



Repeated median and hybrid filters

Roland Fried^{a,*}, Thorsten Bernholt^b, Ursula Gather^c

^a*Department of Statistics, University Carlos III, 28903 Getafe (Madrid), Spain*

^b*Department of Computer Science, University of Dortmund, 44221 Dortmund, Germany*

^c*Department of Statistics, University of Dortmund, 44221 Dortmund, Germany*

Available online 19 January 2005

Abstract

Standard median filters preserve abrupt shifts (edges) and remove impulsive noise (outliers) from a constant signal but they deteriorate in trend periods. Finite impulse response median hybrid (FMH) filters are more flexible and also preserve shifts, but they are much more vulnerable to outliers. Application of robust regression, in particular of the repeated median, has been suggested for removing subsequent outliers from a signal with trends. A fast algorithm for updating the repeated median in linear time using quadratic space has been given by Bernholt and Fried (Inform. Process. Lett. 88 (2003) 111). Repeated median hybrid filters are constructed to combine the robustness of the repeated median with the edge preservation of FMH filters. Analytical properties of these filters are investigated and their performance is compared via simulations. An algorithm for updating the repeated median is presented which needs only linear space.

© 2005 Elsevier B.V. All rights reserved.

Keywords: Signal extraction; Drifts; Jumps; Outliers; Update algorithm

1. Introduction

Robust filtering of online monitoring data is a challenging task. In intensive care the heart rate and blood pressure are measured in short time lags. These data provide important information for medical decision support. The extraction of the underlying, clinically relevant signal from the observed time series is a basic goal. A solution needs to preserve abrupt level

* Corresponding author. Tel.: +34 91 624 5854; fax: +34 91 624 9849.

E-mail addresses: rfried@est-econ.uc3m.es, fried@statistik.uni-dortmund.de (R. Fried).

shifts and monotonic trends, while irrelevant short-term fluctuations and outliers caused by, e.g. patient's movements or measurement artifacts should be removed.

The extraction of a signal (μ_t) from a time series (x_t) can be formalized using a components model like

$$x_t = \mu_t + u_t + v_t, \quad t \in \mathbb{Z},$$

where μ_t is the smoothly varying level of the series, possibly with a few abrupt shifts, u_t is symmetric observational noise with mean zero and variance σ^2 , and v_t is impulsive (spiky) noise from an outlier generating mechanism.

Linear filters such as moving averages track trends and are very efficient for Gaussian noise (u_t), but they are highly vulnerable to impulsive noise (outliers or spikes) and they blur level shifts (also called step changes, edges or jumps). Standard median filters remove outliers and preserve level shifts (Tukey, 1977; Nieminem et al., 1989) but their output does not properly represent linear trends. Finite impulse response median hybrid (FMH) filters (Heinonen and Neuvo, 1987, 1988) combine advantages of linear and nonlinear filters by taking the median value of several linear FIR subfilters. The linear subfilters can be designed to track polynomial trends well, and the use of the central observation as central subfilter allows to preserve level shifts about as well as standard median filters. Davies et al. (2004) suggest application of a robust regression estimator such as the repeated median to overcome the difficulties of the median in trend periods.

In this paper, we investigate robust hybrid filters which are designed to combine the good properties of FMH filters with the robustness of the repeated median in trend periods. Instead of linear subfilters we use robust counterparts, namely half-window medians for a constant signal and half-window repeated medians for a linear trend. For the central subfilter, we can apply either the central observation or the median value of the whole time window. We investigate analytical properties of the resulting hybrid filters and compare their ability to attenuate Gaussian noise, to resist outliers and to preserve abrupt shifts in case of a constant signal and in a trend period via simulations.

In Section 2, we review repeated median and FMH filters and introduce repeated median hybrid filters as a combination of these. In Section 3, we discuss the fast computation of the filters and propose a new algorithm for updating the repeated median. In Section 4, we investigate analytical properties of the filters. In Section 5, we compare their performance in basic data situations via simulations. In Section 6, we present applications to real and simulated data for further comparison before we end up drawing some conclusions.

2. Hybrid filters

We introduce some classes of nonlinear filters. Let the filter input be a time series (x_t). Moving averages and standard median filters are based on the idea of approximating the signal (μ_t) underlying (x_t) by moving a time window of fixed width $n = 2k + 1$ through the series to approximate the level of the time series in the center of the window. Hence, applying any such method causes a time delay of k observations. A short delay can be achieved by a small choice of k , but for the cost of reduced smoothing.

2.1. Filters based on robust regression

Standard median filters (running medians) approximate the signal μ_t in the center of a moving window $(x_{t-k}, \dots, x_{t+k})$ by the median of these observations,

$$SM(x_t) = \tilde{\mu}_t = \text{med}(x_{t-k}, \dots, x_{t+k}), \quad t \in \mathbb{Z}.$$

This means treating the level μ_t as locally almost constant. This assumption is of course not appropriate in trend periods. Therefore, Davies et al. (2004) suggest fitting a local linear trend $\mu_{t+i} = \mu_t + i\beta_t$, $i = -k, \dots, k$, by robust regression. Here, μ_t is again the level and β_t is the slope in the center of the time window. Based on a comparison of methods with high breakdown point they recommend Siegel's (1982) repeated median

$$RM(x_t) = \tilde{\mu}_t^{RM} = \text{med}\left(x_{t-k} + k\tilde{\beta}_t, \dots, x_{t+k} - k\tilde{\beta}_t\right),$$

$$\tilde{\beta}_t^{RM} = \text{med}_{i=-k, \dots, k} \left(\text{med}_{j=-k, \dots, k, j \neq i} \frac{x_{t+i} - x_{t+j}}{i - j} \right).$$

The slope within the time window is estimated by taking repeated medians of pairwise slopes, namely an inner median with one observation being fixed, and then the median of all these inner medians. Bernholt and Fried (2003) present an algorithm for updating the repeated median (RM) filter in linear time using quadratic space when moving the time window by one step, which reduces the computational complexity of a straightforward implementation substantially.

2.2. Linear median hybrid filters

Linear median hybrid filters have been suggested to combine the good properties of linear and median filters by linear and nonlinear operations (Heinonen and Neuvo, 1987, 1988; Astola et al., 1989). They are computationally much less expensive than standard median filters and offer increased flexibility due to the use of linear subfilters Φ_j , $j = 1, \dots, M$, which are applied to the input data before taking the median of the outcomes. When the subfilters respond to a finite number of impulses, the resulting procedure is called linear median hybrid filter with finite impulse response, more briefly FIR median hybrid filter or FMH filter

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

where $\Phi_1(x_t), \dots, \Phi_M(x_t)$ are the outputs of the FIR filters. The proper choice of M and of the subfilters depend on the type of signal and on the demands specified by the operator. FMH filters can even be less biased at edges than standard median filters (Astola et al., 1989).

Kalli et al. (1985) filter blood pressures by a simple FMH filter with $M = 3$ subfilters, namely two one-sided moving-averages and the current observation x_t as central subfilter for edge preservation:

$$\Phi_1(x_t) = \frac{1}{k} \sum_{i=1}^k x_{t-i}, \quad \Phi_2(x_t) = x_t, \quad \Phi_3(x_t) = \frac{1}{k} \sum_{i=1}^k x_{t+i}.$$

Like the standard median this FMH filter implicitly takes the signal as locally almost constant since ordinary location estimates are applied.

Instead of moving averages, predictive FMH filters apply FIR subfilters for one-sided extrapolation of a trend

$$PFMH(x_t) = med(\Phi_F(x_t), x_t, \Phi_B(x_t)),$$

where $\Phi_F(x_t) = \sum_{i=1}^k h_i x_{t-i}$ and $\Phi_B(x_t) = \sum_{i=1}^k h_i x_{t+i}$. Heinson and Neuvo (1988) choose the weights h_i to obtain optimal (in the sense of mean square error MSE) predictions for a polynomial signal which is disturbed by Gaussian white noise imposing the restriction that the exact signal value is obtained in the deterministic case without noise. For a polynomial of degree zero corresponding to a constant signal this reduces to the simple FMH filter given above. For a polynomial of degree one, i.e. a linear trend $\mu_{t+i} = \mu_t + i\beta_t$, the resulting weights are

$$h_i = \frac{4k - 6i + 2}{k(k-1)}, \quad i = 1, \dots, k.$$

The combined FMH filter uses predictions of different degrees

$$CFMH(x_t) = med(\Phi_F(x_t), \Phi_1(x_t), x_t, \Phi_3(x_t), \Phi_B(x_t)),$$

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

2.3. Repeated median hybrid filters

In view of increased computational power, computation time is nowadays a much less severe constraint. In order to increase the robustness of the procedure, we can construct hybrid filters with robust instead of linear subfilters.

It is near at hand to replace the one-sided averages $\Phi_1(x_t)$ and $\Phi_3(x_t)$ by half-window medians $\tilde{\mu}_t^F = med(x_{t-k}, \dots, x_{t-1})$ and $\tilde{\mu}_t^B = med(x_{t+1}, \dots, x_{t+k})$, and the one-sided linear extrapolators $\Phi_F(x_t)$ and $\Phi_B(x_t)$ by robust linear regression estimators $RM^F(x_t)$ and $RM^B(x_t)$, with $RM^F(x_t)$ and $RM^B(x_t)$ denoting the estimate of the level at time t obtained by applying the repeated median to x_{t-k}, \dots, x_{t-1} and to x_{t+1}, \dots, x_{t+k} , respectively

$$\begin{aligned} RM^F(x_t) &= med\left(x_{t-k} + k\tilde{\beta}_t^F, \dots, x_{t-1} + \tilde{\beta}_t^F\right), \\ \tilde{\beta}_t^F &= med_{i=-k, \dots, -1} \left(med_{j=-k, \dots, -1, j \neq i} \frac{x_{t+i} - x_{t+j}}{i - j} \right), \\ RM^B(x_t) &= med\left(x_{t+1} - \tilde{\beta}_t^B, \dots, x_{t+k} - k\tilde{\beta}_t^B\right), \\ \tilde{\beta}_t^B &= med_{i=1, \dots, k} \left(med_{j=1, \dots, k, j \neq i} \frac{x_{t+i} - x_{t+j}}{i - j} \right). \end{aligned}$$

In the resulting repeated median hybrid filters we can use the central observation x_t as central subfilter for edge preservation, as in the FMH filters, or we can replace x_t by the

median $\tilde{\mu}_t$ of the time window. We investigate both possibilities and call them RMH and RMMH filters, respectively. Moreover, we will consider the robust predictive and robust combined versions

$$\begin{aligned} PRMH(x_t) &= med\left(RM^F(x_t), x_t, RM^B(x_t)\right), \\ CRMH(x_t) &= med\left(RM^F(x_t), \tilde{\mu}_t^F, x_t, \tilde{\mu}_t^B, RM^B(x_t)\right), \\ PRMMH(x_t) &= med\left(RM^F(x_t), \tilde{\mu}_t, RM^B(x_t)\right), \\ CRMMH(x_t) &= med\left(RM^F(x_t), \tilde{\mu}_t^F, \tilde{\mu}_t, \tilde{\mu}_t^B, RM^B(x_t)\right). \end{aligned}$$

3. Computation

We briefly explain some ideas for efficient implementation of the procedures. Signal extraction using a moving window allows application of algorithms for updating the results from the previous window to save computation time.

3.1. Median

An algorithm for updating (or ‘maintaining’) the standard median is given in [Cormen et al. \(2001, Chapter 14.1\)](#) This algorithm uses a red-black tree for storing the values in the current time window in a sorted order. Deletion and insertion of a value is supported in $O(\log n)$ worst case time using $O(n)$ space. In each node of the tree, the number of nodes in the left respectively right subtree is maintained. Using this information the median in the updated tree can be found in logarithmic time, by starting at the root of the tree and always branching into the appropriate subtree until the median is reached. This reduces the time needed for calculating the median from scratch using Quickselect, which has an average running time of $O(n)$.

3.2. Repeated median

[Matoušek et al. \(1998\)](#) present a randomized algorithm for computing the repeated median in expected time $O(n \log n)$. An algorithm for updating the repeated median in $O(n)$ time using $O(n^2)$ space is presented by [Bernholt and Fried \(2003\)](#) using a hammock graph for storing the data in the dual space. Here, we present a new algorithm which is based on ideas of [Stein and Werman \(1992\)](#) and which needs $O(n)$ space only. Consider the data $(t+i, x_{t+i})$, $i = -k, \dots, k$, in the current time window along with the observation times as points in the plane. According to the point-line duality the point $(t+i, x_{t+i})$ can be mapped to the dual line ℓ_i defined by $v = (t+i)u + x_{t+i}$ in (u, v) coordinates. In the dual space, we say that the point (u, v) is located *left of* the line ℓ if there exists a constant $c > 0$ such that the point $(u+c, v)$ is located on ℓ . The terms *right of*, *above* and *below* are defined analogously. The pairwise slope $(x_{t+i} - x_{t+j}) / (i - j)$ in the calculation of $\tilde{\beta}_t$ is

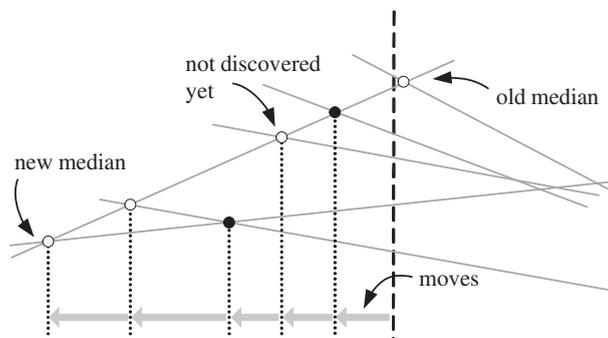


Fig. 1. After a new line is inserted, the imaginary vertical line is moved from the old median towards the new median. The black intersection points are visible to the algorithm at this point in time and stored in the heap. The white intersection points become visible if the two incident lines become neighbored in the permutation.

the u -coordinate of the intersection point (u_{ij}, v_{ij}) of the lines ℓ_i and ℓ_j . The solution with parameters $(\tilde{\mu}_t^{RM}, \tilde{\beta}_t^{RM})$ is mapped to the point $(-\tilde{\beta}_t^{RM}, \tilde{\mu}_t^{RM})$, which is the intersection of two lines. We order the lines ℓ_i according to their occurrence on an imaginary vertical line v at position $-\tilde{\beta}_t$ and denote this permutation of the lines by π . In each update step, one former line has to be deleted from π and a new one inserted into π . If the solution has changed, the permutation at the new position has to be computed. This is done by moving the imaginary line v towards the new solution point as displayed in Fig. 1.

At each intersection point visited, while moving, the two lines intersecting in this point are swapped within π . We will refer to the number of visited intersection points as S . The direction of the line movement is determined by counting for each line ℓ_i the number $c(\ell_i)$ of intersection points on ℓ_i to the left of v . We will refer to $c(\ell_i)$ as the *left counter*. Let

$$L = \#\{i \mid c(\ell_i) > \lfloor (n-1)/2 \rfloor\} \quad \text{and} \\ R = \#\{i \mid c(\ell_i) \leq \lfloor (n-1)/2 \rfloor\}.$$

A new solution $(-\tilde{\beta}_{t+1}^{RM}, \tilde{\mu}_{t+1}^{RM})$ is attained if $L = R$. In the case that two inner medians are located on the same intersection point they are identical and only $|L - R| \leq 1$ can be reached since the corresponding left counters $c(\cdot)$ are both increased or decreased by one if this intersection point is passed. Otherwise, if $L < R$ we need to move to the right, while we need to move to the left if $L > R$.

In order to maintain the permutation π we store the lines in a dictionary T , e.g. a red-black tree or AVL-tree, illustrated in Fig. 2 and described in Cormen et al. (2001). As the depth of such trees is bounded by $O(\log n)$, these data structures support insertion, deletion and search in $O(\log n)$ time. The lines are sorted in the order of their intersection with v . The line intersecting at the smallest v -coordinate is stored as the leftmost node.

Insertion (deletion) of a new (old) line in an update step involves updating the left counters $c(\cdot)$. Call the position of the inserted (deleted) line π_i . As the inserted (deleted) line ℓ^* has the largest (smallest) slope, all lines stored in π_1, \dots, π_{i-1} (π_{i+1}, \dots, π_n) intersect ℓ^* on

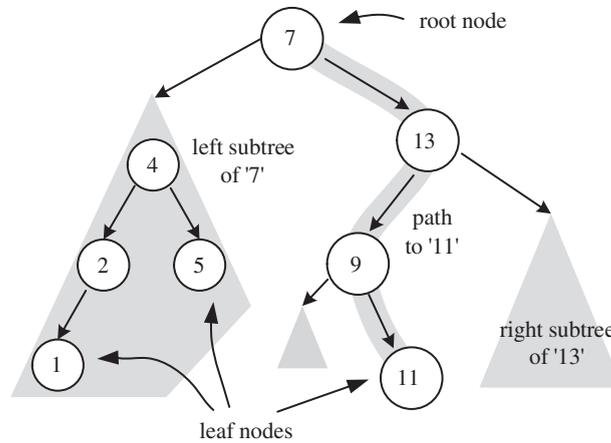


Fig. 2. Each data item is stored in a node of the binary search tree. Each node has pointers to left and/or right child node. If neither left nor right child exists, the node is called a leaf. The depth of the tree is the number of nodes on the longest path from the root to a leaf.

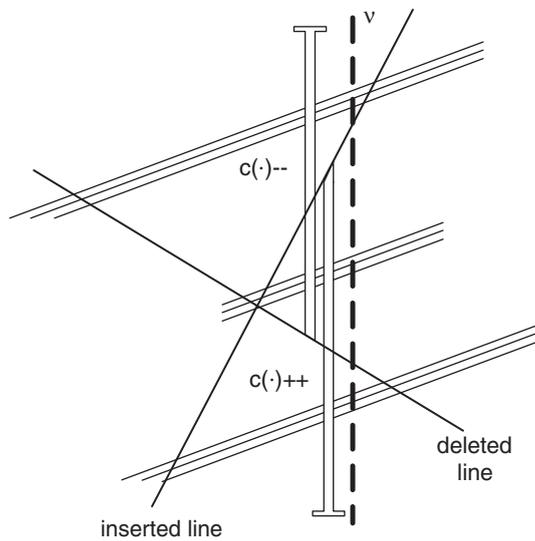


Fig. 3. The left counters $c(\cdot)$ of the lines intersecting the inserted line left of v have to be increased. The left counters $c(\cdot)$ of the lines intersecting the deleted line left of v have to be decreased.

the left-hand side. Therefore the left counters $c(\cdot)$ for all lines that are stored left (right) of ℓ^* in the tree have to be increased (decreased) by one, as displayed in Fig. 3.

The problem is that there are $O(n)$ left counters changing their value and that the correct L has to be computed. In order to achieve a running time of $O(\log n)$, we store an update value $u(v)$ in each node. The value $u(v)$ should be added to the left counters of all nodes in

the subtree of v . Instead of performing this addition directly, an operation accessing a node v has to track the value

$$\Delta U(v) = \sum_{v' \in \text{path}(v)} u(v'),$$

where $\text{path}(v)$ is the set of nodes located on the path from v to the root of the tree including the node v itself. In the node v a value $c'(v)$ is stored. The correct left counter is obtained by $c(v) = c'(v) + \Delta U(v)$.

Here is a complete list of variables needed to perform the operations efficiently:

- $c(v)$: The left counter of the line stored in the node v .
- $u(v)$: The update value of the node v .
- $w_a(v)$: The maximum value of a $c(\cdot)$ less or equal than $\lfloor (n-1)/2 \rfloor$, taken over all the nodes in the subtree of v . Define $w_a(v) = -\infty$ if all $c(\cdot)$ in the subtree are greater than $\lfloor (n-1)/2 \rfloor$.
- $a(v)$: The number of $c(\cdot)$ in the subtree of v which are equal to $w_a(v)$.
- $w_b(v)$: The minimum value of a $c(\cdot)$ greater than $\lfloor (n-1)/2 \rfloor$, taken over all the nodes in the subtree of v . Define $w_b(v) = \infty$ if all $c(\cdot)$ in the subtree are less or equal than $\lfloor (n-1)/2 \rfloor$.
- $b(v)$: The number of $c(\cdot)$ in the subtree of v which are equal to $w_b(v)$.
- $\text{size}(v)$: The number of nodes in the subtree of v .

The variables $u(v)$, $a(v)$, $b(v)$, and $\text{size}(v)$ are directly stored in each node v , the other variables are stored as $c'(v) = c(v) - \Delta U(v)$, $w'_a(v) = w_a(v) - \Delta U(v)$, and $w'_b(v) = w_b(v) - \Delta U(v)$.

The new line is inserted as a node v_{new} into the search tree T . Consider the set $\text{path}^\ell(v_{\text{new}})$ and the left boundary $B^\ell(v_{\text{new}})$, defined as

$$\text{path}^\ell(v_{\text{new}}) = \{v_{\text{new}}\} \cup \{v \mid v \in \text{path}(v_{\text{new}}) \text{ and } v \text{ is left of } v_{\text{new}}\}$$

$$B^\ell(v_{\text{new}}) = \left\{ v' \mid v' \text{ is a left child of } v, v \in \text{path}^\ell(v_{\text{new}}) \right\}.$$

The set $\text{path}^r(\cdot)$ and the right boundary $B^r(\cdot)$ are defined in similar way. The definitions are illustrated in Fig. 4.

During the insertion we increase $u(v)$ in all nodes $v \in B^\ell(v_{\text{new}})$ by one and we increase $w_a(v)$, $w_b(v)$ and $c(v)$ in all nodes $v \in \text{path}^\ell(v_{\text{new}})$ by one. The old line is deleted as the node v_{old} from the search tree. During the deletion we decrease $u(v)$ in all nodes $v \in B^r(v_{\text{old}})$ by one and decrease $w_a(v)$, $w_b(v)$ and $c(v)$ in all nodes $v \in \text{path}^r(v_{\text{old}})$ by one. By this, the variables in the nodes affected by the insertion and deletion are updated in the correct way. As the inserted line has the greatest slope, the variable $c(v_{\text{new}})$ is the number of the nodes stored left of v_{new} in the tree. Therefore the initial value of $c(v_{\text{new}})$ is

$$\left| \text{path}^\ell(v_{\text{new}}) \right| - 1 + \sum_{v \in B^\ell(v_{\text{new}})} \text{size}(v).$$

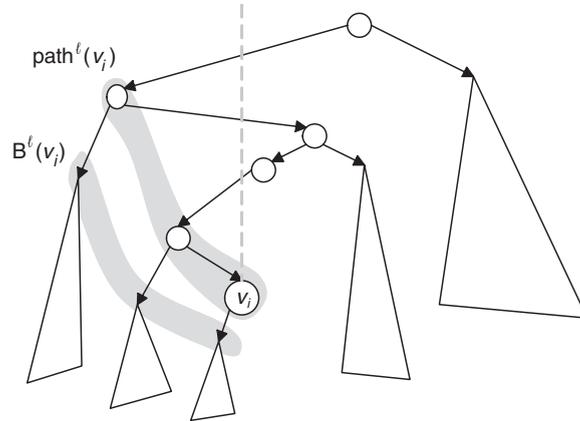


Fig. 4. The variables in a node v left of v_i are adjusted by either manipulating v directly if $v \in \text{path}^l(v_i)$ or indirectly by modifying the update values in the nodes of the set $B^l(v_i)$.

Define $P = \text{path}^l(v_{\text{new}}) \cup \text{path}^r(v_{\text{old}})$. L is increased by ΔL , which is the number of nodes v where $w_a(v)$ or $w_b(v)$ exceeds $\lfloor (n-1)/2 \rfloor$

$$\Delta L = - \sum_{v \in P \text{ and } w_a(v) > \lfloor (n-1)/2 \rfloor} a(v) + \sum_{v \in P \text{ and } w_b(v) \leq \lfloor (n-1)/2 \rfloor} b(v).$$

After this change the values of $w_a(\cdot)$ and $w_b(\cdot)$ may violate their definitions, and so to update the values of $w_a(\cdot)$ and $w_b(\cdot)$, we call for every $v \in P$ Procedure 1. This procedure visits all nodes on the ΔL paths towards nodes where the value of the left counter $c(\cdot)$ has exceeded $\lfloor (n-1)/2 \rfloor$, thus the time for inserting and deleting a line is $O(\Delta L \cdot \log n)$.

Procedure 1 Update(v)

Input:

A node v . Let v_1 and v_2 be the two child nodes.

begin

(The case that one of v_1 and v_2 does not exist is not explicitly mentioned)

if $w_a(v) > \lfloor (n-1)/2 \rfloor$ **then**

$w_b(v) := w_a(v)$ and $b(v) := a(v)$

if v is a leaf **then**

$a(v) := 0$ and $w_a(v) := -\infty$.

else

Call Update(v_1) and Update(v_2)

Call Recompute(v)

end if

end if

The case $w_b(v) \leq \lfloor (n-1)/2 \rfloor$ is handled analogously.

end

After inserting the new line and deleting the old one, we move the vertical line to the new optimum. Note that the next point to be visited, when moving the vertical line, is always an intersection point of two lines which are neighbors in the current permutation π . In order to find the next intersection point in $O(\log n)$ time with respect to the ordering of the u -coordinate, we store the intersection points on both sides of the current imaginary line v in two priority queues Q_L and Q_R , e.g., organized as heaps (Cormen et al., 2001, Chapters 19 and 20). When moving v to the next intersection point two lines are swapped, which creates two new neighborhoods and destroys two old ones. Depending on the location of new (old) intersection points with respect to v they need to be inserted (deleted) in the priority queues, which can be done in $O(\log n)$ time. The next intersection point in the priority queue can also be found in $O(\log n)$ time. A similar technique is used in Cole et al. (1989).

The swap of two lines on v requires to swap two nodes in the tree. Name the nodes, where the two lines are stored in, v_i and v_j . Taking $\Delta U(v_i)$ and $\Delta U(v_j)$ into account, the variables of both nodes can be swapped easily. The left counters $c(\cdot)$ have to be altered by one, depending on the direction of the movement. As the variables in the nodes v_i and v_j may have changed, the variables of the nodes $v \in \text{path}(v_i) \cup \text{path}(v_j)$ have to be recomputed using Procedure 2. The value ΔL can be obtained as described above. Therefore, the swap of two nodes in the tree can be done in time $O(\log n)$.

Procedure 2 Recompute(v)

Input:

A node v . Let v_1 and v_2 be the two child nodes.

begin

$$w_a(v) := \max\{w_a(v_1), w_a(v_2)\}$$

$$a(v) := \begin{cases} a(v_1) + a(v_2) & \text{if } w_a(v_1) = w_a(v_2) \\ a(v_1) & \text{if } w_a(v_1) > w_a(v_2) \\ a(v_2) & \text{otherwise} \end{cases}$$

if $(c(v) = w_a(v))$ **then**

$a(v) := a(v) + 1$

end if

if $(c(v) \leq \lfloor (n-1)/2 \rfloor$ and $c(v) > w_a(v))$ **then**

$w_a(v) := c(v)$ and $a(v) := 1$

end if

$w_b(v)$ and $b(v)$ are computed analogously.

end

The overall runtime of one update is as follows: As the algorithm has to make at least $\Delta L/2$ moves until the new optimum is reached and each move of the vertical line v can be computed in time $O(\log n)$, the overall runtime is bounded by $S \cdot O(\log n)$, where S is the number of moves or visited intersection points while moving v . In Section 6.1 we study S empirically.

The algorithm described above allows updating the regression slope. For updating the extracted signal value we use the fact that

$$RM(x_t) = \text{med}_{i=-k, \dots, k} \left(\text{med}_{j=-k, \dots, k, j \neq i} \frac{jx_{t+i} - ix_{t+j}}{i - j} \right)$$

(Stein and Werman, 1992) and we modify the algorithm in such a way that we use an imaginary horizontal line instead.

3.3. Repeated median hybrid filters

The calculation of the repeated median hybrid filters requires obtaining the median of three or five values and updating the backward extrapolating filter(s) as described above, which are stored and reused for calculating the forward extrapolating filters. Using a FIFO queue, which can be implemented as a linked list, the new element $(RM^B(x_t), \tilde{\beta}_t^B)$ is inserted at the head of the list, and $RM^F(x_t)$ is obtained from the previous $(RM^B(x_{t-k-1}), \tilde{\beta}_{t-k-1}^B)$ at the end of the list as

$$RM^F(x_t) = RM^B(x_{t-k-1}) + (k + 1)\tilde{\beta}_{t-k-1}^B$$

before deleting $(RM^B(x_{t-k-1}), \tilde{\beta}_{t-k-1}^B)$. All this can be done in $O(1)$ time. The value of $\tilde{\mu}_t^F$ is obtained immediately from such a list of the $\tilde{\mu}_t^B$ values. Using the former computations saves a constant factor. Nevertheless, the running time is dominated by the computation of $RM^B(x_t)$.

3.4. FMH filters

Besides calculating the median of three or five values in $O(1)$ time, updating the FMH filters requires maintaining the weighted sums of the observed values. Replacing a value in an ordinary sum for obtaining an updated arithmetic mean can be done in $O(1)$ time. For $\Phi_B(x_t)$ we use

$$\begin{aligned} \Phi_B(x_t) - \Phi_B(x_{t-1}) &= \sum_{i=1}^k h_i x_{t+i} - \sum_{i=1}^k h_i x_{t+i-1} \\ &= h_k x_{t+k} - h_1 x_t + \frac{6}{k(k-1)} \sum_{i=1}^{k-1} x_{t+i}. \end{aligned}$$

This allows the difference to be updated in $O(1)$ time, which is then added to the previous value $\Phi_B(x_{t-1})$.

4. Analytical properties

For a theoretical analysis of the filtering procedures we use the components model (1) and study the behavior in a single time window centered at zero considering the data points $x_{-k}, \dots, x_0, \dots, x_k$ with the same k for all filters.

4.1. Equivariance and invariance

Equivariances and invariances are basic properties of statistical procedures. Location equivariance means that adding a constant to all observations (in the window) changes the estimate accordingly. Similarly, scale equivariance means that multiplication of all observations with a constant changes the output in the same way. All the previous procedures have both properties.

For tracking trends, sometimes linear trend preservation is required, which means that the filter output is identical to the observations if these lie on a straight line. This property is weak and is inherent to all the previous methods. However, it does not consider observational noise, which can hamper the tracking of trends. Another property is preferable instead, which we might call (linear) trend invariance: The extracted signal should not depend on a linear trend as long as the central level remains fixed. This means, when varying the trend in the window, replacing x_{-k}, \dots, x_k by $x_{-k} - k\beta, \dots, x_{-1} - \beta, x_0, x_1 + \beta, \dots, x_k + k\beta$, the estimated central level should remain the same.

The RM, the predictive FMH and the predictive RMH are the only trend invariant procedures among those presented in Section 2. The trend invariance of the RM follows from the regression equivariance of the underlying regression estimator. When adding a constant trend to some data, the backward and forward subfilters of the PFMH and PRMH still result in the same extrapolations, and hence the median is taken from the same values as before.

The standard median is not trend invariant. Consider the case $k = 1$ and the observations 1, 0, -3 with median zero. Adding a trend with slope $\beta = 2$ we obtain the sample $-1, 0, -1$ with median -1 , i.e. the outcome depends on the slope. In the same way, FMH and RMH filters are not trend invariant when subfilters are used which are not invariant like medians or one-sided averages.

4.2. Noise-free situation—a best case analysis

The tracking of trends, the preservation of abrupt shifts and the removal of impulsive noise (outliers) are essential properties of robust filters. The best possible performance of a procedure can be inspected in a noise-free situation, where $\sigma^2 = 0$. The standard median preserves a level shift exactly in this idealized setting, and it removes up to k subsequent spikes completely when using the window width $n = 2k + 1$ and the signal is constant. It even allows the recovery of monotonic trends, but it preserves a shift in a trend period only if the shift has the same direction as the trend. Otherwise the shift gets blurred, and spikes within a trend cause further undesirable effects.

As opposed to standard medians, FMH filters do not delay a level shift when there is an impulse in the opposite direction within k time points after the shift. However, they suffer from smearing then as the height of a shift and a constant signal value close to the

shift change (Heinonen and Neuvo, 1987). Like all methods considered here, FMH filters recover linear trends without noise exactly. The combined FMH removes a single spike and preserves a shift exactly only if the signal is constant or the outlier (shift) has the same direction as the trend. A negative spike (shift) in an upward trend produces smearing. The predictive FMH removes a single spike and preserves a shift exactly within a linear trend irrespective of the directions due to its invariance.

The RM can remove $k - 1$ spikes completely within a window as then the median slope for each of the $k + 2$ clean data points is calculated from $k + 1$ pairs of clean observations. This number is slightly smaller than for the median in case of a constant signal, but for the RM it does not depend on a trend.

The repeated median hybrid filters improve on the FMH filters with respect to the removal of spikes as they can resist up to $\lfloor k/2 \rfloor$ subsequent outliers without any effect. For the predictive RMMH this number is even $2\lfloor k/2 \rfloor - 1$. In order to see this, we note that for the predictive filters two subfilters giving the exact signal value are sufficient to guarantee that this is also true for the hybrid filter, while for the combined filters it is sufficient that there is one more filter with a smaller and one with a larger outcome. Furthermore, each of the extrapolating RM filters gives the exact signal value if it is influenced by at most $\lfloor k/2 \rfloor - 1$ spikes. Now, if there are only $\lfloor k/2 \rfloor$ spikes on the right hand side of the center, the forward extrapolating subfilter and the central subfilter and hence the RMH and the RMMH give the exact signal value (as the outcome of one of the half-window medians in the combined filters will be at most and the other one at least the signal value). If not all of the $\lfloor k/2 \rfloor$ spikes are on the same side of the center both extrapolating RM filters give the exact signal value. However, more than $\lfloor k/2 \rfloor$ negative outliers on the right of the center in an upward trend can change the backward extrapolating filters, and force the combined RMH and RMMH to become smaller than the signal value. Outliers at positions $0, \dots, \lfloor k/2 \rfloor$ change the central subfilter and the backward extrapolating RM filter in the predictive RMH arbitrarily. The predictive RMMH, however, gives the exact signal value as long as there are at most $\lfloor k/2 \rfloor - 1$ outliers on one side of the center and k spikes altogether because then one of the extrapolating filters and the median give the correct value. This is guaranteed if there are at most $2\lfloor k/2 \rfloor - 1$ subsequent spikes. Note that the performance of the combined RMMH depends on the sign of the trend and the spikes: The half-window medians are still to the same side of the signal value as in the case without spikes when being affected by $\lfloor k/2 \rfloor$ spikes in the same direction as the trend, e.g. if both are positive. Therefore, the combined RMMH can remove more spikes completely, namely $\lfloor k/2 \rfloor + 1$, if they are in the same direction as the trend.

Furthermore, the predictive RMH and RMMH preserve shifts exactly. For the combined RMH and RMMH this is true only if the shift is in the same direction as the trend, just like for the combined FMH. Table 1 summarizes these results.

Table 1
Number of subsequent spikes removed from a linear trend without noise

SM	RM	PFMH	CFMH	PRMH	CRMH	PRMMH	CRMMH
0	$k - 1$	1	0	$\lfloor k/2 \rfloor$	$\lfloor k/2 \rfloor$	$2\lfloor k/2 \rfloor - 1$	$\lfloor k/2 \rfloor$

4.3. Continuity—small deviations

The results in the previous subsection are for the case without observational noise. Davies (1993) suggests application of continuous, particularly of Lipschitz continuous procedures for restricting the influence of small changes in the data due to, e.g. rounding or a small amount of observational noise.

All the filters investigated here are Lipschitz continuous. The median, like all order statistics, is Lipschitz continuous with constant 1 as changing every observation by less than a value δ changes the median at most by δ . As every linear filter is Lipschitz continuous with constant $\max |h_i|$, the maximal absolute weight, an FMH filter is Lipschitz continuous with constant $\max |h_i^j|$, the maximal absolute weight given by any subfilter. The repeated median is Lipschitz continuous with constant $2k + 1$: The slope estimate changes at most by 2δ when changing each observation by at most δ , and hence the level estimate is bounded not to change more than $(2k + 1)\delta$ as none of the trend corrected observations changes more. Accordingly, the repeated median hybrid filters are Lipschitz continuous with the same constant $2k + 1$.

4.4. Breakdown—a worst case analysis

We apply the concept of breakdown point to measure the amount of contamination in a single time window which makes the extracted signal value worthless, i.e. arbitrarily large or small. The finite-sample breakdown point is obtained by changing a fraction of the data points, thus getting a contaminated sample. It is then defined to be the smallest fraction of contaminated points such that the estimate can take on any value.

When applying the median to $n = 2k + 1$ observations x_{-k}, \dots, x_k , the breakdown point is $(k + 1)/n$ telling us that at least half of the sample needs to be ‘outlying’ in order to make the level approximation completely unreliable.

The breakdown point for the level estimation by each of the FMH filters is only $2/n$. The outcome of the central subfilter and of the subfilters for forward extrapolation and hence the outcome of the FMH filter can be made arbitrarily large by replacing the observation in the center and another observation which has positive weight for all predicting subfilters, e.g. the observation just to the left of the center. Hence, two subsequent spikes cause the same, only somewhat dampened pattern in the extracted signal.

Breakdown of the RM affords contamination of at least k observations, while for the predictive RMH the observation in the center and at least $\lfloor k/2 \rfloor$ observations on one side of it need to be replaced. For the combined RMH, one more observation is necessary to destroy also the corresponding half-window median. Therefore, the breakdown points are $(\lfloor k/2 \rfloor + 1)/n$ and $(\lfloor k/2 \rfloor + 2)/n$ corresponding to replacing e.g. 6 and 7 out of $n = 21$ observations, respectively. For the PRMMH, at least $\lfloor k/2 \rfloor$ observations both on the left-hand side and on the right-hand side of the center need to be replaced to destroy both extrapolating subfilters. Thus, the breakdown point is $2\lfloor k/2 \rfloor/n$ meaning that, e.g. 10 bad observations are necessary to make the PRMMH break down if $n \in \{21, 23\}$. For the CRMMH, destruction of both subfilters to one side affords replacement of $\lfloor k/2 \rfloor + 1$ observations. Destruction of the linear extrapolating filter on the other side needs another

Table 2
Breakdown points in a single time window

SM	RM	FMH	PRMH	CRMH	PRMMH	CRMMH
$\frac{k+1}{n}$	$\frac{k}{n}$	$\frac{2}{n}$	$\frac{\lfloor k/2 \rfloor + 1}{n}$	$\frac{\lfloor k/2 \rfloor + 2}{n}$	$\frac{2\lfloor k/2 \rfloor}{n}$	$\frac{2\lfloor k/2 \rfloor + 1}{n}$

$\lfloor k/2 \rfloor$ modifications, summing up to $2\lfloor k/2 \rfloor + 1$ out of n . This is 11 both for $n = 21$ and 23, which is the same as for the median or slightly smaller. Table 2 summarizes these results.

5. Simulation experiments

We performed some Monte Carlo experiments for a comparison of the filter procedures in basic data situations and in the presence of observational noise. We used the components model (1) with $\mu_t = \mu_0 + t\beta$ generating the observational noise from an autoregressive model of order one, AR(1), $u_t = \phi u_{t-1} + \varepsilon_t$, where the innovations ε_t form Gaussian white noise with mean zero and variance σ^2 . AR(1) models are a convenient choice for modelling autocorrelations. Because of location and scale equivariance we set $\mu_0 = 0$ and $\sigma^2 = 1$ without loss of generality. The results for the RM, the PFMH and the PRMH do not depend on the choice of β since these methods are trend invariant.

The suitable choice of the window width $n = 2k + 1$ depends on the application, i.e. on the situations a filtering procedure needs to handle. For removing patches of subsequent outliers we must choose k sufficiently large, while upper limits for k are imposed by the duration of periods in which trends can be assumed to be approximately linear and by the admissible time delay. We chose a width of $n = 21$ and investigated the attenuation of Gaussian noise, the preservation of level shifts and the removal of outlier patches.

5.1. Efficiency

Firstly we compared the attenuation of Gaussian noise. All methods are unbiased then because of symmetry. We simulate 50,000 windows for each slope $\beta = 0, 0.05, \dots, 0.5$ and correlation $\rho \in \{0, 0.6\}$ to compare the efficiencies of the filters as measured by the percentage MSE of the arithmetic mean relatively to each method, see Fig. 5.

The standard median loses a lot of efficiency with increasing slope, while the RM is almost as efficient as the median for a constant signal throughout. The one-sided averages or medians increase the efficiencies of the combined filters substantially for a constant signal, but this gain gets lost with increasing slope as the filters become dominated by the subfilters extrapolating a linear trend. The combined and the predictive versions become nearly undistinguishable for a moderate trend with $\beta = 0.2$ in case of zero autocorrelations, while positive autocorrelations slow this decrease down. Each RMMH filter is more efficient than the corresponding FMH filter, while the corresponding RMH filter is slightly less

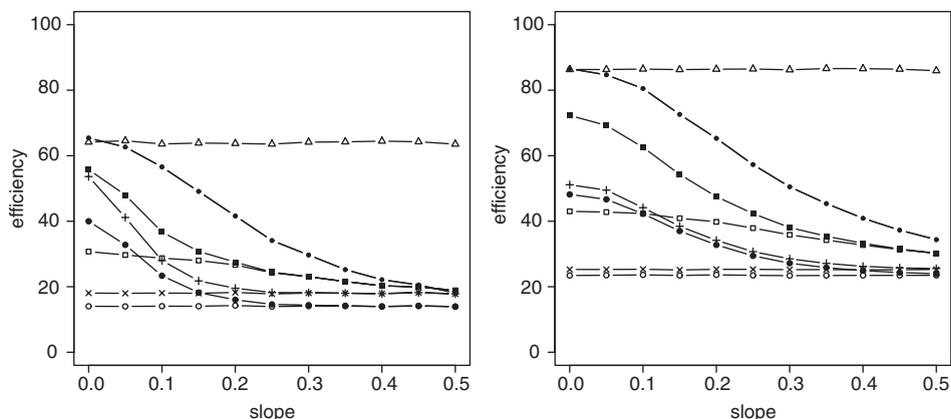


Fig. 5. Relative efficiencies for Gaussian noise, autocorrelations $\rho = 0.0$ (left) and $\rho = 0.6$ (right): median (\cdot), RM (Δ), PFMH (\times), CFMH ($+$), PRMH (\circ), CRMH (\bullet), PRMMH (\square) and CRMMH (\blacksquare).

efficient. Positive autocorrelations increase the relative efficiencies of the (repeated) median-based filters substantially.

5.2. Preservation of shifts

Level shifts need to be localized and tracked as exactly as possible. In order to investigate the preservation of a shift within a linear trend we simulated signals with $\beta \in \{-0.5, 0, 0.5\}$ and generated data situations which mimic the intrusion of a shift into the window replacing an increasing number 1, 2, ..., 10 of observations at the end of the window by additive outliers of size $s \in \{2, 4, \dots, 10\}$. For each setting we simulated 2000 windows and calculated the bias, the standard deviation and the root of the mean square error RMSE for all methods. We found larger outliers to have stronger effects and compared the maximal bias, standard deviation and RMSE for a fixed number of outliers of size 10, i.e. observations affected by a shift of this size.

Fig. 6 depicts the RMSE for a constant signal ($\beta = 0$) and for a steep upward or downward trend ($\beta = \pm 0.5$). Like in most of the following comparisons the RMSE is dominated by the bias, while the standard deviations are much less sensitive to an increasing number of outliers. We restrict to the RMSE therefore. If the shift and the trend have the same direction, e.g. if both are upwards, the RM is the only filter to become strongly biased, namely when more than 5 (25%) observations are affected. The median and the hybrid methods perform very well then, the FMH and the RMH filters being even better than the median in case of a constant signal. This picture changes substantially if the shift and the trend have opposite directions. As outlined in Section 4.2, the median smooths the shift then. The CRMMH resists up to 4 (20%), the CRMH and the PRMMH up to 6 (30%) and the CFMH up to 7 (34%) affected observations. Only the PFMH and the PRMH do not smooth a shift at all, not even during a trend, due to their trend invariance. In the case of positive autocorrelations $\phi = 0.6$, we find an increase of variance for all methods, but the results are essentially the same otherwise.

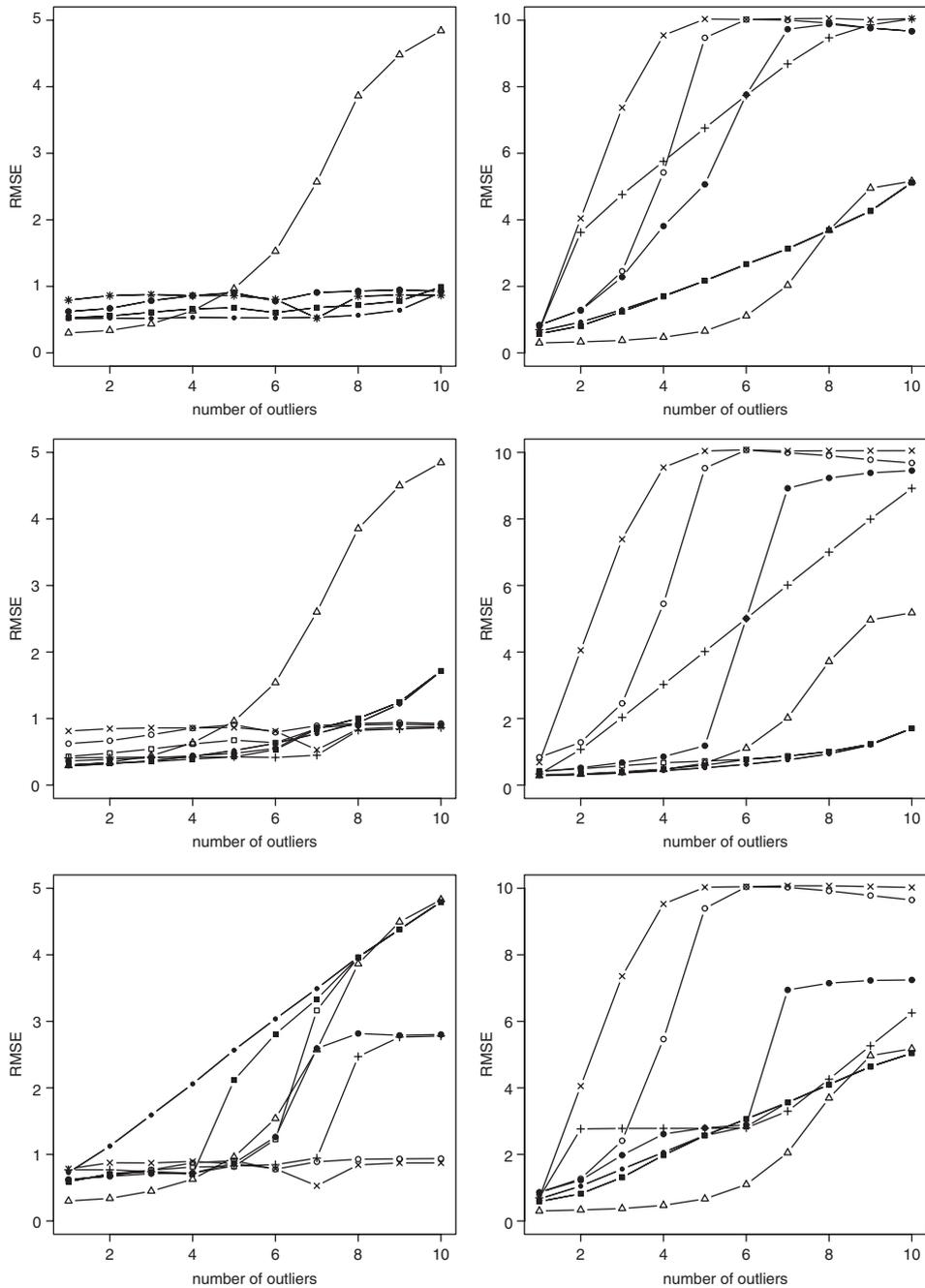


Fig. 6. RMSE for the intrusion of a level shift (left) and for an outlier patch just to the right of the center (right), slope $\beta = 0.5$ (top), $\beta = 0.0$ (center) and $\beta = -0.5$ (bottom): median (·), RM (Δ), PFMH (\times), CFMH (+), PRMH (\circ), CRMH (\bullet), PRMMH (\square) and CRMMH (\blacksquare).

5.3. Removal of patchy impulsive noise

We also investigated the removal of impulsive noise (outliers). In applications like intensive care, we find irrelevant patches of several subsequent outliers of similar size, which should not influence the filter outcome. Sometimes even more than one outlier patch occurs within a short time period.

For some of the methods the location of the outliers is essential for the effect they have. The hybrid methods are to be expected to perform well when all outliers are located on the same side of the window since then only one and two subfilters are affected, respectively. The FMH and RMH filters are influenced most if the central observation is outlying. Since the window is moved through the data an outlier patch will be found at any location in the window at some time point. We considered several settings with either one outlier patch in the center of the window or with two separated outlier patches. The situation with one outlier patch at the start or the end of the window has been treated implicitly in Section 5.2. In each situation we replace an increasing number 1, 2, . . . , 10 of observations by additive outliers of size $s \in \{2, 4, \dots, 10\}$ and calculate the RMSE, the bias and the standard deviation from 2000 simulation runs for each setting.

Fig. 6 reports the RMSEs for one outlier patch next to the center of the window, where we replace the observations at time points 0, 1, 2, . . . , 9 by positive outliers. We find this situation to be a worst case for most of the methods. The RMSEs are again mainly due to bias, only the PRMH shows an increase of variance for 3–5 outliers. The PFMH is strongly biased already in case of 2 outliers, while the PRMH resists them much better. The performance of the other methods depends on the underlying slope. Two positive outliers within an upward trend damage the CFMH considerably, while the CRMH resists 2 (10%) outliers in case of a trend and 5 (25%) outliers in case of a constant signal satisfactorily. The RMMH filters perform very similarly to the median showing a weakly increasing bias. The best method in case of a trend is the RM, which resists about 6 (30%) outliers in the center.

When the time window is moved forward so that the outlier patch gets right in the center of the window at time points 0, 1, -1, 2, -2, . . . , 5 (not shown here), the RMH filters perform somewhat better than before. Only the PRMMH is strongly biased in case of more than 5 (25%) outliers even in case of a constant signal. In a trend period, the RM is again the most resistant method.

The repeated median hybrid filters also resist two patches of additive outliers within half a window width distance at $t=0$, 10, -1, 9, -2, 8, -3, 7, -4, 6 much better than the FMH filters, see Fig. 7. If both patches have the same sign, the PRMH resists up to four outliers, which is twice as many as the FMH filters. The other filters perform very well in case of a constant signal, but only the RM removes more than six positive outliers from a negative trend. If the patches have different sign (not shown here), the PFMH is strongly affected already by two outliers, just like the CFMH in case of a trend.

Regarding situations with an outlier patch at each end of the window we replaced the observations at time points -10, 10, -9, 9, . . . , -6, 6 by additive outliers, see Fig. 7. Outliers at both ends of the window have a smaller effect on the hybrid filters than outliers in the center. If both patches have the same sign, it needs more than three outliers for the PFMH and more than 7 for the PRMH and the PRMMH to become considerably affected,

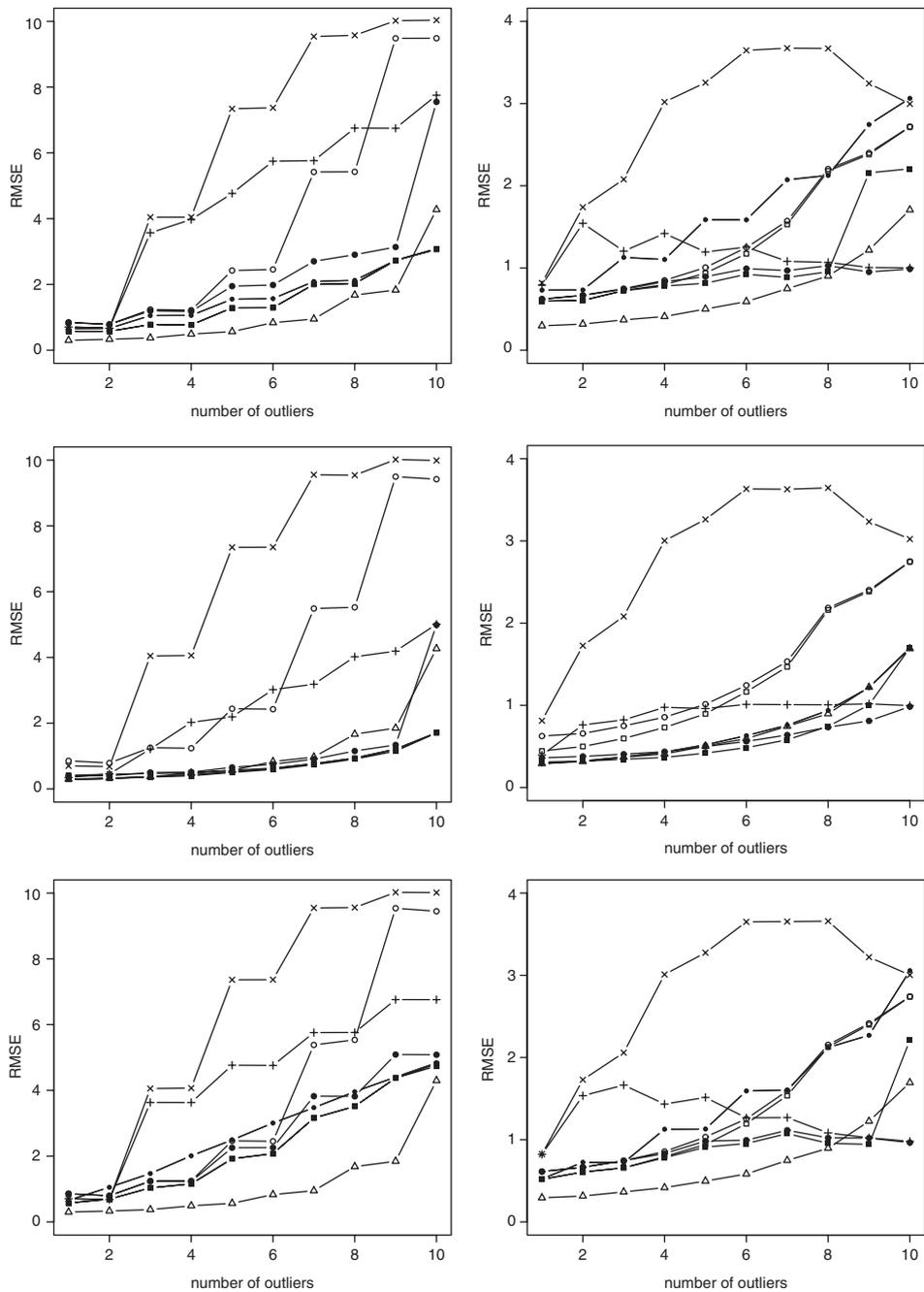


Fig. 7. RMSE for positive patches lagged by a half (left) and a full window width, slope $\beta = 0.5$ (top), $\beta = 0.0$ (center) and $\beta = -0.5$ (bottom): median (\cdot), RM (Δ), PFMH (\times), CFMH (+), PRMH (\circ), CRMH (\bullet), PRMMH (\square) and CRMMH (\blacksquare).

which is similar to the median in a trend period. The CRMMH improves on the median as it accommodates well up to 8 outliers within a trend. The RM and the CRMH even resist up to almost 50% outliers without being significantly affected. Outlier patches with different sign cause much smaller problems here as the bias effects of the outliers nearly cancel each other out.

6. Application to time series

We finally applied the procedures to time series with an underlying complex signal. Again we chose the window width as $n = 21$. At the start and the end of the series we appended the first and the last value, respectively, for signal extraction except for the RM, for which we use the regression lines in the first and the last window.

6.1. Simulated time series

We simulated a time series of length 300 with an underlying signal containing constant as well as trend periods and three shifts, which was all overlaid by $N(0, 1)$ white noise, see Fig. 8. 5% of the observations were replaced by outliers of size -10 , which were positioned as three isolated, three pairs and two triples of outliers. The time points for the patches were chosen at random. The RM preserves the shift within the constant signal worse, but those within trends better than the median, and it is smoother during the trends. The FMH and the PRMH are more variable, but they preserve the shifts much better. The PFMH shows similar spikes as the time series when these have a length of at least two. The PRMH is only mildly affected by these and we find it to perform similarly to a standard median with a suitably chosen short window width. However, for the latter the choice of n is more critical as it tracks outlier patches with a length more than half this width, while the PRMH accommodates them mostly. The CRMH and the PRMMH preserve the shifts and the extremes better than the RM, but they are more volatile.

We use this example for getting information on the number S of moves needed for updating the repeated median when using the new algorithm, which only needs linear space. Fig. 9 shows percentiles of S obtained from 1000 time series generated from the same model as described above, with the outliers at the same positions, but different observational noise. It seems that S increases linearly with increasing width n for a steady state or a trend period, but quadratically when a shift or a local extreme occur.

To shed more light on the influence of the window width n , we analyze time periods representing a linear trend, including outliers, being affected by a level shift, or containing a slope change. We vary the window width within $n \in \{21, 31, 51, 101, 201, 501, 1001\}$. Fig. 10 depicts the mean number of moves needed for updating the repeated median in each of these situations, calculated from averaging across $\lfloor n/2 \rfloor + 1$ time points around the specific pattern and 1000 independent time series. The slope of a fitted regression line directly corresponds to the estimated power of n since the plot is on a logarithmic scale. A straight line gives a satisfactory description of the influence of $\lg(n)$ on $\lg(S)$ in all situations, except for the occurrence of outliers, for which we observe some alternation. During a linear trend, the number of moves needed for each update indeed seems to increase linearly in n ,

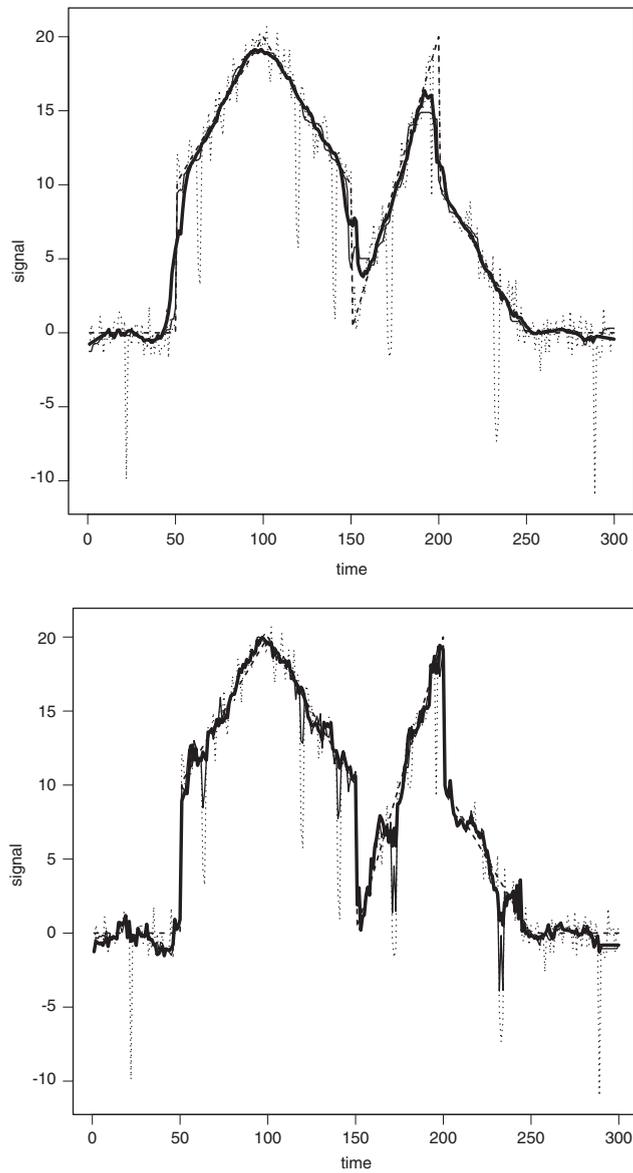


Fig. 8. Simulated time series with outlier patches of length up to three (dotted) and underlying signal (dashed). Top: SM (thin solid), RM (bold solid). Bottom: PFMH (thin solid), PRMH (bold solid).

and the occurrence of a moderate number of outliers apparently does not change this order. However, we get further evidence that the increase close to a level shift or a slope change is stronger since the slope of a fitted regression line is significantly larger than 1. Repeating this experiment for different noise variances, we found that this additional effort may be

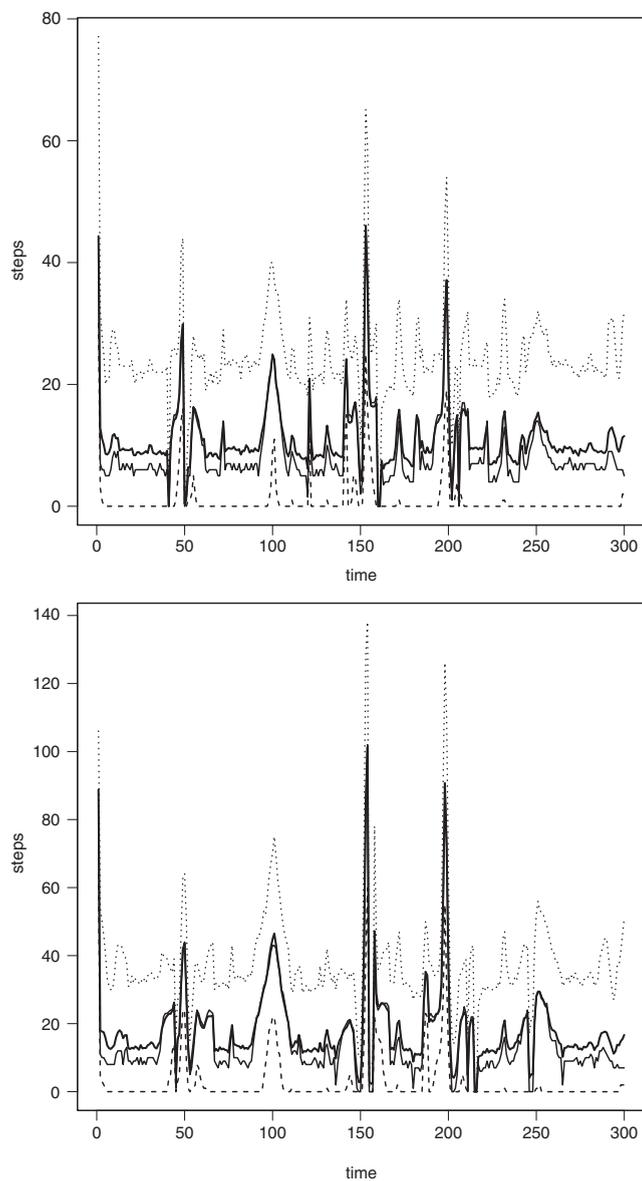


Fig. 9. Number S of moves needed for updating the repeated median at each time point, window width $n = 21$ (top) and $n = 31$ (bottom), upper 10 percentage point (dotted), median (thin solid), mean (bold solid), and lower 10 percentage point (dashed).

spread over a number of time points, or it may concentrate in a single update if the variance is small relative to the size of the shift or change. Under the assumption that level shifts or

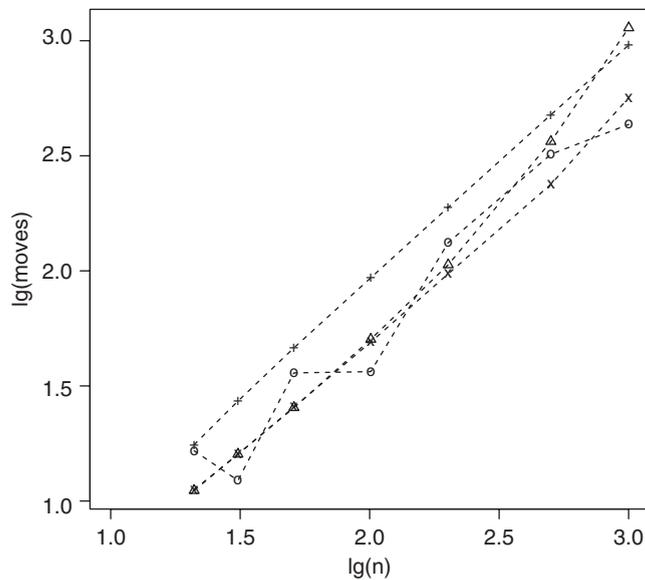


Fig. 10. Average number S of moves needed for updating the repeated median as a function of the window width, logarithms to the power of 10: constant trend (x), outliers present (o), at a level shift ($+$) and at a slope change (Δ).

slope changes occur at most every $O(n)$ time points, we claim from our experiments that the running time is $O(n \log n)$ in an amortized sense.

6.2. Real-time series

We analyzed two time series representing arterial blood pressure, see Figs. 11 and 12. Again the repeated median (not shown here) results in the smoothest signal, but like the median it underestimates the extremes. The PFMH and the PRMH preserve the local extremes and track long fluctuations, and the PRMH is not affected by short patches. The PRMMH and the CRMH are in between the PFMH, the median and the repeated median. The CRMH performs very well in the second example.

7. Summary

As expected none of the filters is overall optimal. Standard median filters are highly robust and preserve edges in a constant signal, but they worsen in trend periods. Repeated median filters are not affected by trends and attenuate both Gaussian and spiky noise well. However, they smooth shifts and extremes considerably. Predictive FMH filters are also not affected by trends and preserve shifts better than the median, but they are neither efficient for Gaussian noise nor robust. Combined FMH filters attenuate Gaussian noise more efficiently than predictive FMH filters in case of a constant signal and they track shifts better than the median, but worse than the predictive FMH.

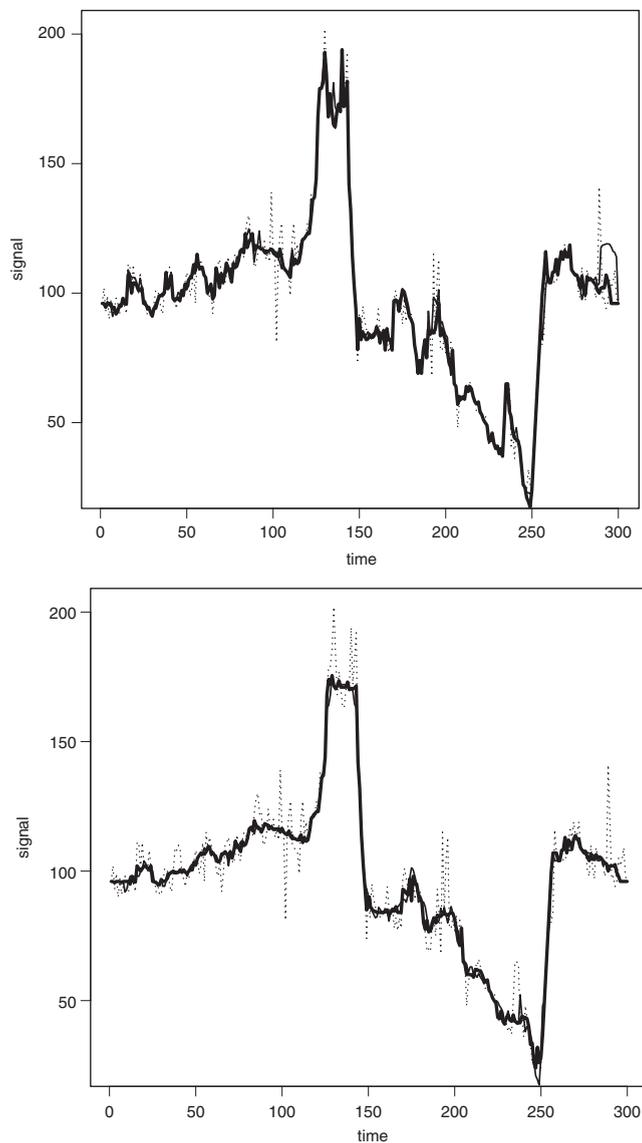


Fig. 11. Time series representing arterial pressure (dotted) and extracted signals. Top: PFMH (thin solid), PRMH (bold solid). Bottom: PRMMH (thin solid), CRMH (bold solid).

We prefer repeated median hybrid filters to the FMH filters since they provide the same benefits and are considerably more robust. Update algorithms allow fast computation of these filters. This becomes very important when analyzing high-frequency data using time windows longer than those considered here. The predictive RMH filter is not affected by trends and can preserve shifts exactly. Its major drawback might be its low efficiency, which

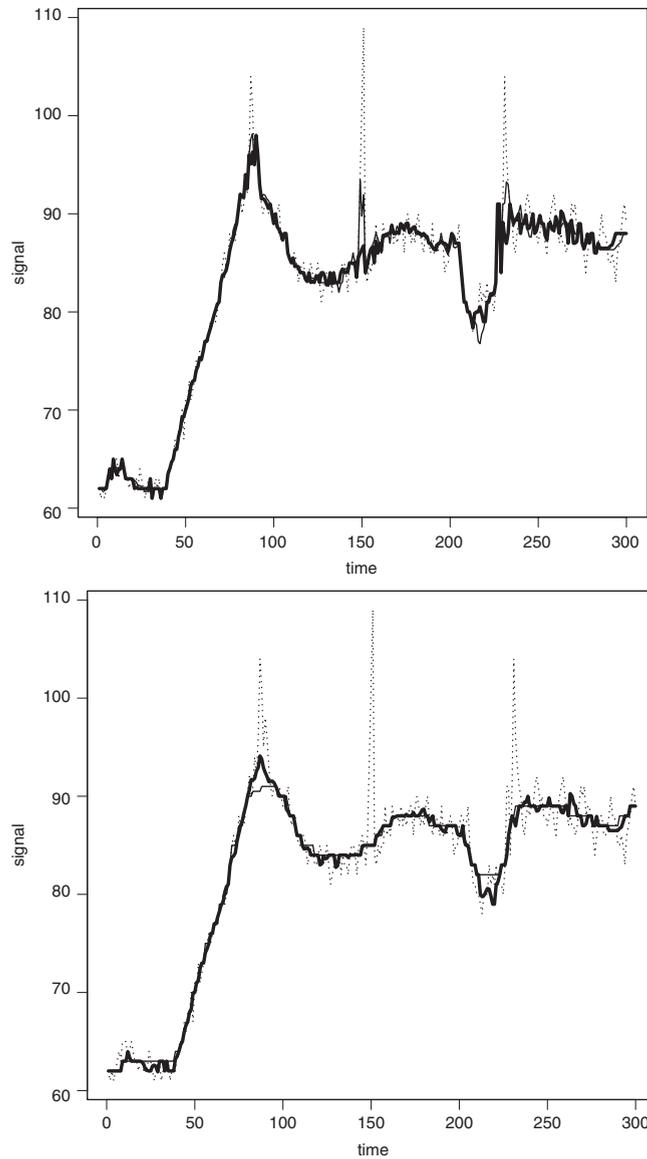


Fig. 12. Time series representing arterial pressure (dotted) and extracted signals. Top: PFMH (thin solid), PRMH (bold solid). Bottom: PRMMH (thin solid), CRMH (bold solid).

might be increased in a second step by applying a linear filter to its outcome. The combined RMMH shows a performance close to the median offering similar robustness properties and slightly better performance in trend periods, but being less efficient. The combined RMH and the predictive RMMH can be seen as compromises between these extremes. They offer

larger efficiency and higher robustness than the predictive RMH and preserve shifts better than the median. The predictive RMMH has the better analytical properties since more spikes are needed to influence its results and since it can preserve a shift irrespective of the directions of the shift and an underlying trend. Nevertheless, we have found the combined RMH to perform better in practice.

Acknowledgements

The financial support of the Deutsche Forschungsgemeinschaft (SFB 475, “Reduction of complexity in multivariate data structures”) is gratefully acknowledged. This work was initiated while the first author visited the University College London supported by the European Community’s Human Potential Programme under contract HPRN-CT-2000-00100 (DYNSTOCH). The authors thank Eleni Mitropoulou for additional data analysis and the referees for their suggestions, which were helpful to improve the presentation.

References

- Astola, J., Heinonen, P., Neuvo, Y., 1989. Linear median hybrid filters. *IEEE Trans. Circuits Systems* 36, 1430–1438.
- Bernholt, T., Fried, R., 2003. Computing the update of the repeated median regression line in linear time. *Inform. Process. Lett.* 88, 111–117.
- Cole, R., Salowe, J.S., Steiger, W.L., Szemerédi, E., 1989. An optimal-time algorithm for slope selection. *SIAM J. Comput.* 18 (4), 792–810.
- Cormen, T.H., Leiserson, C.E., Rivest, R.L., Stein, C., 2001. *Introduction to Algorithms*, second ed. MIT Press, Cambridge, MA, and McGraw-Hill Book Company, New York.
- Davies, P.L., 1993. Aspects of robust linear regression. *Ann. Statist.* 21, 1843–1899.
- Davies, P.L., Fried, R., Gather, U., 2004. Robust signal extraction for on-line monitoring data. *J. Statist. Plann. Inference* 122, 65–78.
- Heinonen, P., Neuvo, Y., 1987. FIR-median hybrid filters. *IEEE Trans. Acoust. Speech Signal Process.* 35, 832–838.
- Heinonen, P., Neuvo, Y., 1988. FIR-median hybrid filters with predictive FIR substructures. *IEEE Trans. Acoust. Speech Signal Process.* 36, 892–899.
- Kalli, S., Heinonen, P., Turjanmaa, V., Saranummi, N., 1985. Identifying patterns and profiles in 24-hour blood pressure recordings. In: *Proceedings of the Fifth International Symposium on Ambulatory Monitoring*, pp. 113–118.
- Matoušek, J., Mount, D.M., Netanyahu, N.S., 1998. Efficient randomized algorithms for the repeated median line estimator. *Algorithmica* 20, 136–150.
- Nieminen, A., Neuvo, Y., Mitra, U., 1989. Algorithms for real-time trend detection. *Signal Process.* 18, 1–15.
- Siegel, A.F., 1982. Robust regression using repeated medians. *Biometrika* 69 (1), 242–244.
- Stein, A., Werman, M., 1992. Finding the repeated median regression line. In: *SODA: ACM-SIAM Symposium on Discrete Algorithms*, pp. 409–413.
- Tukey, J.W., 1977. *Exploratory Data Analysis*. Addison-Wesley, Reading, MA.