

Robust detail-preserving signal extraction

Ursula Gather¹, Roland Fried², Vivian Lanius^{1,1}

¹Department of Statistics
University of Dortmund
Germany

²Department of Statistics
University Carlos III de Madrid
Spain

Abstract

We discuss robust filtering procedures for signal extraction from noisy time series. Particular attention is paid to the preservation of relevant signal details like abrupt shifts. moving averages and running medians are widely used but have shortcomings when large spikes (outliers) or trends occur. Modifications like modified trimmed means and linear median hybrid filters combine advantages of both approaches, but they do not completely overcome the difficulties. Better solutions can be based on robust regression techniques, which even work in real time because of increased computational power and faster algorithms. Reviewing previous work we present filters for robust signal extraction and discuss their merits for preserving trends, abrupt shifts and local extremes as well as for the removal of outliers.

1 Introduction

Linear filters have long been the primary device for the extraction of a time-varying level (a 'signal') from time series because of the profound theory of linear systems, computational ease, simple design, and optimal attenuation of additive Gaussian noise. However, they are neither suitable if there are sudden changes from one signal level to another nor in case of impulsive noise generating strongly deviant outliers ('spikes') caused by measurement problems for instance. Change points are often the most important information and should be preserved, while at the same time a substantial amount of outliers should be resisted, since previous data-cleaning is not possible in automatic application. Tukey [89] suggests standard median filters ('running medians') for these purposes, but these still have some shortcomings as we will point out in the following.

To fix notation, we assume a simple data-generating model. Let (y_t) be a time series, observed at discrete time points $t \in \mathbb{Z}$. Of course there will be only a finite number of

¹Financial support of the Deutsche Forschungsgemeinschaft (SFB 475, "Reduction of complexity in multivariate data structures") is gratefully acknowledged.

measurements y_1, \dots, y_N available, but the main difference is that additional rules are needed for handling the endpoints. Extrapolation of the results from the first and last window or adding the first and the last observed value a sufficient number of times are possible ways to deal with finite sets of data. We assume that the data are generated as

$$y_t = \mu_t + u_t + v_t, \quad t \in \mathbb{Z}, \quad (1)$$

where the sequence (μ_t) is the signal, while u_t is 'ordinary' observational noise with constant median zero and variance σ_t^2 . Sporadic measurement problems are represented by the impulsive (spiky) noise v_t from an outlier generating mechanism. It is zero most of the time, but can take very large absolute values occasionally.

The construction of filtering procedures is usually guided by some demands. One aim is to preserve certain signal characteristics, e.g. like linear or more generally monotonic trends and abrupt, long-term level shifts. Good noise attenuation is not enough to yield acceptable signal quality. Filters with optimal noise reduction could be derived under restrictions guaranteeing detail preservation if we were willing to specify a distribution for the noise, or at least a suitably small family of distributions. However, knowledge about the noise distribution is often scarce, in particular when we are faced with measurement problems resulting in large, irrelevant spikes. Moreover, we are often confronted with other phenomena like heteroscedasticity due to time varying 'environmental' conditions. Therefore we propagate robust filters, which perform reasonably well under a broad range of conditions and do not strongly rely on a completely specified model which is most likely misspecified.

To illustrate the previous arguments we give a simple example: If we just impose that a time invariant constant signal value $\mu_t = \mu$ is to be approximated and assume the observational noise to be independently Gaussian distributed, the most efficient method in terms of the error variance is the sample mean, i.e. the arithmetic average of all available observations. However, it is well-known that the sample mean is not at all robust against deviations from normality. A simple measure of robustness is the finite sample breakdown point of an estimator, which gives us the minimal fraction of deviant observations possibly making the estimate completely meaningless [21]. It is well-known that a single outlier has an unbounded effect on the sample mean, resulting in a finite sample breakdown point of $1/N$. A possible solution are M-estimators, which achieve some robustness and large efficiency within so-called contamination neighborhoods of the Gaussian distribution Φ [48]. These neighborhoods contain all mixtures $(1 - \epsilon)\Phi + \epsilon F$ with a constant $\epsilon \in (0, 1)$ and an arbitrary distribution F . However, there is a trade-off between efficiency and robustness: designing the estimator for a larger neighborhood increases robustness, but reduces the efficiency at the Gaussian. The median finally is the Huber M-estimator with maximal asymptotic breakdown point 50%, guaranteeing optimal protection among all reasonable location estimators: about half of the sample needs to be contaminated for the effects to become arbitrarily large. We will focus on methods with high breakdown points.

Instead of optimizing a single criterion, statistical procedures intended to deal with real world data should behave well in many different aspects [17]. Common criteria in routine application are the existence of a unique solution, low computation time, the preservation of important signal details, high robustness against outliers and satisfactory finite sample efficiency under Gaussian or other prototype distributions. We restrict ourselves to filters fulfilling the first two demands, and compare candidate methods w.r.t. the latter three properties.

In the general situation of a time-varying signal there are different approaches to filter construction: Recursive filters update the estimate for the previous time point including the information from the incoming observation. Exponentially weighted moving averages (EWMA) are perhaps the most common example. These filters are designed for sequential ('online') application, where one approximates the signal value at the most recent time point without delay. The resulting estimates are optimal w.r.t. a weighted least squares loss and very vulnerable to outliers. Robustifications based e.g. on weighted least absolute deviations are possible, but computationally expensive and their statistical properties are difficult to analyze [16]. Recursive filters like EWMA tend to follow changes like abrupt level shifts or monotonic trends with some delay since they only include past observations. A further major difficulty is to construct filters which preserve fine signal details like temporary shifts, while removing short sequences of irrelevant outliers.

Moving window techniques slide a time window through the series for local approximation of the signal from the data in the window. moving averages and running medians are prominent representatives of such filters. However, moving averages and linear filters in general are not suitable for removing outliers and they always blur level shifts (also called 'step changes' or 'jumps'), see Fig. 1.

Moving window techniques can be designed for retrospective (fixed sample) or online (sequential) application. In retrospective application, a time delay does not cause a problem. Here one approximates the signal value in the center of the window, including both past and future observations in the calculations. In online analysis one approximates the signal at the most recent time point, i.e. at the end of the window. To unify notation, we denote the time window used for approximation of the signal value μ_t at time t by $y_{t-m}, \dots, y_{t+\tilde{m}}$, where $\tilde{m} = m$ in the symmetric retrospective and $\tilde{m} = 0$ in the online situation. For determination of the window width $n = m + \tilde{m} + 1$ we need to choose a suitable value of m .

We discuss moving window techniques which allow to preserve relevant signal details like level shifts and provide considerable robustness against deviations from the modelling assumptions, particularly against outlying spikes. For distinguishing between relevant temporary level shifts and irrelevant sequences of spikes we assume the latter to have shorter durations. The filter can be designed accordingly by choosing appropriate substructures and window widths. For more extensive reviews of (robust) non-linear filters see [5, 33, 53, 74, 75].

This chapter is organized as follows: Section 2 illustrates robust detail-preserving signal extraction using location-based filters like running medians. Section 3 proposes regression-based procedures which achieve large improvements in trend periods. Section 4 presents ideas for modification and combination of the filters studied before. Section 5 draws some conclusions.

2 Filters based on local constant fits

Location-based filters apply a location estimator for the approximation of the signal value μ_t from $y_{t-m}, \dots, y_{t+\tilde{m}}$. Such methods implicitly assume the signal to be almost constant within each time window, i.e. $\mu_{t-m} \approx \dots \approx \mu_t \approx \dots \approx \mu_{t+\tilde{m}}$ for all t . This assumption can be justified when choosing m small since the signal is assumed to vary slowly, but the cost is reduced smoothing. Generally, the window width needs to be chosen by a compromise

between several aims: On the one hand, the assumption of a constant level within each window is less reasonable for large m . This causes problems particularly in the online situation as we rely on a simple extrapolation then. On the other hand, a large width stands for smaller variability, produces smoother estimates and increases robustness.

2.1 Standard median filters

Standard median filters, also called running medians, have been introduced by Tukey [89] and are perhaps the most prominent robust location-based filters. They approximate μ_t by the median of $y_{t-m}, \dots, y_{t+\tilde{m}}$,

$$StM(y_t) = \tilde{\mu}_t = \text{med}(y_{t-m}, \dots, y_{t+\tilde{m}}), t \in \mathbb{Z}.$$

Like all filters based on 'reasonable' location estimators, standard median filters are location- and scale-equivariant, meaning that adding a constant or multiplying by a constant changes the filter output in the same way. The quality of filters with these properties does hence not depend on the underlying measurement scale.

The asymptotic variance of the median is $1/(4nf^2(0))$ if the noise has a density f with median zero. Accordingly, its asymptotic efficiency relatively to the mean is 63.7% for the Gaussian, but 200% for the Laplace distribution.

The finite sample breakdown point of the median applied to n observations is $\lfloor (n+1)/2 \rfloor / n$, where $\lfloor c \rfloor$ represents the largest integer not larger than c . This means that at least half of the data needs to be shifted to completely change the estimate. This property can be used for designing running medians: To remove sequences of up to ℓ outliers and preserve level shifts with a duration of at least $\ell+1$ observations, we can apply a running median with window width $n = 2\ell + 1$.

The exact fit point provides information on the preservation of relevant signal details and the removal of spikes under idealized conditions with no observational noise, i.e. $\sigma_t^2 \equiv 0$. Applied to a regression functional $T : \mathbb{R}^n \rightarrow \mathbb{R}^p$, the exact fit point corresponds to the smallest possible fraction of contamination which can cause T to deviate from a fit $\tilde{\gamma} \in \mathbb{R}^p$. Consider a sample $\mathbf{y}_n = \{(x_1, y_1), \dots, (x_n, y_n)\}$ of n observations of a response y and a p -variate regressor x such that $y_i = \tilde{\gamma}'x_i$ for all $i = 1, \dots, n$, and let $\mathbf{y}_{k,n}$ be a sample where k out of the n observations in \mathbf{y}_n are replaced by arbitrary values. The exact fit point of T then becomes

$$\delta_n^*(T, \mathbf{y}_n) = \min_k \left\{ \frac{k}{n} \mid \text{there exists a sample } \mathbf{y}_{k,n} \text{ such that } T(\mathbf{y}_{k,n}) \neq \tilde{\gamma} \right\}.$$

The median, like all location estimators, regresses on a constant only, i.e. $p = 1$ and $x_i = 1, i = 1, \dots, n$. Its exact fit point is equal to its finite-sample breakdown point. While the latter corresponds to the minimal number of spikes which can render the extracted value meaningless, the former yields the number of spikes a filter can remove completely in the absence of observational noise. A running median with width $n = 2\ell + 1$ can hence remove up to ℓ subsequent spikes completely if $\sigma_t^2 \equiv 0$. In retrospective application, it can preserve a level shift from one constant signal value to another one exactly if it lasts at least for $\ell+1$ observations, while in online application the shift is delayed by ℓ observations. Another notable property

of the running median in retrospective application is that it recovers monotonic trends exactly under noise-free conditions.

The exact preservation of signal characteristics as described above applies only under idealized conditions. Nevertheless, the deviations can be expected to be small in the presence of little (as compared e.g. to the height of shifts) observational noise since the median is Lipschitz-continuous with constant 1: The median deviates at most by δ from μ_t if for all i $|u_{t+i}| < \delta$ and if not more than one of the following occurs in the window: at most ℓ spikes,, a single level shift, or a monotonic trend. Lipschitz-continuous functionals are to be recommended in general since this property restricts the influence of minor changes in the data due to small observational noise or rounding [17].

Nevertheless, the performance of running medians becomes worse at monotonic changes (edges): it suppresses noise less efficiently there, and it shows a bias which is related to the noise power and the height of the edge. Further problems arise when more than one data pattern occurs in a single window: Running medians suffer from edge jitter, i.e. they move shifts towards preceding close-by spikes into the same direction. A shift during a monotonic trend can be preserved only if the shift and the trend point into the same direction. The shift gets blurred otherwise, and a single spike within a trend causes smearing [12, 9, 3, 68, 29]. Median filters with an adaptive window width have been suggested to reduce edge and plateau jitter caused by spikes close to edges [76, 62, 54]. The window width can be chosen using criteria like the current signal slope [54], the length of detected outlier sequences [62, 50], or a variance decomposition assuming the noise variance to be stationary [84, 64].

When designing a filter we often want certain signals to pass the filter unperturbed. For a linear filter, such eigenfunctions can be characterized in the frequency domain by its passband and stopband. Signals which pass a nonlinear filter unchanged are called its roots and can be analyzed in the time domain. The roots of a running median with width $n = 2\ell + 1$ contain only edges of monotonic increase or decrease, separated by at least $\ell + 1$ constant values [34]. Thus, the roots of a running median are also roots of all running medians with smaller width. A running median reduces any time series within a finite number of repetitions to one of its roots.

Recursive medians are a simple variation of running medians, replacing the observations before time t by the already filtered values when calculating the output at t . A recursive median possesses the same set of roots as a running median with the same window width, but a time series may be filtered to different roots by the two filters. Recursive medians reduce every series to become a root in a single step; they provide better smoothing and they are more robust than running medians, but they distort edges more strongly [93, 55].

2.2 Modified order statistic filters

Instead of the median, other order statistics (OS) can be applied for filtering as well. Switching to a higher or lower order statistic can improve the preservation of shifts [85]. More generally, OS-Filters, or L -filters, are based on linear combinations of order statistics [10]. Using a set of weights w_1, \dots, w_n summing up to 1, the filter output is calculated as

$$OS(y_t) = \sum_{i=1}^n w_i y_{t(i)}, \quad (2)$$

where $y_{t(1)}, \dots, y_{t(n)}$ are the ordered observations within the window. A suitable choice of the weights allows to dampen noise with different tail behavior efficiently [10, 41]. Order statistic filters are location- and scale-equivariant. They preserve linear trends exactly in retrospective application and under noise-free conditions if the weights are chosen symmetric, $w_i = w_{n-i+1}$, $i = 1, \dots, n$. As special cases we obtain the mean ($w_i = 1/n$ for $i = 1, \dots, n$), the median ($w_{\ell+1} = 1$ and $w_i = 0$ otherwise for odd $n = 2\ell + 1$), the midpoint ($w_1 = w_n = 1/2$ and all other $w_i = 0$) and the α -trimmed means ($w_i = 1/(n - 2\lfloor \alpha n \rfloor)$ for $i = \lfloor \alpha n \rfloor + 1, \dots, n - \lfloor \alpha n \rfloor$ and $w_i = 0$ otherwise). Order statistic filters with nearly minimal MSE for a given error distribution can be designed using analytical approximations even in real time application [70].

α -trimmed means (α -TM) have received considerable attention since they constitute a compromise between the mean ($\alpha = 0$) and the median ($\alpha = 0.5$). Often $\alpha \in [0.2, 0.275]$ is suggested to yield good efficiency for a broad family of distributions including the Gaussian [63, 28], i.e. we trim between 20% and 27.5% of the smallest and the largest observations. The price to be paid for increased efficiency close to the Gaussian as compared to the median is a smaller resistance to outliers: The breakdown point of an α -TM is asymptotically $2\alpha \cdot 100\%$. Accordingly, an α -TM-filter with $\alpha < 0.5$ smoothes a level shift to a ramp edge with $(1 - \alpha)n$ observations [6, 59, 73, 77]. More generally, running medians are the only order statistic filters which can preserve shifts exactly [63].

Order statistic filters with data-adaptive choice of the weights have been suggested to overcome this deficiency. They achieve considerable robustness against outliers and at the same time high efficiency under a broad range of conditions including time-varying, heterogeneous noise. Modified trimmed mean (MTM)-filters are defined in analogy to trimmed means, but they choose the fraction of trimming α depending on the data in the current window. Observations which are further away than a distance q_t from the local median are trimmed and the average of the remaining observations is taken as filter output:

$$\begin{aligned} MTM(y_t) &= \frac{1}{|I_t|} \sum_{i \in I_t} y_{t+i}, \\ I_t &= \{i = -m, \dots, \tilde{m} : |y_{t+i} - \tilde{\mu}_t| \leq q_t\} \\ \tilde{\mu}_t &= \text{med}(y_{t-m}, \dots, y_{t+\tilde{m}}), \quad t \in \mathbb{Z}. \end{aligned} \quad (3)$$

Hence, MTM-filters are a data-adaptive compromise between the running median ($q_t = 0$) and the moving average ($q_t = \infty$), compare also Fig. 1. An a-priori choice of q_t can be based on the expected height of the shifts. A data-adaptive alternative can be formulated using a robust scale estimate like the local median absolute deviation about the median (MAD),

$$\tilde{\sigma}_t^M = c_n \cdot \text{med}(|y_{t-m} - \tilde{\mu}_t|, \dots, |y_{t+\tilde{m}} - \tilde{\mu}_t|).$$

Here, c_n is a correction factor depending on the window width n , usually chosen to achieve unbiasedness in case of Gaussian noise. For n not very small we set $c_n = 1.483$. A reasonable range of choices is $q_t \in [2\tilde{\sigma}_t^M, 3\tilde{\sigma}_t^M]$, see [59, 44].

Double window modified trimmed mean (DWMTM)-filters are a variant of MTM-filters. They apply two windows with different widths. The median and the MAD are calculated from a short signal window with width $k < n$ to retain signal details. Then all observations deviating more than q_t from this median are trimmed from the larger window with width n ,

before the remaining values are averaged for better attenuation of observational noise. MTM-filters can be seen as DWMTM-filters with $k = n$. DWMTM-filters with adaptive choice of that factor, by which the local MAD is multiplied, have been suggested for removing signal-dependent noise [19].

Analysing the breakdown and exact fit points shows that a DWMTM can remove up to $\lfloor k/2 \rfloor$ subsequent spikes from a constant signal in noise-free conditions. The smaller window width k should hence be chosen depending on the minimal duration of relevant signal details. Using a short inner window improves the preservation of shifts, see Fig. 1, but reduces the attenuation of noise. A DWMTM-filter can be tuned to be considerably more efficient for Gaussian noise and preserve large shifts better than a running median with the same n choosing k and q_t large enough [44, 75].

DWMTM-filters calculate the mean of a subsample which is chosen according to the distances from an initial estimate. This means a kind of hard-thresholding and implies discontinuity. Soft thresholding with a smooth transition between acceptance and rejection can be obtained in form of weighted averages with weights chosen according to the initial distances [76]. This idea will be explained in more detail at the end of the next section.

Further location estimators have been proposed for filtering. The Hodges-Lehman-Bickel (HLB) estimate of location is the median of averages of symmetrically placed order statistics,

$$HLB_n(y_{t-m}, \dots, y_{t+\tilde{m}}) = \text{med}_{i=1, \dots, \lfloor (n+1)/2 \rfloor} \frac{y_{t(i)} + y_{t(n-i+1)}}{2}.$$

It is location- and scale-equivariant, preserves trends in the absence of noise, has a breakdown point of 25% Nevertheless, application of suitably trimmed means seems preferable [28].

2.3 Weighted median filters

The standard median and (modified) order statistic filters defined in the previous sections do not take into account the temporal distances between the target point t at which we estimate the signal and the observation times of the measurements included in the calculation. This causes problems if the implicit assumption of a locally constant signal within each window is not fulfilled. A remedy is to weight the observations according to their temporal distances [51], giving smaller weight to observations more distant from the target point.

We focus on weighted median (WM)-filters: while the median minimizes the L_1 -distance (the sum of the absolute deviations) to the data points, the weighted median of $y_{t-m}, \dots, y_{t+\tilde{m}}$ for arbitrary positive real weights $w_{-m}, \dots, w_{\tilde{m}}$ minimizes the weighted L_1 -distance

$$WM(y_t) = \underset{\mu}{\text{argmin}} \sum_{i=-m}^{\tilde{m}} w_i \cdot |y_{t+i} - \mu|. \quad (4)$$

running medians correspond to uniform weights $w_i = 1$, $i = -m, \dots, \tilde{m}$. WM-filters have become popular because of their high flexibility: a running median necessarily applies a window of width $n = 2\ell + 1$ to preserve signal details of length $\ell + 1$ and to remove up to ℓ outlying spikes. Weighting of the observations allows to use longer windows and thus yields a stronger noise reduction [51, 93].

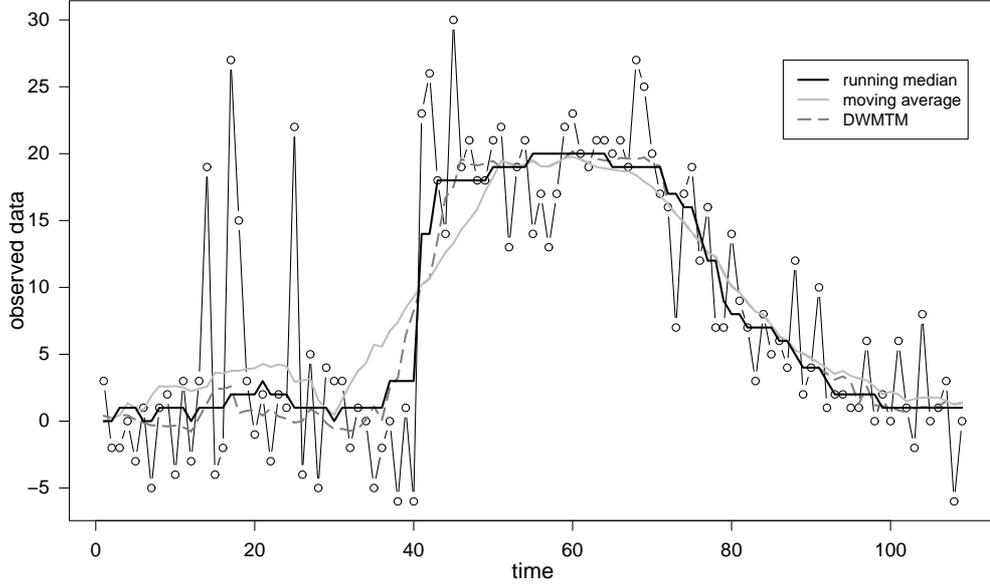


Figure 1: Observed time series with outliers, level shift and trend together with approximations of the signal by means of location-based filters: moving average, median and DWMTM-filter with $m = \tilde{m} = 10$, i.e. $n = 21$, and $k = 9$. In this example the median performs very well; however, it approximates the linear trend by a step function.

Denoting the ordered observations in the window by $y_{t(1)} \leq \dots \leq y_{t(n)}$ and the corresponding positive weights by $w_{(1)}, \dots, w_{(n)}$, the weighted median corresponds to the k -th order statistic $\hat{\mu} = y_{t(k)}$, where

$$k = \max \left\{ h : \sum_{i=h}^n w_{(i)} \geq \frac{1}{2} \sum_{i=1}^n w_{(i)} \right\}. \quad (5)$$

For example, the WM of 1, 2, 3, 9 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, (5) and (4) yield the same results. However, the whole interval $[y_{(k-1)}, y_{(k)}]$ solves (4) 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 (5) by summing from the bottom instead of from the top and taking the minimum instead of the maximum. This ambiguity can be solved as usual by choosing the center of the interval (the only choice which gives affine equivariance). For non-negative integer valued weights w_1, \dots, w_n , a simple equivalent representation of the weighted median of $y_{t-m}, \dots, y_{t+\tilde{m}}$ is

$$WM(y_t) = \text{med}(w_{-m} \diamond y_{t-m}, \dots, w_{\tilde{m}} \diamond y_{t+\tilde{m}}), \quad (6)$$

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

Even though there is an infinite number of real weights, there is only a finite number of WM-filters for a given window width. In particular, for every WM with arbitrary positive real weights there is an equivalent WM with integer weights [69]. Two weighted medians with respective weights w_1, \dots, w_n and $\tilde{w}_1, \dots, \tilde{w}_n$ are called equivalent iff they give the same result for every sample. This is the case iff for every subset $I \subset \{1, \dots, n\}$ of indices we have

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

In particular, to get equivalence to the standard median it is crucial 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. For an overview on the equivalence of WMs see [72].

WM-filters are unbiased for the mean in case of symmetric noise. Formula for output central moments and the variance of WM-filters can be found in [91], as well as an algorithm to obtain WM-filters which minimize symmetrically distributed noise under the constraint that certain signal details are to be preserved in noise-free conditions. The optimal WM-filter does not depend on the underlying error distribution, and it is optimal both under the mean squared error (MSE) and under the mean absolute error (MAE) criterion. In the absence of structural constraints, the WM-filter with minimal MAE and MSE for a given window width is the running median.

Root signal properties of general WM-filters are much more difficult to derive than those of running medians [93]. Weighted median filters are basically low-pass filters, like the other filter classes treated here. The frequency response of selection type non-linear filters like WMs can be analysed by comparison with a linear filter having the sample selection probabilities as coefficients [65, 47]. Weighted median filters can be used for high-pass and band-pass filtering by allowing for negative weights [2] or by linear combination of several weighted medians [15]. Ideas for robust periodograms and robust short-time Fourier transforms based on M-estimators in general and medians in particular can be found in [20]. For similarities between WM-filters and linear filters with finite impulse response (FIR) see [93].

Weighting according to the temporal distances can of course also be applied to location estimators different from the median. DWMTM-filters with additional weighting according to the temporal order of the observations can retain desired signal frequencies in addition to edge preservation and impulse suppression [36]. Again we can also apply the soft-thresholding described at the end of Subsection 2.2 [26]. Let $w_{-m}, \dots, w_{\tilde{m}}$ be weights according to the temporal distances in the design space as before. Further, additional weights for the distances in the observation space are derived using an unimodal affinity function A , which is controlled by initial robust estimates of location and spread, e.g. the median and the MAD. Then the resulting weighted order statistic (WOS) affine FIR filter reads

$$WAF(y_t) = \frac{\sum_{i=-m}^{\tilde{m}} w_i A_i^{\mu, \gamma} y_{t+i}}{\sum_{i=-m}^{\tilde{m}} w_i A_i^{\mu, \gamma}}.$$

Filters defined like this are data-adaptive and location-equivariant, and they can preserve trends and shifts exactly in noise-free conditions. For the preservation of shifts and the suppression of spikes the affinity function needs to decay sufficiently fast to zero.

Very general filter classes have been derived by linear combination of all order statistics with weighting according to both the temporal and the rank order [71, 36], but these are difficult to design except for multiplicative weights. Similarly, generalized Wilcoxon filters can be constructed combining linear rank statistics and temporal weighting, but they seem to be inferior to DWMTM-filters both w.r.t. edge preservation and noise attenuation [35].

3 Filters based on local linear fits

Location-based filters like those discussed before have difficulties in trend periods since the assumption of a local constant level is only appropriate when using very short time windows. These filters lose both efficiency and robustness in trend periods. Neither can they preserve arbitrary shifts during trends, nor can they remove spikes completely, not even under idealized conditions. Only DWMTMs can keep their good properties during trends if the inner window is sufficiently short [18, 8, 29].

It suggests itself that local linear fits are preferable to local constant fits [25], as they improve the approximation. In the context of time series filtering this means that we assume the data in a moving time window to be locally well approximated by a linear trend, $\mu_{t+i} = \mu_t + i\beta_t$, $i = -m, \dots, \tilde{m}$. For estimation of the level μ_t and the slope β_t at time t we can apply robust linear regression to fit this local model, see also Fig. 2. In addition to the location- and scale-equivariance of location-based filters, a filter thus obtained offers invariance to (linear) trends [29] when using a regression-equivariant functional. This property guarantees that the quality of signal extraction does not depend on an underlying local linear trend. When varying the trend in the window, i.e. replacing $y_{t-m}, \dots, y_{t+\tilde{m}}$ by $y_{t-m} - mc, \dots, y_{t-1} - c, y_t, y_{t+1} + c, \dots, y_{t+\tilde{m}} + \tilde{m}c$, the level estimate at time t remains the same, while the slope estimate increases by c .

3.1 Filters based on robust regression

Contrary to the median for robust estimation of location, no generally accepted unique standard exists for robust linear regression. Comparisons of common robust regression techniques in the retrospective and in the online situation, respectively, can be found in [18, 39].

Like the median, standard L_1 -regression minimizes the least absolute deviations (LAD)

$$(\hat{\mu}_t^{L1}, \hat{\beta}_t^{L1}) = \operatorname{argmin}\{(\mu, \beta) : \sum_{i=-m}^{\tilde{m}} |y_{t+i} - \mu - \beta(t+i)|\}. \quad (7)$$

The hierarchical repeated median (RM) [83] at the target point t is

$$\begin{aligned} \hat{\beta}_t^{RM} &= \operatorname{med}_{j=-m, \dots, \tilde{m}} \left(\operatorname{med}_{i \neq j} \frac{y_{t+i} - y_{t+j}}{i - j} \right), \\ \hat{\mu}_t^{RM} &= \operatorname{med} \left(y_{t-m} + m\hat{\beta}_t^{RM}, \dots, y_{t+\tilde{m}} - \tilde{m}\hat{\beta}_t^{RM} \right). \end{aligned} \quad (8)$$

The RM firstly calculates a slope estimate $\hat{\beta}_t^{RM}$ by taking repeated medians of all pairwise slopes in the window, and then a level estimate $\hat{\mu}_t^{RM}$ as the median of the trend-corrected observations. It has turned out to outperform standard L_1 -regression in most respects.

The Hampel-Rousseeuw least median of squares (LMS) [40, 79] minimizes the median of the squared distances,

$$(\hat{\mu}_t^{LMS}, \hat{\beta}_t^{LMS}) = \operatorname{argmin}\{(\mu, \beta) : \operatorname{med}_{i=-m, \dots, \tilde{m}}(y_{t+i} - \mu - \beta(t+i))^2\}. \quad (9)$$

A generalization is the least quantile of squares (LQS), replacing the median by another quantile.

The RM and the LMS both possess the maximal breakdown point $\lfloor n/2 \rfloor / n$ for regression-equivariant estimators calculated from a sample of size n . This implies the same asymptotic 50% breakdown point as for the standard median. The breakdown point of L_1 -regression is smaller than this and asymptotically not larger than 25% in case of an equidistant design like in time series filtering [24].

For regression- and scale-equivariant functionals, the exact fit point is never smaller than the finite-sample breakdown point [82]. In the case of linear regression, an exact fit point of k/n means that whenever $y_{t+i} = \tilde{\mu} + \tilde{\beta}i$ fits at least $n - k$ of the n observations exactly, then the estimate becomes $(\tilde{\mu}, \tilde{\beta})$ whatever the other k observations are. The exact fit point of the LMS is $\lceil n/2 \rceil / n$, see [82], while for the RM it is $\lfloor n/2 \rfloor / n$, i.e. one less observation is needed to pull the fit away if the sample size is odd.

The RM and the LMS have the same breakdown point, but the LMS better resists many large outliers as even almost 50% outliers of any size do not cause it to be strongly biased. Accordingly, it is able to preserve a level shift almost exactly in retrospective application. The strong negative bias of the corresponding scale estimate can be used to determine an LQS adaptively by comparison with the residual standard deviation [60]. To its disadvantages belongs its computational complexity of order n^2 [22], yielding computation times rapidly increasing with the window width. Besides, the LMS filter output is very wiggly since it is not continuous and its Gaussian efficiency is less than 25% in small samples, and even decreasing in n .

In spite of the benefits of the LMS, the repeated median can be recommended both for retrospective and online application [18, 39]. It offers almost the same Gaussian efficiency of about 65% as the standard median, but independently of the underlying slope, Lipschitz-continuity implying stability in case of small changes in the data, and reasonable robustness as it resists well up to about 30% outliers in a single window. It is computationally faster than the LMS, particularly so since a fast update algorithm is available which allows calculating the next filter value in linear time when moving the window forward [7]. Its main disadvantage consists probably in increased smoothing in case of a level shift, see Fig. 2. A common phenomenon of local linear fits in online application is an overshoot of the new signal value after a shift [23]. The RM performs considerably better than the LMS in this respect. Further improvements retaining the robustness can be achieved by an adaptive choice of the window width based on residual sign tests [38].

Least trimmed squares (LTS) regression [78] can be seen as a modification of the LMS and has also been suggested for filtering purposes [56]. The LTS has better asymptotic properties than the LMS, especially a non-zero Gaussian efficiency, but it is computationally even more expensive and performs similarly to the LMS in finite samples [39]. In the same way, no significant advantages of deepest regression [81] have been found as compared to the RM.

3.2 Modified repeated median filters

In analogy to the modified trimmed mean filters defined in Section 2.2, we can fit a least squares regression line, trimming or more generally down-weighting observations with large residuals in a preliminary robust regression step. This allows to retain the breakdown point of the initial estimate when giving observations with huge residuals zero weight. Reweighted least squares (RLS) based on an initial LMS fit is popular for robust regression since it increases the Gaussian efficiency of the LMS considerably, but RLS can be unstable like the LMS because of its inherent lack of continuity.

Trimmed repeated median (TRM)-filters suggested in [8] use the RM in the initial step and apply least squares to the trimmed observations in a second step. A suitable trimming constant q_t can be obtained by estimating the variability about the RM regression line e.g. by the MAD of the regression residuals [37].

Since TRM-filters apply regression-equivariant functionals in both steps, they are not only location- and scale-equivariant, but like RM-filters also trend-invariant. Instabilities have not been observed, although RM-filters are not Lipschitz-continuous because of the hard thresholding [8]. A TRM-filter can be substantially more efficient than the RM with the same width n , depending on the amount of trimming. Choosing q_t as three (two) times the MAD yields e.g. Gaussian efficiency of 92% (76%). TRM-filtering is computationally feasible since an update algorithm can be applied for the initial RM [7].

Double window filters with a shorter inner window width $k < n$ in the initial step improve the preservation of signal details, especially of abrupt shifts, see Fig. 2. The choice of k should depend on the length of outlier patches the filter should cope with: up to $\lfloor k/2 \rfloor - 1$ outliers in the inner window can be resisted before the output can be completely wrong, according to the breakdown point of the initial RM. In practice, k should even be chosen about three times the length of outlier patches to be removed since one third outliers can have a big, though limited influence on the RM. The benefits obtained in case of a level shift increase with the difference $n - k$ between the outer and the inner window width.

3.3 Weighted repeated median filters

Application of a regression instead of a location estimator to the data in a moving window (implicitly) replaces the assumption of a locally constant level by that of a locally constant slope. Using ideas similar to those underlying weighted medians, we can weight the observations according to their temporal distances. Doing so we aim at increasing the window width of standard robust regression filters, without increasing the bias when the signal slope is time-varying.

Weighted repeated median (WRM)-filters and weighted L_1 -filters for detail-preserving robust filtering are investigated in [30]. Weighting reduces the breakdown point of the repeated median, while it can increase that of L_1 -regression when downweighting observations far away from the target point t . The breakdown point can be further increased when confining to an approximative weighted L_1 -solution: Starting from a high breakdown fit like the standard RM, we can iterate a finite number of steps between maximization w.r.t. μ given β and vice versa. In case of standard L_1 -regression, this increases the breakdown point asymptotically to $1 - 1/\sqrt{2} \approx 0.293$.

The WRM in combination with so called Epanechnikov weights $w_i^{(1)} = 1 - [|i|/(m+1)]^2$, $i = -m, \dots, 0$, is well-adapted for online application, while L_1 -regression with weights $w_i^{(2)} = (1 + |i - t|)^{-1/2}$, $i = -m, \dots, m$ performs even better in the retrospective case. Similar to weighted medians, the weighting in combination with the possible longer window widths increases considerably the Gaussian efficiency of these filters in the respective situation.

A simple WRM designed for preserving level shifts in retrospective application uses a shorter window for the initial slope estimation, applying uniform weights [8]:

$$\begin{aligned} DWRM(y_t) &= \text{med}(y_{t-m} + m\hat{\beta}_t, \dots, y_{t+m} - m\hat{\beta}_t) \\ \hat{\beta}_t &= \text{med}_{i=-h, \dots, h} \left(\text{med}_{j=-h, \dots, h, j \neq i} \frac{y_{t+i} - y_{t+j}}{i - j} \right). \end{aligned} \quad (10)$$

The DWRM-slope is little affected until the shift intrudes into the inner window, resisting a shift almost as good as a standard median in case of a constant signal if $h \ll m$. It is almost as efficient as the median with the same width $2m + 1$ in case of a constant signal, but it is trend-invariant like all WRMs. Different from the double window filters presented in Section 3.2, the DWRM is Lipschitz-continuous with constant $2h + 1$.

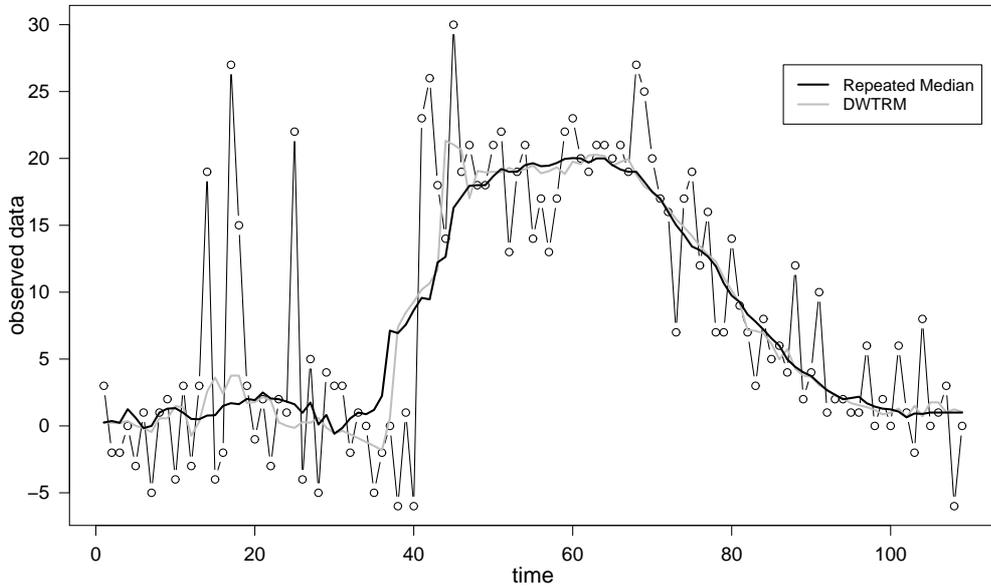


Figure 2: Observed time series with outliers, level shift and trend together with approximations of the signal by means of filters based on local linear fits: simple repeated median filter and DWTRM-filter with $m = \hat{m} = 10$, i.e. $n = 21$, and $k = 9$.

4 Modifications for better preservation of shifts

A major disadvantage of the previous filters based on the repeated median is the smearing of level shifts, which is stronger than for median based filters in case of a constant signal. Double window filters reduce this effect, but they do not solve the problem completely. In the following we present some possibilities for further improvement.

4.1 Linear median hybrid filters

Linear median hybrid (LMH)-filters take the median value of linear subfilters Φ_1, \dots, Φ_M as the filter output [42, 43, 4]. When all subfilters give non-zero weight to only a finite number of observations, the resulting procedure is called linear median hybrid filter with finite impulse response, briefly FMH-filter:

$$FMH(y_t) = \text{med}(\Phi_1(y_t), \Phi_2(y_t), \dots, \Phi_M(y_t)), \quad t \in \mathbb{Z}.$$

The linear subfilters used for preprocessing reduce the computational costs as compared to a running median with the same width, and they provide increased flexibility due to the many choices possible. They can be designed to track well polynomial trends of different degrees p .

Simple FMH-filters are adapted to a constant signal ($p = 1$), using $M = 3$ subfilters, namely two one-sided moving-averages and the current observation y_t :

$$\Phi_1(y_t) = \frac{1}{m} \sum_{i=1}^m y_{t-i}, \quad \Phi_2(y_t) = y_t, \quad \Phi_3(y_t) = \frac{1}{m} \sum_{i=1}^m y_{t+i}.$$

Including the central observation as central subfilter allows to preserve level shifts even better than running medians [4]. FMH-filters are suitable only in retrospective application when using backward forecasting filters. We thus set $\tilde{m} = m$ in general.

Predictive FMH-filters apply subfilters for one-sided extrapolation of linear trends ($p = 1$):

$$PFMH(y_t) = \text{med}(\Phi_F(y_t), y_t, \Phi_B(y_t)),$$

where $\Phi_F(y_t) = \sum_{i=1}^m h_i y_{t-i}$ and $\Phi_B(y_t) = \sum_{i=1}^m h_i y_{t+i}$. The minimal MSE predictions for a linear trend in the case of Gaussian noise under the restriction that the exact signal value is obtained in the deterministic situation without noise use the weights $h_i = \frac{4m-6i+2}{m(m-1)}$, $i = 1, \dots, m$, see [43].

Combined FMH-filters use predictions of different degrees,

$$CFMH(y_t) = \text{med}(\Phi_F(y_t), \Phi_1(y_t), y_t, \Phi_3(y_t), \Phi_B(y_t)),$$

where $\Phi_1(y_t)$, $\Phi_3(y_t)$, $\Phi_F(y_t)$ and $\Phi_B(y_t)$ are the subfilters for forward and backward extrapolation of a constant signal or a linear trend given above.

FMH-filters have a smaller bias error at level shifts than running medians at the expense of a larger variance around the shift [4]. They do not suffer from edge jitter, but a spike - distant at most m time points from a shift - causes some smearing as the height of a shift and a constant signal value close to the shift change [42]. FMH-filters recover linear trends in the

absence of noise exactly, but only the PFMHs are trend-invariant and thus can preserve shifts within trends as good as in constant periods.

However, PFMHs are neither very efficient for Gaussian noise nor very robust. All FMH-filters dampen isolated outliers better than running medians [43], but already two outliers can affect them strongly [29]. The CFMH-filters improve the Gaussian efficiency of PFMHs considerably when the signal is constant, becoming about as efficient as a simple FMH or a median with the same width [42]. However, this advantage gets lost with increasing signal slope. Every FMH-filter is Lipschitz-continuous with constant $\max |h_i^j|$, the maximal absolute weight given by one of the subfilters.

Different from running medians, FMH-filters create new values and can smooth oscillations between two measurements. Besides signals consisting only of local constant neighborhoods and edges, among the roots of FMH-filters we find e.g. triangular waves, which are not roots of running medians [42, 43]. Repeated filtering with increasing window widths helps to overcome the typical triangular wave form of FMH-filtered time series.

Variations of FMH-filters have been proposed. Recursive FMH-filters apply the previously filtered values in the forward predictions. They provide better noise reduction than their non-recursive counterparts and running medians, but they distort edges because of larger bias errors [4]. In-place growing FMH-filters use a cascade of FMH-filters of different widths [90],

$$\begin{aligned} z_t^{(0)} &= y_t \\ z_t^{(j)} &= \text{med}(\Phi_{l_j}(y_t), z_t^{(j-1)}, \Phi_{r_j}(y_t)) \end{aligned}$$

with subfilters Φ_{l_j} and Φ_{r_j} of width increasing in j . These filters preserve shifts better than median and (recursive) FMH-filters. Similar variants have been suggested for improved trend elaboration [67]. Finally, weights can be given to the linear subfilters. An optimal FIR-WOS hybrid filter under the MAE criterion can be found by an adaptive algorithm [93].

A general framework for adaptive order statistic, i.e. location based filtering is developed in [45]. Similar to the hybrid filters discussed before, the idea is to use test statistics for selecting one of the location estimates obtained from different subwindows, or more generally, to obtain a weighted linear combination of all of them. A triple window median filter turned out to perform particularly well for retrospective elimination of impulsive noise and edge preservation.

4.2 Repeated median hybrid filters

To overcome the lack of robustness of FMH-filters, we can construct hybrid filters with robust instead of linear subfilters [29]. We replace the half-window averages in the simple and the combined FMH by half-window medians, and use half-window repeated medians RM^F and

RM^B for a linear trend :

$$\begin{aligned}
 RM^F(y_t) &= \text{med} \left(y_{t-m} + m\hat{\beta}_t^F, \dots, y_{t-1} + \hat{\beta}_t^F \right), \\
 \hat{\beta}_t^F &= \text{med}_{i=-m, \dots, -1} \left(\text{med}_{j=-m, \dots, -1, j \neq i} \frac{y_{t+i} - y_{t+j}}{i - j} \right), \\
 RM^B(y_t) &= \text{med} \left(y_{t+1} - \hat{\beta}_t^B, \dots, y_{t+m} - m\hat{\beta}_t^B \right), \\
 \hat{\beta}_t^B &= \text{med}_{i=1, \dots, m} \left(\text{med}_{j=1, \dots, m, j \neq i} \frac{y_{t+i} - y_{t+j}}{i - j} \right).
 \end{aligned}$$

The resulting repeated median hybrid (RMH)-filters are Lipschitz-continuous with the same constant $2m + 1$ as the RM. Fast update algorithms are available for the computation [29]. RMH-filters are location- and scale-equivariant, but as for FMH-filters only the predictive version is trend-invariant. Replacing the central observation by the median of the whole window increases robustness and the Gaussian efficiency while also preserving shifts, but destroys the trend-invariance, see Fig. 3.

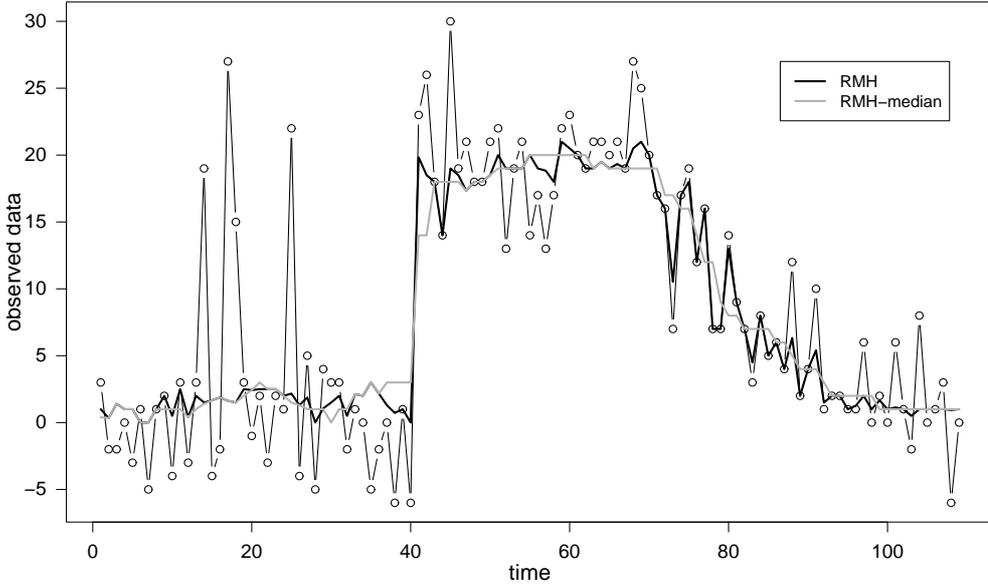


Figure 3: Observed time series with outliers, level shift and trend together with approximations of the signal by means of median hybrid filters: RMH-filter and robustified RMH-filter (median replacing the central observation) with $m = \tilde{m} = 10$.

RMH-filters have the same nice properties w.r.t. shift preservation as FMH-filters, while improving upon them w.r.t. the removal of spikes. RMH-filters preserve shifts better than the median even if the signal is constant. Only the predictive RMH preserves shifts irrespectively

of a trend because of its invariance, while the combined RMH has problems with shifts into the opposite direction of a trend, but less than the median.

The predictive RMH and the combined RMH have breakdown points $(\lfloor m/2 \rfloor + 1)/n$ and $(\lfloor m/2 \rfloor + 2)/n$, respectively, so that they guarantee some protection against up to five and six outliers, respectively, within $n = 21$ observations, while FMH-filters do so only for a single outlier. The situation with a single long outlier patch starting right in the center of the window turns out to be a worst case situation for most of these hybrid methods, while several short patches have much smaller effects. In simulations, two outliers are found to damage the FMHs considerably, while the RMHs resist them substantially better. The combined RMH even resists about 25% outliers when the signal is constant. However, the RMHs can be more affected by a patch of successive outliers than the standard RM.

With respect to Gaussian efficiency, RMH-filters are only slightly worse than the respective FMH-filters. Again, the combined versions are more efficient than the predictive ones if the signal is constant, but this gain gets lost with increasing slope.

Summarizing, RMH-filters are preferable to FMH-filters since they provide the same benefits and are considerably more robust for the prize of only a small loss in efficiency under the Gaussian. As compared to the standard RM, they attenuate Gaussian noise and long sequences of spikes less efficiently and are more variable, but they can preserve shifts and local extremes even better than the median when the signal is constant. The combined RMH improves the efficiency and the robustness of the predictive RMH in case of a constant signal, but the latter preserves shifts irrespectively of a trend.

4.3 Level shift detection

Instead of designing filters for improved shift preservation, we can incorporate rules for shift detection so that appropriate actions can be taken. Accordingly, an abundance of edge detection rules has been suggested. Some kind of low-pass filtering followed by differentiation is a common approach. FMH-detectors combining several edge detection rules are outlined in [66]. However, detection rules based on differences which are optimized e.g. for the Gaussian distribution can be adversely affected by deviations from this assumption, and in particular such non-robust rules often confuse spikes with shifts.

Robust shift detection can be based on a comparison of two robust level estimates. Considerable robustness with only a small loss under the Gaussian can already be achieved when using an F-test comparing trimmed means and using a winsorized variance for standardization [94, 46]. A retrospective multilevel filter for edge detection and efficient suppression of different types of noise is suggested in [49]. If two half-sided median subfilters deviate largely, an edge is detected and the filter output is calculated as the median of these half-sided medians and the current observation. Alternatively, it is decided if the shift has happened at this or the previous time point depending on whether or not a shift had been detected before, and the filter output is chosen accordingly as one of the half-sided medians. If no shift is detected, the filter output is the average of the half-sided medians. Optionally, a preliminary impulse detection step can be added and the half-sided medians can be replaced by half-sided averages or midpoints to suppress noise with normal or light tails more efficiently. The deviation between the half-sided medians from which on a shift is detected can be determined in a Bayesian framework by specifying the a-priori probability of a shift. These filters assume

a piecewise constant signal, but in simulations they perform better than running medians also during trends. Gradient estimates formed from differences of medians, or more generally trimmed means, have also been suggested as robust alternatives to ordinary means [13].

Edges can also be identified via detecting an increase of the local variability. Quasi-ranges $y_{t(n-i+1)} - y_{t(i)}$ are simple scale estimates, and the interquartile range is a usual robust standard. Double-window Hodges-Lehman-Bickel (HLB) and HLB median hybrid filters for improved edge preservation and noise suppression based on this principle are derived in [57]. The hybrid filters replace the HLB-estimate by the median when a shift is detected. The double window filter trim all observations which are far from the median before calculating the HLB-estimate. The HLB double window filters are found to provide better noise suppression than running medians, but they are outperformed by HLB hybrid filters. The interquartile range is also applied in [77]. The filter output is taken to be either the median or a trimmed mean with an adaptive amount of trimming, depending on whether a shift has been detected or not. A basic problem is the choice of the threshold for shift detection. A comparison of the neighbors of the median for edge detection is suggested in [61], i.e. $y_{t(m+2)} - y_{t(m)}$, where $n = 2m + 1$. If additionally the difference between the central observation and the median is large, it is concluded that a spike has occurred in addition to the shift, and a modified trimmed mean centered at a suitably chosen neighbor of the median is calculated to reduce edge shifting [61].

The empirical variance s_t^2 within the time window can also be applied for edge detection. Adaptive L-filters (AL) can be constructed as a convex combination of the local mean and median for retrospective application, where $\tilde{m} = m$ [86]. The weight of the median increases with s_t^2 since a shift is regarded as more likely then,

$$AL(y_t) = w_t y_{t(m+1)} + (1 - w_t) \bar{y}_t,$$

where $w_t = v_t^2 / (\sigma_u^2 + v_t^2)$, σ_u^2 is the noise variance, and $v_t^2 = s_t^2 - \sigma_u^2$ estimates the local variation of the signal. The noise variance is assumed to be constant and needs to be known or estimated from smooth signal regions. The lack of robustness of the empirical variance does not cause a problem since the filter output tends to the median as s_t^2 goes to infinity. The filter is unbiased in case of symmetric noise and inherits the good properties of the median, namely edge preservation and removal of spikes, while offering larger efficiency under the Gaussian distribution. A modified version replacing the median by a weighted median for using larger windows is also suggested. The current observation y_t could be taken instead of the median for better edge preservation [87], but then the filter loses its robustness completely. Improvements are possible, e.g. by replacing the mean by an adaptively chosen trimmed mean based on the tails of the noise. For edge detection we can use $(s_t^2 / \sigma_u^2)(s_t^2 / \sigma_u^2 - 1)$. This quantity is close to zero if $s_t^2 \approx \sigma_u^2$, and largely positive if $s_t^2 \gg \sigma_u^2$, but there could occur again a confusion of spikes and shifts.

Further possibilities for edge detection are tests based on linear rank statistics, particularly the Wilcoxon and the median test [11]. The former is almost as effective as tests based on averages in case of Gaussian noise, but it is more robust to deviations from the Gaussian assumption, while the latter performs well even in presence of a substantial amount of impulses. The main disadvantage of these tests is probably that for short time windows a given significance level is difficult to obtain, because of the discretization due to using ranks. A

comparative study shows that the rank based tests and the tests comparing two robust level estimates outperform tests based on the local variability [32]. The latter seem interesting mainly for 2d-signal (image) processing since they do not need specification of a direction for the shift.

All the tests described above rely on shifts to arise from one constant signal value to another. robust regression techniques can be applied to adapt these rules for being suitable during trends. We will just outline one such possibility, which is described in more detail in [27].

Robust shift detection within trends is made possible by a simple majority rule, applied to the repeated median residuals in the current window. A positive level shift is detected if more than half of the most recent RM residuals, or another appropriate large fraction of them, is larger than a multiple of a robust estimate of the variability, e.g. the MAD about the regression line. An analogous rule is used for negative shifts. Using twice the scale estimate for the threshold is a reasonable standard choice since small shifts are often irrelevant and can be accommodated otherwise. If we base the shift detection on the most recent $\lfloor n/2 \rfloor$ observations, requiring that at least half of them deviate widely from the regression fit, one quarter of outlying observations in the current window can have arbitrarily large effects. This means an indispensable loss of robustness when adding such a rule, but a shift can still be detected with a short delay if almost a quarter of the observations after the shift are outliers.

When a shift is detected, suitable actions need to be taken and the procedure be restarted. Shortening the window minimizes the blurring of edges [52]. For restarting we typically need to specify the time point at which the shift has happened. A simple possibility is to use the first time point at which a signal was triggered, or the first time point at which we found a large deviating residual when applying the above majority rule. Adaptive exponential smoothing for improved filtering close to shifts is proposed in [92], constructing a convex combination of the current observation and the previous level estimate with weights depending on the last time point of a shift. However, such schemes are sensitive to outliers. As pointed out by the author himself, a robustification would be desirable.

4.4 Impulse detection

Rules for the detection of spikes can be applied as well. Spikes, also called impulses in the literature, are sometimes interesting for their own sake, or simply because we can replace detected impulses by a cleaned value to increase the robustness of the basic procedure.

In the location context, a couple of approaches for impulse detection has been suggested. Distribution-free rules can be based on the rank of the inspected observation within the time window since outliers are expected to be among the most extreme observations [41]. Difficulties are a high false-detection rate if not only the smallest and the largest observation are regarded as outlying, as well as a lack of detection power in case of outlier sequences. Another possibility is to use the distance to the median for measuring outlyingness, but such rules cannot distinguish between outliers and shifts. To overcome these problems, rank-based and distance-based rules should be combined [1].

The robustness of the repeated median can be further increased by adding automatic rules for outlier detection and replacement based on robust scale estimators like the MAD [27]. We can check whether the incoming observation $y_{t+\tilde{m}+1}$ is outlying by comparing its residual

$r_{\tilde{m}+1} = y_{t+\tilde{m}+1} - \hat{\mu}_t - \hat{\beta}_t(\tilde{m} + 1)$ w.r.t. the current regression line to the estimate $\hat{\sigma}_t$ of the standard deviation about the regression line. A promising alternative to the classical MAD for robust scale estimation is [80]

$$\hat{\sigma}_t^{QN} = d_n \cdot \{|r_i - r_j| : -m \leq i < j \leq \tilde{m}\}_{(h)}, \quad h = \binom{m+1}{2},$$

where $r_i = y_{t+i} - \tilde{\mu}_t - i\tilde{\beta}_t$, $i = -m, \dots, \tilde{m}$. This estimate shows excellent performance at the occurrence of level shifts and performs better than the MAD in the presence of identical measurements (inliers) due to e.g. rounding. Here, d_n is another finite-sample correction factor depending on the window width $n = m + \tilde{m} + 1$. Replacing detected outliers by their prediction $\hat{\mu}_t + \hat{\beta}_t(\tilde{m} + 1)$ gives almost the same robustness as least median of squares (LMS) regression even in extreme situations, but additional rules need to be added since e.g. level shifts remain undetected otherwise because all shifted observations are replaced. Such combined procedures seem preferable to the LMS because of the much better performance in moderate outlier situations and the smaller computational costs.

Many outlier detection rules like the previous one are based on a single difference between the inspected observation and a level estimate. For location-based filters multiple comparison to several weighted medians has been proposed in [14].

5 Conclusions

Starting with running medians, many filters have been suggested for detail-preserving robust signal extraction from noisy time series. Many contributions in the literature focus on the attenuation of different types of noise, just imposing that desired signal details like trends are preserved under idealized conditions like the complete absence of observational noise, or that certain signals are roots of the filter. These restrictions are rather weak. Thus, a substantial loss of filtering quality, namely both bias and increased variability, may occur. Requiring appropriate equivariances and invariances whenever possible allows to construct filters which keep their performance at the occurrence of the interesting signal details.

In particular, locally linear trends can be dealt with using robust regression. Such techniques additionally allow to overcome the inherent delay which hampers the online application of location-based filters to signals which are not piecewise constant. The repeated median has been regarded a promising method for time series filtering in a couple of investigations. Fast update algorithms are available allowing its application even online and to high frequency data. Similar as for the standard median, modifications are possible for better preservation of shifts and local extremes. Repeated median hybrid filters offer excellent performance in this respect, but they lose robustness and Gaussian efficiency. A reasonable compromise can be achieved by double-window trimmed repeated medians. Weighted repeated medians seem very promising for online analysis.

Many interesting aspects could not be addressed in this chapter. Like many other studies we have restricted to the case of independent errors. Here it can be said that the positive auto-correlations found in many applications further increase the efficiencies of robust estimators under Gaussian assumptions as compared to least squares techniques [8, 28, 29, 31]. The filters discussed here are designed to improve preservation of certain signal details. Specially

designed adaptive order statistic filters even allow to recover certain signal details, which have been lost before e.g. due to linear filtering [58]. Finally, repeated medians can also be applied for highly robust frequency domain analysis, fitting robust sine and cosine coefficients [88].

References

- [1] I. Aizenberg, and C. Butakoff, Effective impulse detector based on rank-order criteria, *IEEE Signal Processing Letters* 11, 363-366 (2004).
- [2] G.R. Arce, A general weighted median filter structure admitting real-valued weights, *IEEE Transactions on Signal Processing* 46, 3195-3205 (1989).
- [3] G.R. Arce, R.E. Foster, Detail preserving ranked-order based filters for image processing, *IEEE Transactions on Acoustics, Speech, and Signal Processing* 37, 83-98 (1989).
- [4] J. Astola, P. Heinonen, and Y. Neuvo, Linear median hybrid filters, *IEEE Transactions on Circuit and Systems* 36, 1430-1438 (1989).
- [5] J. Astola, P. Kuosmanen, *Fundamentals of Nonlinear Digital Filtering*, (CRC, Boca Raton, FL, 1997).
- [6] J.B. Bednar, and T.L. Watt, Alpha-trimmed means and their relationship to median filters, *IEEE Transactions on Acoustics, Speech, and Signal Processing* 32, 145-153 (1984).
- [7] T. Bernholt, and R. Fried, Computing the update of the repeated median regression line in linear time, *Information Processing Letters* 88, 111-117 (2003).
- [8] T. Bernholt, R. Fried, U. Gather, and I. Wegener, Modified repeated median filters, *Technical Report 46/2004, SFB 475, University of Dortmund, Germany* (2004).
- [9] A.C. Bovik, Streaking in median filtered images, *IEEE Transactions on Acoustics, Speech, and Signal Processing* 35, 493-503 (1987).
- [10] A.C. Bovik, T.S. Huang, and D.C. Munson Jr., A generalization of median filtering using linear combinations of order statistics, *IEEE Transactions on Acoustics, Speech, and Signal Processing* 31, 1342-1350 (1983).
- [11] A.C. Bovik, T.S. Huang, and D.C. Munson Jr., Nonparametric tests for edge detection in noise, *Pattern recognition* 19, 209-219 (1986).
- [12] A.C. Bovik, T.S. Huang, and D.C. Munson Jr., The effect of median filtering on edge estimation and detection, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 9, 181-194 (1987).
- [13] A.C. Bovik, and D.C. Munson Jr., Edge detection using median comparisons, *Computer Vision, Graphics, and Image Processing* 33, 377-389 (1986).
- [14] T. Chen, and H.R. Wu, Adaptive impulse detection using center-weighted median filters, *IEEE Signal Processing Letters* 8, 1-3 (2001).
- [15] K.-S. Choi, A.W. Morales, and S.-J. Ko, Design of linear combinations of weighted medians, *IEEE Transactions on Signal Processing* 49, 1940-1952 (2001).
- [16] T. Cipra, Robust exponential smoothing, *Journal of Forecasting* 11, 57-69 (1992).
- [17] P.L. Davies, Aspects of robust linear regression, *Annals of Statistics* 21, 1843-1899 (1993).
- [18] P.L. Davies, R. Fried, and U. Gather, Robust signal extraction for on-line monitoring data, *Journal of Statistical Planning and Inference* 122, 65-78 (2004).

- [19] R. Ding, and A.N. Venetsanopoulos, Generalized homomorphic and adaptive order statistic filters for the removal of impulsive and signal-dependent noise, *IEEE Transactions on Circuits and Systems* 34, 948-955 (1987).
- [20] I. Djurovic, V. Katkovnik, and L. Stankovic, Median filter based realizations of the robust time-frequency distributions, *Signal Processing* 81, 1771-1776 (2001).
- [21] D.L. Donoho, and P.J. Huber, 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, 157-184 (1983).
- [22] H. Edelsbrunner, and D.L. Souvaine, Computing least median of squares regression lines and guided topological sweep, *Journal of the American Statistical Association* 85, 115-119 (1990).
- [23] J. Einbeck, and G. Kauermann, Online monitoring with local smoothing methods and adaptive ridging, *Journal of Statistical Computation and Simulation* 73, 913-929 (2003).
- [24] S.P. Ellis, and S. Morgenthaler (1992), Leverage and breakdown in L1 regression, *Journal of the American Statistical Association* 87, 143-148 (1991).
- [25] J. Fan, T.-C. Hu, and Y.K. Truong, Robust nonparametric function estimation, *Scandinavian Journal of Statistics* 21, 433-446 (1994).
- [26] A. Flaig, G.R. Arce, and K.E. Barner, Affine order statistic filters: "Medianization" of linear FIR filters, *IEEE Transactions on Signal Processing* 46, 2101-2112 (1998).
- [27] R. Fried, Robust filtering of time series with trends, *Journal of Nonparametric Statistics* 16, 313-328 (2004).
- [28] R. Fried, Robust location estimation under dependence, *Journal of Statistical Computation and Simulation*, to appear (2005).
- [29] R. Fried, T. Bernholt, and U. Gather, Repeated median and hybrid filters, *Computational Statistics & Data Analysis*, to appear (2005).
- [30] R. Fried, J. Einbeck, and U. Gather, Weighted repeated median smoothing and filtering, *Technical Report 33/2005, SFB 475, University of Dortmund, Germany* (2005).
- [31] R. Fried, and U. Gather, Robust trend estimation for AR(1) disturbances, *Austrian Journal of Statistics* 34, 139-151 (2005).
- [32] R. Fried, and U. Gather, Robust shift detection in time series, *Working Paper*, Department of Statistics, University of Dortmund, Germany (2006).
- [33] M. Gabbouj, E.J. Coyle, and N.C. Gallagher, Jr., An overview of median and stack filtering, *Circuits Systems Signal Process* 11, 7-45 (1992).
- [34] N.C. Gallagher, and Wise, G.L., A theoretical analysis of the properties of median filters, *IEEE Transactions of Acoustic Speech, and Signal Processing* 29, 1136-1141 (1981).
- [35] P.P. Gahndi, I. Song and S.A. Kassam, Nonlinear smoothing filters based on rank estimates of location, *IEEE Transactions on Acoustics, Speech, and Signal Processing* 37, 1359-1379 (1989).
- [36] P.P. Gahndi, and S.A. Kassam, Design and performance of combination filters for signal restoration, *IEEE Transactions on Signal Processing* 39, 1524-1540 (1991).
- [37] U. Gather, and R. Fried, Robust estimation of scale for local linear temporal trends, *Tatra Mountains Mathematical Publications* 26, 87-101 (2003).

- [38] U. Gather, and R. Fried, Methods and algorithms for robust filtering, in *Proceedings in Computational Statistics COMPSTAT 2004* 159-170, ed. J. Antoch, (Physica-Verlag, Heidelberg, 2004).
- [39] U. Gather, K. Schettlinger, and R. Fried, Online signal extraction by robust linear regression, *Computational Statistics* to appear (2005).
- [40] F.R. Hampel, Beyond location parameters: Robust concepts and methods. *Bulletin of the International Statistical Institute* 46, 375–382 (1975).
- [41] A.B. Hamza, and H. Krim, Image denoising: A nonlinear robust statistical approach, *IEEE Transactions on Signal Processing* 49, 3045-3054 (2001).
- [42] P. Heinonen, and Y. Neuvo, FIR-median hybrid filters, *IEEE Transaction on Acoustic, Speech, and Signal Processing* 35, 832-838 (1987).
- [43] P. Heinonen, and Y. Neuvo, FIR-median hybrid filters with predictive FIR substructures, *IEEE Transaction on Acoustic, Speech, and Signal Processing* 36, 892-899 (1988).
- [44] N. Himayat, and S.A. Kassam, Approximate performance analysis of edge preserving filters, *IEEE Transaction on Signal Processing* 41, 2764-2777 (1993).
- [45] N. Himayat, and S.A. Kassam, A structure for adaptive order statistics filtering, *IEEE Transaction on Image Processing* 3, 265-280 (1994).
- [46] Z. Hou, and T.S. Koh, Robust edge detection, *Pattern Recognition* 36, 2083-2091 (2003).
- [47] S. Hoyos, J. Bacca, and G.R. Arce, Spectral design of weighted median filters: A general iterative approach, *IEEE Transaction on Signal Processing* 53, 1045-1056 (2005).
- [48] P.J. Huber, The Notation of Breakdown Point, *Robust Statistics*, (John Wiley, New York, 1981).
- [49] H. Hwang, and R.A. Haddad, Multilevel nonlinear filters for edge detection and noise suppression, *IEEE Transactions on Signal Processing* 42, 249-258 (1994).
- [50] H. Hwang, and R.A. Haddad, Adaptive median filter: New algorithms and results, *IEEE Transactions on Image Processing* 4, 499-502 (1995).
- [51] B.I. Justusson, Median filtering: Statistical properties, in *Two-Dimensional Digital Signal Processing II*, ed. T. S. Huang, (Springer, New York, 1981).
- [52] S.A. Kassam, Robust hypothesis testing and robust time series interpolation and regression, *Journal of Time Series Analysis* 3, 185-194 (1982).
- [53] S.A. Kassam, and H.V. Poor, Robust techniques for signal processing: A survey, *Proc. IEEE* 73, 433-481 (1985).
- [54] V. Katkovnik, K., Egiazarin, and J. Astola, Application of the ICI principle to window size adaptive median filtering, *Signal Processing* 83, 251-257 (2003).
- [55] P. Koivisto, O. Yli-Harja, A. Niemisto, and I. Shmulevich, Breakdown probabilities of recursive stack filters, *Signal Processing* 81, 227-231 (2001).
- [56] V. Koivunen, A robust nonlinear filter for image restoration, *IEEE Transactions on Image Processing* 4, 569-578 (1995).
- [57] A. Kundu, and W.-R. Wu, Double-window Hodges-Lehman (D) filter and hybrid D-median filter for robust image smoothing, *IEEE Transactions on Acoustics, Speech, and Signal Processing* 37, 1293-1298 (1989).
- [58] Y.H. Lee, and A.T. Fam, An edge gradient enhancing adaptive order statistics filter, *IEEE Transactions on Acoustics, Speech, and Signal Processing* 35, 680-693 (1987).

- [59] Y.H. Lee, and S.A. Kassam, Generalized median filtering and related nonlinear filtering techniques, *IEEE Transactions on Acoustics, Speech, and Signal Processing* 33, 672-683 (1985).
- [60] K.-M. Lee, P. Meer, and R.-H. Park, Robust adaptive segmentation of range images, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20, 200-205 (1998).
- [61] Y.H. Lee, and S. Tantaratana, Decision-based order statistic filters, *IEEE Transactions on Acoustics, Speech, and Signal Processing* 38, 406-420 (1990).
- [62] H.M. Lin, A.N. Willson, Jr., Median filters with adaptive length, *IEEE Transactions on Circuits and Systems* 35 (1988).
- [63] H.G. Longbotham, and A.C. Bovik, Theory of order statistic filters and their relationship to linear FIR filters, *IEEE Transactions on Acoustics, Speech, and Signal Processing* 37, 275-287 (1989).
- [64] T. Loupas, W.N. McDicken, and P.L. Allan, An adaptive weighted median filter for speckle suppression in medical ultrasonic images, *IEEE Transactions on Circuits and Systems* 36, 129-135 (1989).
- [65] C.L. Mallows, Some Theory of nonlinear smoothers, *Annals of Statistics* 8, 695-715 (1980).
- [66] Y. Neuvo, P. Heinonen, and I. Defee, Linear-median hybrid edge detectors, *IEEE Transactions on Circuits and Systems* 34, 1337-1343 (1987).
- [67] A. Nieminem, Y. Neuvo, and U. Mitra, Algorithms for real-time trend detection, *Signal Processing* 18, 1-15 (1989).
- [68] A. Nieminem, P. Heinonen, and Y. Neuvo, A new class of detail-preserving filters for image processing, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 9, 74-90 (1987).
- [69] J. Nieweglowski, M. Gabbouj, and Y. Neuvo, Weighted medians-positive boolean functions conversion, *Signal Processing* 34, 149-162 (1987).
- [70] R. Öten, and R.J.P. Figueiredo, An efficient method for L-filter design, *IEEE Transactions on Signal Processing* 51, 193-203 (2003).
- [71] F. Palmieri, and C.G. Boncelet, LI-filters - a new class of order statistic filters, *IEEE Transactions on Acoustics, Speech, and Signal Processing* 37, 691-701 (1989).
- [72] M.K. Prasad, and Y.H. Lee, Weighted Median Filters: Generation and Properties, in *Proceedings of the IEEE International Symposium on Circuits and Systems ISCAS 89*, 425-428 (1989).
- [73] S.R. Peterson, Y.H., Lee, and S.A. Kassam, Some statistical properties of alpha-trimmed mean and standard-type M filters, *IEEE Transactions on Acoustics, Speech, and Signal Processing* 36, 707-713 (1988).
- [74] I. Pitas, and A.N. Venetsanopoulous, *Nonlinear Digital Filters: Principles and Applications*, (Kluwer, Boston, 1990).
- [75] I. Pitas, and A.N. Venetsanopoulous, Order statistics in digital image processing, *Proceedings of the IEEE* 80, 1893-1921 (1992).
- [76] C.A. Pomalaza-Raez, and C.D. McGillem, An adaptive, nonlinear edge-preserving filter, *IEEE Transactions on Acoustics, Speech, and Signal Processing* 32, 571-576 (1984).

- [77] A. Restrepo, and A.C. Bovik, Adaptive trimmed mean filters for image restoration, *IEEE Transactions on Acoustic, Speech, and Signal Processing* 36, 1326-1337 (1988).
- [78] P.J. Rousseeuw, Multivariate Estimation with High Breakdown Point, in *Proceedings of the 4th Pannonian Symposium on Mathematical Statistics and Probability, Vol. B*, eds. W. Grossmann, G. Pflug, I. Vincze, W. Wertz, (D. Reidel Publishing Company, Dordrecht, 1983).
- [79] P.J. Rousseeuw, Least median of squares regression, *Journal of the American Statistical Association* 79, 871-880 (1984).
- [80] P.J. Rousseeuw, and Ch. Croux, Alternatives to the median absolute deviation, *Journal of the American Statistical Association* 88, 1273-1283 (1993).
- [81] P.J. Rousseeuw, and M. Hubert, Regression Depth, *Journal of the American Statistical Association* 94, 388-402 (1999).
- [82] P.J. Rousseeuw, and A. Leroy, A., *Robust regression and outlier detection*, (Wiley, New York, 1987).
- [83] A.F. Siegel, Robust regression using repeated medians, *Biometrika* 68, 242-244 (1982).
- [84] W.-J. Song, and W.A. Pearlman, Edge-preserving noise filtering based on adaptive windowing, *IEEE Transactions on Circuits and Systems* 35, 1048-1055 (1988).
- [85] R.A. Stein, and T.J. Fowlow, The use of median filters for edge detection in noisy time series, *Proceedings of ISCAS 85*, 1331-1334 (1985).
- [86] T. Sun, M., Gabbouj, and Y. Neuvo, Adaptive L-filters with applications in signal and image processing, *Signal Processing* 38, 331-344 (1994).
- [87] X.Z. Sun, and A.N. Venetsanopoulos, Adaptive schemes for noise filtering and edge detection by use of local statistics, *IEEE Transactions on Circuits and Systems* 35, 57-69 (1988).
- [88] L.G. Tatum, and C.M. Hurvich, High breakdown methods of time series analysis, *Journal of the Royal Statistical Society: Series B* 55, 881-896 (1993).
- [89] J. W. Tukey, *Exploratory Data Analysis*, (Addison-Wesley, Reading, Mass., 1977, preliminary edition 1971).
- [90] R. Wichman, J.T. Astola, P.J. Heinonen, and Y.A. Neuvo, FIR-Median hybrid filters with excellent transient response in noisy conditions *IEEE Transactions on Acoustics, Speech, and Signal Processing* 38, 2108-2117 (1990).
- [91] R. Yang, L. Yin, M. Gabbouj, J. Astola, and Y. Neuvo, Optimal weighted median filtering under structural constraints, *IEEE Transactions on Signal Processing* 43, 591-604 (1995).
- [92] E. Yashchin, Change-point models in industrial applications, *Nonlinear Analysis, Theory, Methods and Applications* 30, 3997-4006 (1997).
- [93] L. Yin, R. Yang, M. Gabbouj, and Y. Neuvo, Weighted median filters: A tutorial, *IEEE Transactions on Circuits and Systems II* 43, 157-192 (1996).
- [94] K.K. Yuen, K.K., The two-sample trimmed t for unequal population variances, *Biometrika* 61, 165-170 (1974).