero. Since the total weight at $x_{i-k}, \ldots, x_i$ is larger than the rest, the WRM slope is larger than zero and the WRM line deviates from the horizontal axis.
If $\tilde{w}_1 + \ldots + \tilde{w}_{i-k} < \tilde{w}_{i+1} + \ldots + \tilde{w}_n$, set the observations at $x_{i-k+1}, \ldots, x_i$ to an increasing line through $(x_{i+1}, 0)$ and use the same arguments as before interchanging the role of $x_1, \ldots, x_{i-k}$ and $x_{i+1}, \ldots, x_n$.
If $k_l < k_s$, set the k observations with largest w_i to 1. From the proof of Proposition 1a) follows that the slope estimate is zero, but the level estimate is at least 0.5. □

There are also WRMs which attain their respective upper bound, e.g. the one for $n = 5$ mentioned above. The previous results allow to determine weighted L_1- and WRM filters which remove outlier patches up to a given length completely while exactly preserving longer shifts under idealized conditions ($\sigma^2 = 0$).

We consider two weighting schemes and an equidistant design $x_1 = 1, \ldots, x_n = n$ latter on: The first scheme $w_i^{(1)}(x) = 1 - [|x - x_i|/(m+1)]^2$, with x being the target point, stems from the Epanechnikov kernel. The second one motivated by L_1-regression is $w_i^{(2)}(x) = (1 + |x - x_i|)^{-1/2}$. The Epanechnikov weights $w^{(1)}$ are flat close to x and decay strongly away from it, while it is the other way round for $w^{(2)}$. The standardized weighting schemes are compared in Figure 1.

Figure 1: Standardized symmetric bell-shaped weights $w^{(1)}$ obtained from the Epanechnikov kernel (○), $w^{(2)}$ (▽) and the uniform weights of the standard version (□). The weights $w^{(1)}$ are flat close to the target point, where $x_i - x = 0$, and strongly decaying away from it, while it is the other way round for $w^{(2)}$.

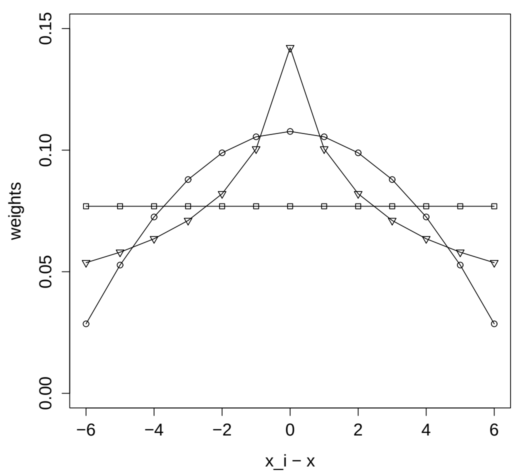


Table 1 gives the minimal window widths n necessary to remove outlier patches of different lengths for standard and weighted L_1- and RM filtering. We observe that n increases for the WRM as compared to the standard RM, while weighting allows to decrease n for online L_1-filtering because of increased robustness. Nevertheless, L_1-regression does not achieve the optimal robustness of the standard RM when using these weighting schemes and needs somewhat larger n. The second scheme affords somewhat larger n in online, and slightly smaller n in retrospective RM filtering.

Table 1: Minimal window width n necessary to remove outlier patches of length ℓ in online (left) and retrospective (right) application, L_1- (top) and RM-regression (bottom).

ℓ	1	2	3	4	5	6	1	2	3	4	5	6
standard L_1	5	8	11	15	18	22	5	7	9	11	13	15
$w_i^{(1)}(x) = 1 - (\lvert x_i - x\rvert/(m+1))^2$	4	7	10	12	15	18	5	7	9	11	15	17
$w_i^{(2)}(x) = (1 + \lvert x_i - x\rvert)^{-1/2}$	4	7	10	12	16	19	5	7	9	11	15	17
standard RM	4	6	8	10	12	14	5	7	9	11	13	15
$w_i^{(1)}(x) = 1 - (\lvert x_i - x\rvert/(m+1))^2$	4	7	10	13	16	19	5	7	11	13	15	19
$w_i^{(2)}(x) = (1 + \lvert x_i - x\rvert)^{-1/2}$	4	7	11	14	17	21	5	7	9	13	15	19

3.3 Continuity

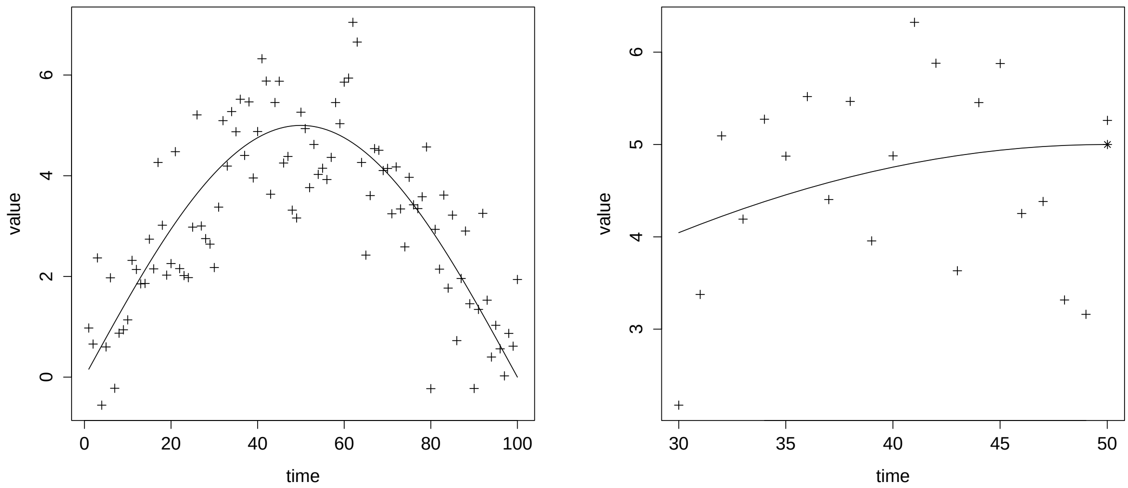
(Lipschitz) continuity guarantees local stability to small changes in the data due to observational noise or rounding. Every WM is Lipschitz continuous with constant 1 as changing every observation by less than δ changes any order statistic at most by δ, and a WM always corresponds to one of these. For fixed design, the slope estimate of a WRM changes at most by $2\delta/\min_{i=1,\dots,n}(x_i - x_{i-1})$, so that the WRM level estimate is Lipschitz continuous with constant $2\max\{|x_1 - x|, |x_n - x|\}/\min_{i=2,\dots,n}(x_i - x_{i-1})$ since none of the slope corrected observations changes more.

4 Monte Carlo study

A common demand for robust filters discussed in Section 3.2 is that long-term shifts should be preserved, while irrelevant sequences of spikes should be removed. We compare the performance of the filters in simulations, concentrating on equidistant designs as encountered in time series filtering. Data are generated from model (5) with standard Gaussian white noise u_i. The signal is a sine function, $g(x_i) = \nu \cdot 0.5 \cdot \sin(i \cdot \pi/100)$, $i = 1, \dots, 100$, where $\nu \in \{0, 1, \dots, 20\}$ determines the degree of non-linearity (Figure 2). We treat a single window with target point $x = 50$.

A rule of thumb in intensive care says that five subsequent strongly deviant observations in hemodynamic time series point at a clinically relevant shift, while shorter sequences are typically irrelevant (Imhoff et al. 2002). Accordingly, we fix window widths with the aim of preserving shifts lasting at least $\ell = 5$ observations.

Figure 2: Senoidal signal $\mu_t = 5 \cdot \sin(t \cdot \pi/100)$, $t = 1, \ldots, 100$, overlaid by Gaussian white noise with unit variance (left), and exemplary data window of width 21 used for online approximation of the signal value ($*$) at the target point $t = 50$ (right).



4.1 Online signal extraction

We start with the online versions of the procedures, choosing suitable window widths from Table 1. For the standard RM and L_1-regression we select $n = 11$ and $n = 15$, respectively. For the WRM with weights $w^{(1)}$ ($w^{(2)}$) we use a larger $n = 15$ ($n = 16$), while for L_1-regression weighted by $w^{(1)}$ ($w^{(2)}$) we choose $n = 14$ ($n = 15$).

Comparing the ability of the procedures to distinguish relevant from irrelevant deviating sequences, we generate data resembling the intrusion of a shift into the window. We simulate data as described above, adding the same constant c to an increasing number of observations at the window end. In accordance to the above demands, up to four shifted observations are regarded as outliers and should not affect the estimation, while from five shifted observations on the shift should be reproduced.

Figure 3 compares the bias of the approximation of the non-shifted signal caused by $\ell = 1, 2, \ldots, 11$ shifted observations at time points $t = 50, 49, \ldots, 40$, calculated from 2000 windows each. A curve would be optimal if it stayed at zero up to $\ell = 4$, and then increased abruptly to the added constant c representing the new level.

We observe that all versions of L_1-regression have difficulties in distinguishing relevant from irrelevant patterns. Although the widths are chosen to achieve adequate breakdown and exact fit points, the versions which down-weight remote observations are strongly influenced already by four shifted observations, particularly $w^{(2)}$. The desired delay of tracking is only obtained for huge shifts, according to the breakdown asymptotics. Standard L_1-regression resists too many remote observations and neither tracks shifts properly, although n is at the lower limit for $\ell = 4$ outliers, see

Figure 3: Online application: Bias for the level (left) and the slope (right) due to an increasing number of observations shifted by $c = 10$ (top), $c = 100$ (center) and $c = 10000$ (bottom) at the end of the window: RM (solid lines) and L_1-regression (dashed), standard version ($\Box$), weights $w^{(1)}$ (o) and $w^{(2)}$ ($\triangledown$).

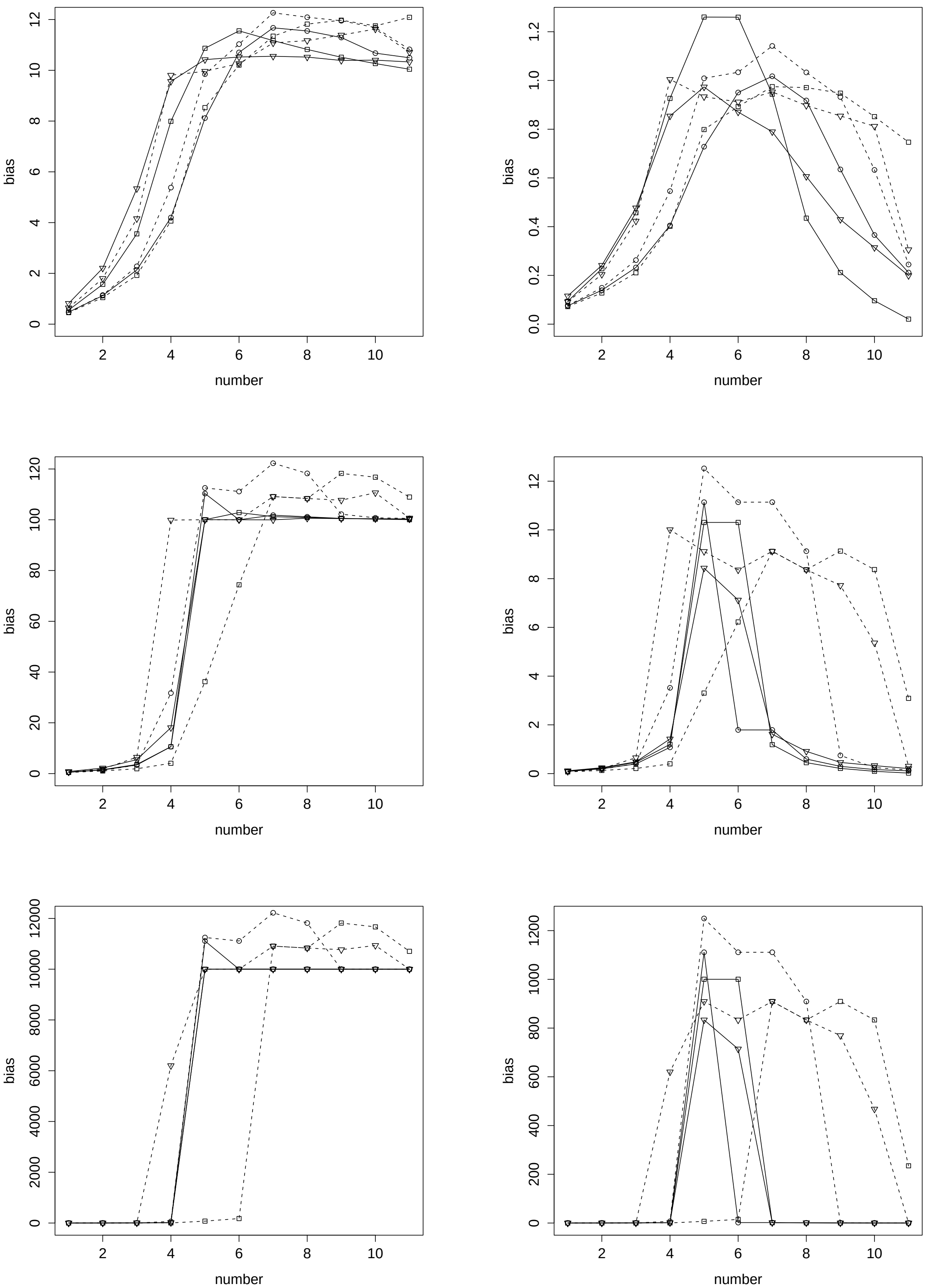


Table 1, as opposed to the widths for the other methods. L_1-regression is influenced by leverage points, i.e. by the oldest observations in the online situation. Incoming shifted observations are not in worst case positions. Rules for shift detection or smaller widths were needed, but the latter increase all outlier effects. Additionally, all versions of L_1-regression overshoot the signal value after the shift.

The WRMs preserve the shifts better, obtaining the desired delay of tracking. Their superiority is further increased when taking the variability of the estimates (not shown) into account, which is for about four outlying observations much less than for (weighted) L_1-regression. The WRM with weights $w^{(1)}$ performs best for moderate shifts, but it overshoots huge shifts. Weighting by $w^{(2)}$ preserves huge shifts better.

For the slope, we would consider a bias curve as optimal if it stayed constantly at zero. The results, depicted also in Figure 3, are similar as for the level. Weighting reduces the bias of the RM, particularly for moderate shifts. For large shifts, the WRMs, in particular with $w^{(1)}$, tend to be less biased than (weighted) L_1-regression.

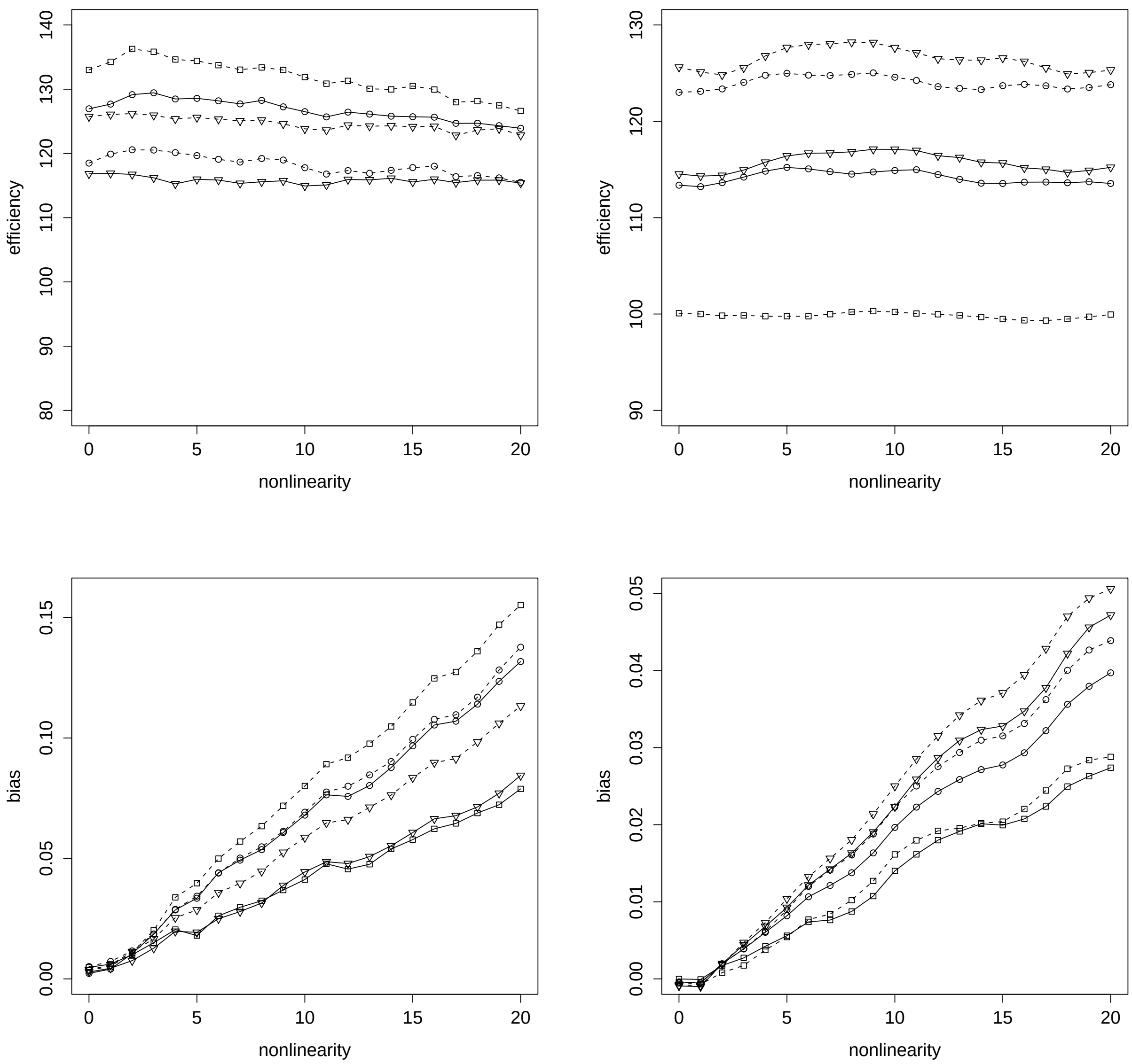
Figure 4 compares the efficiencies for Gaussian noise in dependence on the nonlinearity ν, in the absence of outliers and shifts. Because of the bias for $\nu \neq 0$ we measure the efficiency by the percentage mean square error MSE as compared to the standard RM, obtained from 10000 runs for each $\nu = 0, \ldots, 20$. Standard L_1-regression turns out to be more efficient than its weighted versions, although this advantage decreases in ν due to a more increasing bias. For the RM, weighting and the therefore possible larger n increase the efficiency, and even more so for the slope. The Epanechnikov weights $w^{(1)}$ give the highest efficiencies for the WRM, but the smallest for L_1. The latter aspect can be due to the smaller n, and the former to the fact that more observations close to the target x get large weights.

4.2 Retrospective signal extraction

We also compare the filters in the retrospective situation using symmetric bell-shaped weights. Deviating from Table 1, we use $n = 9$ for standard RM and L_1-regression. Relevant shifts would be smoothed a lot otherwise. The widths in Table 1 do not guide detail-preserving retrospective smoothing by the RM and L_1. Table 1 corresponds to the worst-case, but centric outliers affect the slope estimation only mildly and are not worst-case. Accordingly, we use shorter windows corresponding to those necessary for weighted medians with the respective weights.

Figure 5 depicts the results for a window centered at the target point $x = 50$ and an outlier sequence at $x_i = 50, 51, \ldots$, i.e. just starting in the center, using $n = 13$ for the RM and the L_1 with either set of weights. All procedures reduce a shift at its starting point, irrespective of its duration. The standard versions of the filters are

Figure 4: Percent efficiency for the level relatively to the standard RM (top) and absolute bias (bottom) in online (left) and retrospective (right) application in the case of Gaussian noise as function of the amount of non-linearity ν: RM (solid lines) and L_1-regression (dashed lines), standard version ($\square$), weights $w^{(1)}$ ($\circ$) and $w^{(2)}$ ($\triangledown$).



sensitive to three or four outliers. The WRMs perform better in this respect and close to the weighted L_1-procedures for moderate outliers, but they resist too many huge outliers delaying such shifts. When reducing the width to $n = 11$ for the WRMs to overcome this delay, the WRMs are considerably more affected in case of three or four moderately large outliers and close to the standard RM. L_1-regression with weights $w^{(2)}$ performs best, particularly for moderate outliers.

All procedures have good discriminatory power when the sequence is in the window center at positions $x_i = 50, 49, 51, 48, 52, \ldots$, see also Figure 5. Up to four outliers are dampened substantially, while the shift is preserved well from that on. The weighted filters with $n = 13$ provide improved suppression of four centric outliers, with weighted L_1 being somewhat better than the WRMs.

L_1 with weights $w^{(2)}$ offers also the best Gaussian efficiency, see Figure 4. In spite of its slightly larger bias, it is somewhat more efficient than L_1 with $w^{(1)}$, and quite a bit more than the WRMs. The WRMs with $n = 13$ are somewhat more efficient for the level than the standard L_1 and RM, which are close to each other. For the slope, the L_1 (WRM) with $w^{(2)}$ reaches about 300% (220%) of the efficiency of the standard RM. The WRMs with $n = 11$ (not shown here) are somewhat less efficient than the standard RM for the level, but somewhat more for the slope.

5 Application to time series

For further comparison we apply the filters to some time series. The simulated data depicted in Figure 6 are generated by overlaying a senoidal signal of length $N = 250$ with a shift by standard Gaussian white noise. A temporary shift of duration six is inserted at $x_i = 70$ to investigate the preservation of relevant patterns.

The online procedures are challenged by inserting irrelevant sequences of up to three outliers of size ten. In view of the results from Section 4.1, we choose widths suitable for removing $\ell = 5$ outliers, i.e. $n = 13$ for the standard RM, $n = 18$ ($n = 20$) for the WRM with weights $w^{(1)}$ ($w^{(2)}$), $n = 18$ for standard L_1, $n = 17$ ($n = 18$) for L_1 with weights $w^{(1)}$ ($w^{(2)}$), see Table 1. Accordingly, all filters resist the irrelevant outliers well, but delay the relevant shifts by five observations. As was to be expected, the L_1-filters overshoot the shifts, particularly the standard L_1. The WRMs provide less wiggly outcomes, with the WRM with weights $w^{(1)}$ performing best.

To further increase the challenge in the retrospective case, we replace up to four subsequent observations by irrelevant outliers. The standard RM and L_1 with $n = 9$, the WRMs with $n = 11$ and the weighted L_1 with $n = 13$ preserve the shift and the temporary shift well. However, the standard RM and L_1 are strongly affected by three or four subsequent outliers, and the WRMs do only slightly better. The L_1 weighted by $w^{(2)}$ performs best and is only affected by the four outliers at $x_i = 40$.

Figure 5: Retrospective application: Bias for the level due to an increasing number of observations shifted by $c = 10$ (top and bottom) and $c = 1000$ (center) starting in the center (top and center) and right in the center (bottom): RM (solid lines) and L_1-regression (dashed lines), standard version (□), weights $w^{(1)}$ (○) and $w^{(2)}$ (▽). WRM with width $n = 13$ (left) and $n = 11$ (right).

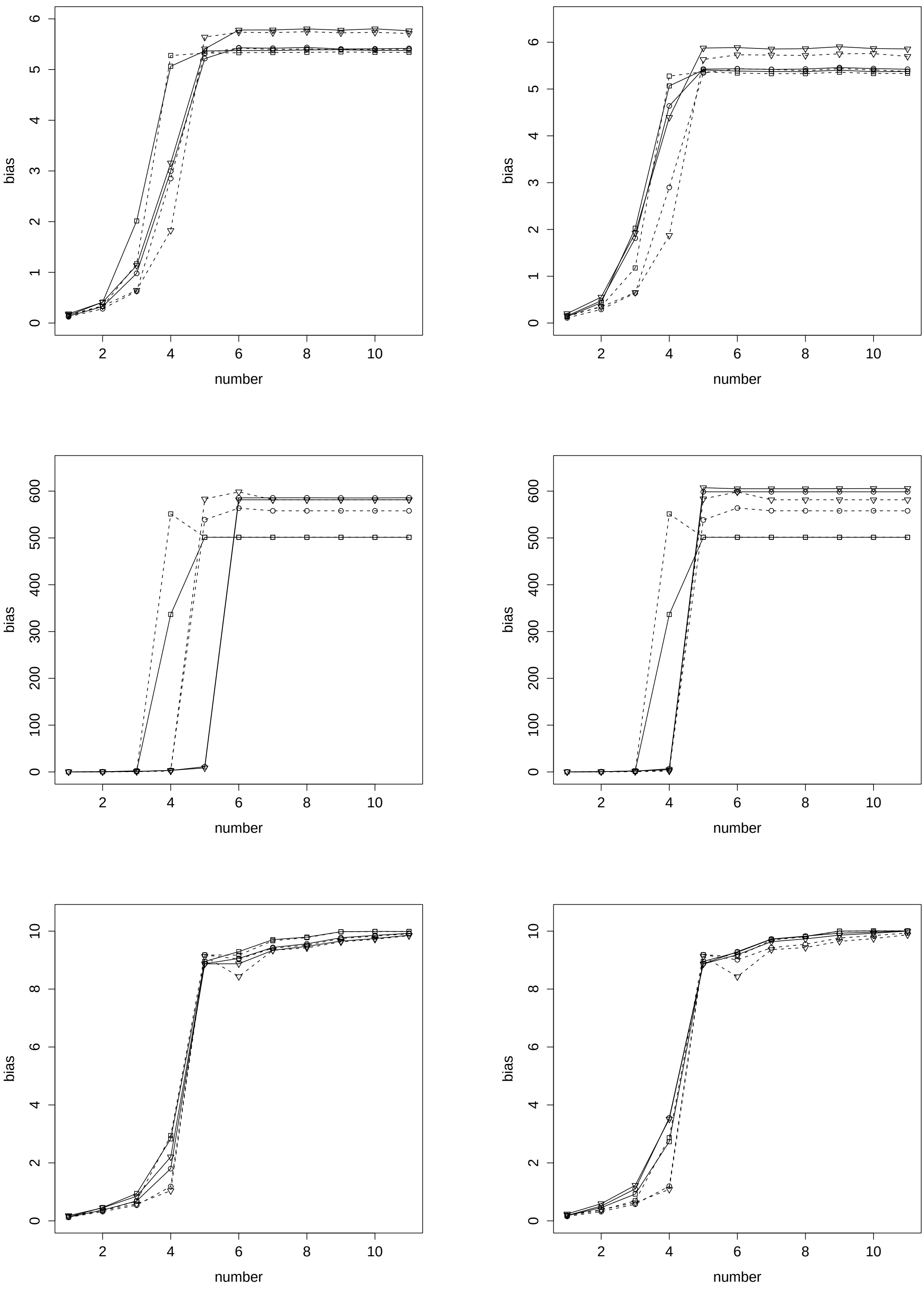


Figure 6: Online (top) and retrospective (bottom) L_1 (left) and RM (right) filtering: simulated time series $+$, underlying signal (bold dashed), standard (thin solid) and weighted version (bold solid). Weight function $w^{(1)}$ is used in the online, $w^{(2)}$ in the retrospective application.

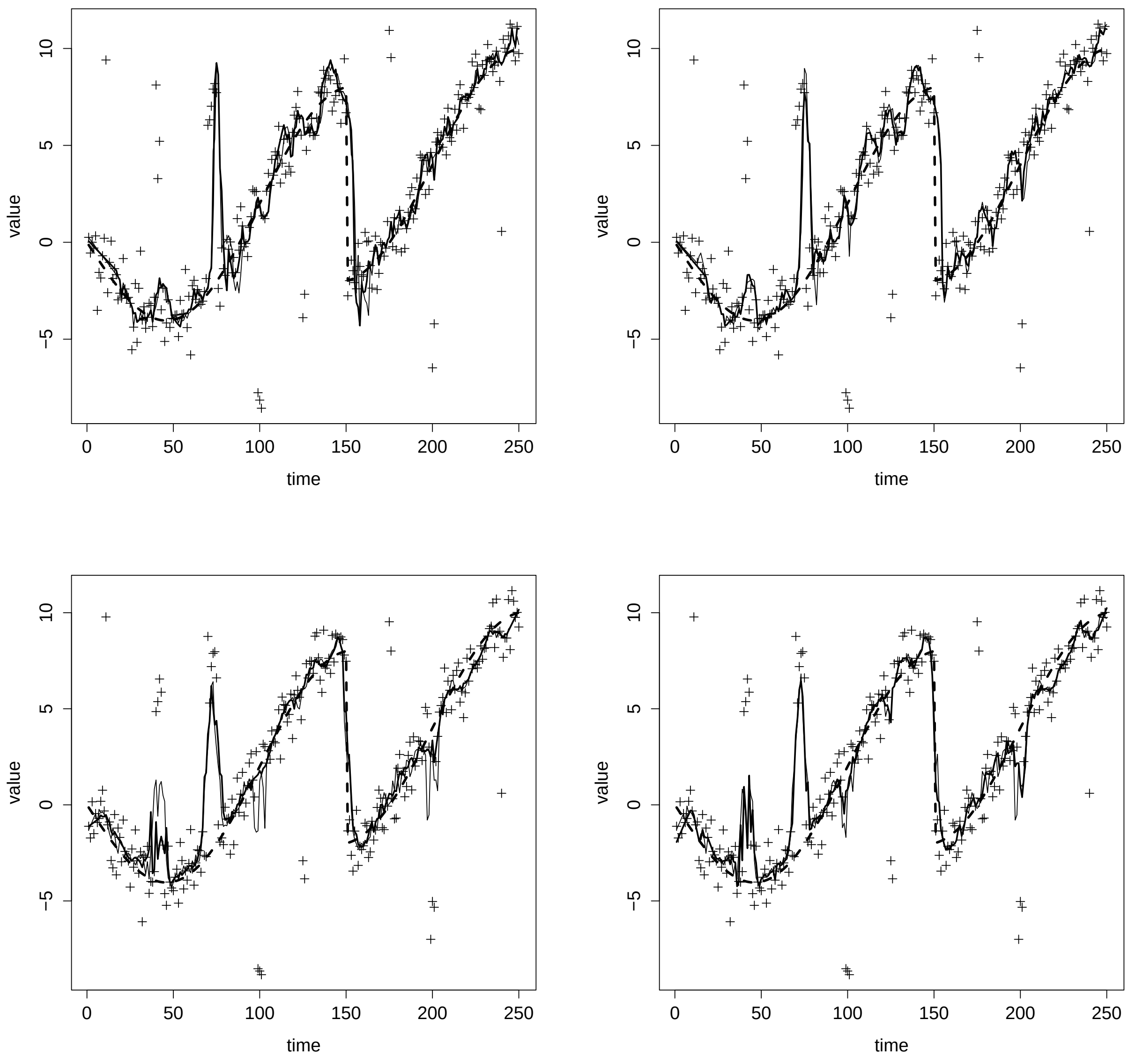
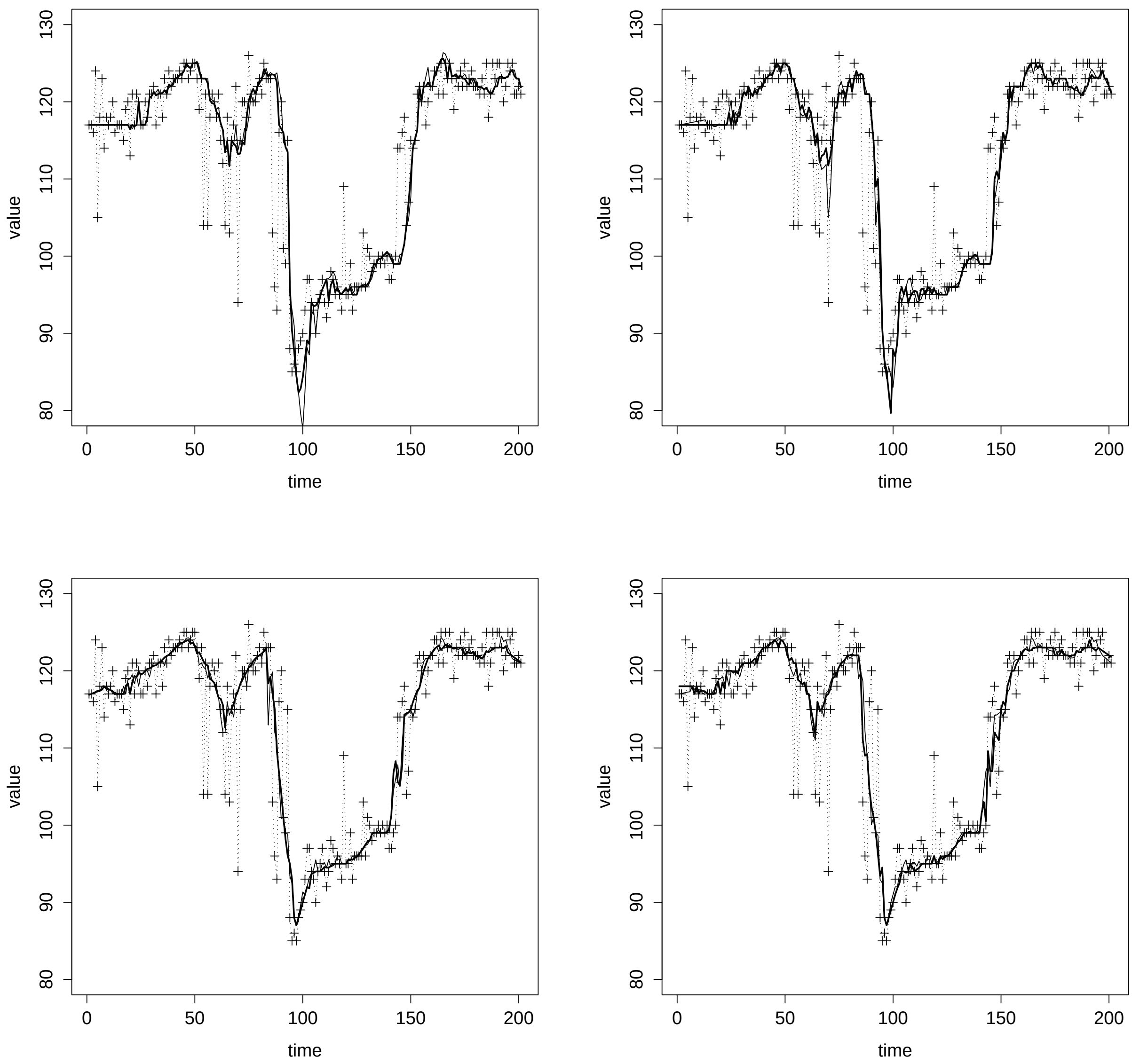


Figure 7: Online (top) and retrospective (bottom) L_1 (left) and RM (right) filtering: time series $+$, underlying signal (bold dashed), standard (thin solid) and weighted version (bold solid). Weight function $w^{(1)}$ is used in the online, $w^{(2)}$ in the retrospective application.



We also consider real data representing the arterial pressures of patients in intensive care, see Figures 7 and 8. The filters are applied using the same widths as before, but increasing n to 13 for the retrospective WRMs. The online L_1-filters again overshoot the downward shift at $x_t = 100$ in the first and the shifts in the second example, particularly the standard L_1. The standard RM is affected by some outliers occurring at about $x_t = 70$ in the first example. The online WRMs provide generally better results. In retrospective application, the standard versions are more affected by the outliers, particularly by the tripel at $x_i = 100$. The weighted versions perform again better and less wiggly.

6 Conclusions

We have investigated weighted repeated median and weighted L_1-filters for robust detail-preserving smoothing of noisy data with underlying trends. In case of the repeated median, weighting the observations according to their distance in the design space improves the local adaption to nonlinear regression functions, allows to use longer windows and increases efficiency as compared to the unweighted version, retaining the suppression of outlying spikes and the preservation of relevant shifts. Weighted repeated medians provide substantial benefits particularly in the challenging online situation. In case of L_1-regression, weighting can increase the robustness and the discrimination between sequences of relevant and irrelevant length. In retrospective application large efficiency gains are possible due to longer windows.

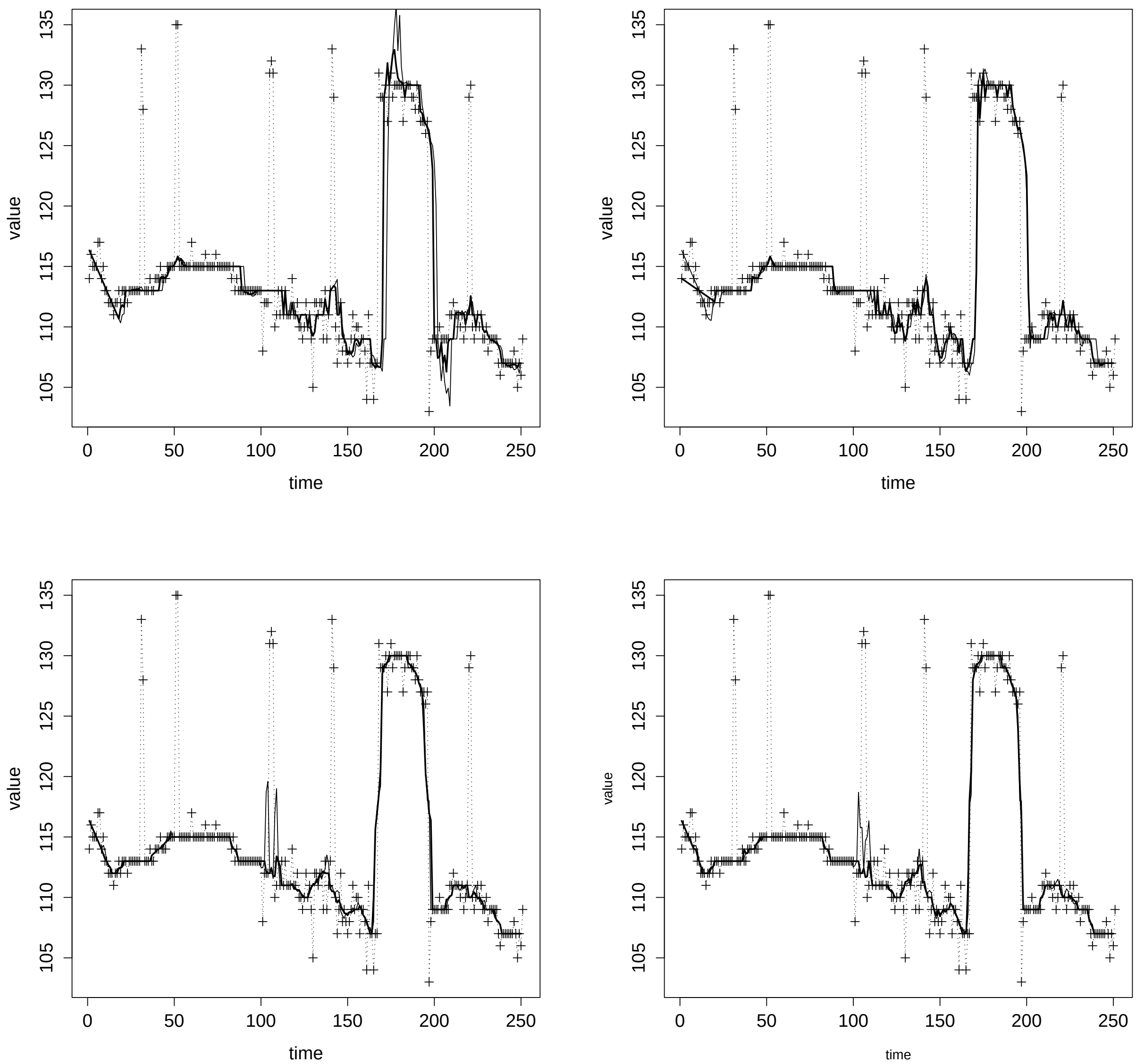
An open issue is the optimal choice of the weights under some error criterion. In general, the most suitable choice of the filtering procedure is likely to depend on the circumstances. Probably important aspects are the expected sizes of outliers and shifts as well as the curvature of the regression function. Under the criteria inspected here, the repeated median with Epanechnikov weights $w^{(1)}$ can be recommended for online, and weighted L_1-regression with $w^{(2)}$ for retrospective application.

These results rely on outlier patches being well separated. When such patches occur close to each other, using a standard repeated median with a reasonable width may still be the best decision since it can deal with the largest fraction of outliers.

Acknowledgements

The financial support of the Deutsche Forschungsgemeinschaft (SFB 475, "Reduction of complexity in multivariate data structures") is gratefully acknowledged.

Figure 8: Online (top) and retrospective (bottom) L_1 (left) and RM (right) filtering: time series +, underlying signal (bold dashed), standard (thin solid) and weighted version (bold solid). Weight function $w^{(1)}$ is used in the online, $w^{(2)}$ in the retrospective application.



References

Davies, P. L., Fried, R., and Gather, U. (2004), "Robust Signal Extraction for On-line Monitoring Data," *Journal of Statistical Planning and Inference*, 122, 65–78.

Davies, P. L., and Gather, U. (2005), "Breakdown and Groups," *Annals of Statistics*, 33, 977–1035.

Donoho, D. L., and Huber, P. J. (1983), "The Notation of Breakdown Point", in *A Festschrift for Erich Lehmann*, eds. P. J. Bickel, K. Doksum, and J. L. Hodges Jr., Belmont, CA: Wadsworth, pp. 157–184.

Edgeworth, F. Y. (1887), "A New Method of Reducing Observations Relating to Several Questions," *Phil. Mag.*, 24, 184–191.

Einbeck, J., and Kauermann, G. (2003), "Online Monitoring with Local Smoothing Methods and Adaptive Ridging", *Journal of Statistical Computation and Simulation*, 73, 913–929.

Ellis, S. P., and Morgenthaler, S. (1992), "Leverage and Breakdown in L1 Regression," *Journal of the American Statistical Association*, 87, 143–148.

Fan, J. (1992), "Design-adaptive Nonparametric Regression," *Journal of the American Statistical Association*, 87, 998–1004.

Fan, J., and Hall, P. (1994), "On Curve Estimation by Minimizing Mean Absolute Deviation and its Implications," *Annals of Statistics*, 22, 867–885.

Fan, J., Hu, T.-C., and Truong, Y. K. (1994), "Robust Nonparametric Function Estimation," *Scandinavian Journal of Statistics*, 21, 433–446.

Fried, R., Bernholt, T., and Gather, U. (2006), "Repeated Median and Hybrid Filters," *Computational Statistics and Data Analysis*, to appear.

Gather, U., Schettlinger, K., and Fried, R. (2006), "Online Signal Extraction by Robust Linear Regression," *Computational Statistics*, to appear.

Giloni, A., and Padberg, M. (2004), "The Finite Sample Breakdown Point of L1-regression," *SIAM Journal of Optimization*, 14, 1028–1042.

Härdle, W., and Gasser, T. (1984), "Robust Non-parametric Function Fitting," *Journal of the Royal Statistical Society*, Ser. B, 46, 42–51.

Hastie, T. and Loader, C. (1993), "Local Regression: Automatic Kernel Carpentry," *Statistical Science*, 8, 120–129.

He, X., Jureckova, J., Koenker, R. and Portnoy, S. (1990), "Tail Behavior of Regression Estimators and Their Breakdown Points," *Econometrica*, 58, 1195–1214.

Imhoff, M., Bauer, M., Gather, U., and Fried, R. (2002), "Pattern Detection in Intensive Care Monitoring Time Series with Autoregressive Models: Influence of the Model

Order," *Biometrical Journal*, 44, 746–761.

Koenker, R. (2005), *Quantile Regression*, Cambridge: Cambridge University Press.

Mizera, I., and Müller, Ch. H. (1999), Breakdown Points and Variation Exponents of Robust M-Estimators in Linear Models, *Annals of Statistics*, 27, 1164–1177.

Portnoy, S., and Koenker, R. (1997), The Gaussian Hare and the Laplacian Tortoise: Computability of Squared-Error versus Absolute-Error Estimators, *Statistical Science*, 12, 279–300.

Rousseeuw, P. J., and Leroy, A. (1987), *Robust Regression and Outlier Detection*, New York: Wiley.

Scholz, F.-W. (1978), "Weighted Median Regression Estimates," *Annals of Statistics*, 6, 603–609.

Siegel, A. F. (1982), "Robust Regression using Repeated Medians," *Biometrika*, 68, 242–244.

Simpson, D. G., and Yohai, V. J. (1998), "Functional Stability of One-Step GM-Estimators in Approximately Linear Regression," *Annals of Statistics*, 26, 1147–1169.

Tukey, J. W. (1977), *Exploratory Data Analysis*, Reading, Mass.: Addison-Wesley (preliminary edition 1971).

Wang, F. T., Scott, D. W. (1994), "The L_1 Method for Robust Nonparametric Regression," *Journal of the American Statistical Association*, 89, 65–76.

Yang, R., Yin, L., Gabbouj, M., Astola, J., and Neuvo, Y. (1995), "Optimal Weighted Median Filtering under Structural Constraints," *IEEE Transactions on Signal Processing*, 43, 591–604.