Getting started with `robfilter`

Roland Fried, Karen Schettlinger and Matthias Borowski

1 Introduction

`robfilter` is a package of functions for robust extraction of an underlying signal from time series. The general idea is to approximate the data in a moving time window by a constant level (location-based methods) or a linear trend (regression-based methods). The several functions differ with respect to the signal characteristics and the outlier patterns they can deal with. Some of the procedures are available both for delayed filtering by approximating the signal in the window centre and for full online analysis by estimating the signal value at the end of the time window. The former versions generally lead to better approximations, but for the costs of a time delay corresponding to half a window width.

2 Overview

This package contains the following functions:

`robreg.filter`	simple regression filters (delayed and online)
`dw.filter`	location or regression based filters working in two steps with possibly different window widths (delayed and online)
`hybrid.filter`	median and repeated median hybrid filters (delayed)
`robust.filter`	regression filter with additional rules for outlier replacement and level shift tracing (delayed and online)
`adore.filter`	repeated median filter with adaptive choice of the window width (online)
`madore.filter`	a multivariate version of the `adore.filter` (online)
`scarm.filter`	repeated median filter with adaptive choice of the window width (online)
`mscarm.filter`	a multivariate version of the `scarm.filter` (online)
`wrm.filter`	weighted repeated median filters (delayed and online)
`wrm.smooth`	weighted repeated median smoother (delayed and online)

A more detailed explanation of the single functions can be found below.

2.1 robreg.filter

Running medians are the classical window-based robust filtering procedures. They simply calculate the median of the data in each window for the filter output. This leads to good results only when the signal is approximately constant within each time window, and it generally leads to delayed estimates. Running medians should only be applied using very short time windows in case of time trends. Better results can be achieved by replacing the median by a robust regression method.

The `robreg.filter` function applies simple robust linear regression for fitting a straight line to the data within each time window and estimates the current level of the time series either by the fitted value in the window centre (`online=FALSE` where `width` must be odd) or by the fitted value at the recent most time point within the time window (option `online=TRUE`).

Possible options for the robust regression (`method`) applied within one window are: the repeated median of Siegel (1982) (`"RM"`), least median of squares (`"LMS"`; Hampel, 1975; Rousseeuw, 1984), least trimmed squares (`"LTS"`; Rousseeuw, 1983), least quartile differences (`"LQD"`; Croux, Rousseeuw and Hössjer, 1994) and deepest regression (`"DR"`; Rousseeuw and Hubert, 1999). For comparison, the simple running median can also be calculated setting `method="MED"`.

Comparisons of the performance of certain of the above named methods can be found in Fried and Gather (2002) and Davies, Fried and Gather (2004) for the delayed filter versions, and in Gather, Schettlinger and Fried (2006) for the online versions. Fast algorithms for the running `RM` and `LQD` are described by Bernholt and Fried (2003) and Bernholt, Nunkesser and Schettlinger (2007), respectively.

2.2 dw.filter

Double window filters improve on simple regression-based filters at the occurrence of level shifts. They are based on the idea of trimming all observations deviating too much from a rough initial fit obtained from a possibly shorter inner time window (Bernholt, Fried, Gather, Wegener, 2006). Only the observations which have not been trimmed in this preliminary step are included in the calculation of the final filter output. Both location-based and regression-based procedures have been implemented, as well as full online and delayed versions.

2.3 hybrid.filter

Repeated median hybrid filters further improve the preservation of level shifts and local extremes of the signal value ('turning points') by calculating several subfilters from different parts of the window (Fried, Bernholt, Gather, 2006). The price of this is reduced smoothing and robustness. Different versions of this approach have been designed for locally constant or locally linear signals, all versions working with some delay.

2.4 robust.filter

The function `robust.filter` uses rules for outlier and shift detection to improve the results of an ordinary repeated median filter (Fried, 2004). For this, a robust scale estimator and suitable multiples of the estimated standard deviation need to be chosen. Additionally, the window width can be allowed to be time-varying so that the procedure adapts itself to time-varying slopes and local signal characteristics. The adaption of the window width is described by Gather and Fried (2004). For more explanations on shift detection, see Fried and Gather (2007).

2.5 `adore.filter`

The `adore.filter` overcomes the problem of the window width choice and the accompanying bias variance trade-off for the signal level estimations by choosing the window width adaptively, depending on the current data situation. Therefore, it provides an alternative to online filters applying rules for shift or change detection. The window width adaption is predicated on a test which is based on the signs of the residuals of a Repeated Median regression. The filter produces online signal estimates, and if wanted also accompanying robust online scale estimates. For more details see Schettlinger, Fried and Gather (2010).

2.6 `madore.filter`

The `madore.filter` is an adaptive online filter for multivariate time series, based on repeated median and multivariate least squares regression. The test procedure of the `adore.filter` (see above) is used to adapt the window width at each time point, depending on the current data situation. The signal is estimated within the adapted time window by a slight modification of the multivariate *Trimmed Repeated Median-Least Squares* regression (Lanius, Gather, 2010). In contrast to the `adore.filter`, the `madore.filter` takes account of the local correlations between the components of the multivariate time series. A more detailed description of the filter can be found in Borowski, Schettlinger, Gather (2009).

2.7 `scarm.filter`

The `scarm.filter` overcomes the problem of the window width choice and the accompanying bias variance trade-off for the signal level estimations by choosing the window width adaptively, depending on the current data situation. Therefore, it provides an alternative to online filters applying rules for shift or change detection. The window width adaption is predicated on a test which is based on the difference of Repeated Median slopes computed in separated time windows. The filter produces online signal estimates, and if wanted also accompanying robust online scale estimates. For more details see Borowski and Fried (2012).

2.8 `mscarm.filter`

The `mscarm.filter` is an adaptive online filter for multivariate time series, based on repeated median and multivariate least squares regression. The test procedure of the `scarm.filter` (see above) is used to adapt the window width at each time point, depending on the current data situation. The signal is estimated within the adapted time window by a slight modification of the multivariate *Trimmed Repeated Median-Least Squares* regression (Lanius, Gather, 2010). In contrast to the `scarm.filter`, the `mscarm.filter` takes account of the local correlations between the components of the multivariate time series. In contrast to the `madore.filter`, the `mscarm.filter` compares the estimated trends at each time point to identify parallel running time series components. These information are given to the user and used to improve the signal extraction. For more details see Borowski (2012).

2.9 `wrm.filter`

Weighted repeated median filters weight the observations in the time window according to their temporal distance to the current target point at which the signal is estimated (Fried, Einbeck, Gather, 2007). This weighting allows to increase the window width as compared to repeated median filters, and thus it increases the efficiency of noise reduction without increasing the bias at level shifts. They can be particularly recommended for robust detail-preserving signal extraction without time delay.

2.10 `wrm.smooth`

Weighted repeated median smoothing can be seen as a generalisation of weighted repeated median filtering allowing for non-equidistant time points. More generally, it can be applied for robust non-parametric regression of a dependent variable y on an independent (regressor) variable x and preserves discontinuities in the regression function. Accordingly, this function needs two variables, a dependent and an independent one.

A comparison of some of the above named methods can be found in Gather and Fried (2004) (including some regression filters and some double window and hybrid filters) and an overview over all methods is given in Schettlinger, Fried and Gather (2006).

3 Contact

Roland Fried
Department of Statistics, TU Dortmund University, 44221 Dortmund, Germany
E-Mail: `fried@statistik.tu-dortmund.de`

Karen Schettlinger
Department of Statistics, TU Dortmund University, 44221 Dortmund, Germany
E-Mail: `schettlinger@statistik.tu-dortmund.de`

Matthias Borowski
Department of Statistics, TU Dortmund University, 44221 Dortmund, Germany
E-Mail: `borowski@statistik.tu-dortmund.de`

4 References

Bernholt, T., Fried, R. (2003). Computing the Update of the Repeated Median Regression Line in Linear Time, *Information Processing Letters* **88**(3), 111-117.

Bernholt, T., Fried, R., Gather, U., Wegener, I. (2006). Modified Repeated Median Filters, *Statistics and Computing* **16**, 177-192.

Bernholt, T., Nunkesser, R., Schettlinger, K. (2007). Computing the Least Quartile Difference Estimator in the Plane, *Computational Statistics and Data Analysis* **52**(2), 763-772.

Borowski, M. (2012). Echtzeit-Extraktion relevanter Information aus multivariaten Zeitreihen basierend auf robuster Regression, *PhD thesis, TU Dortmund University (in German).*

Borowski, M., Fried, R. (2012). Online signal extraction by robust regression in moving windows with data-adaptive width selection. *Submitted to Statistics and Computing.*

Borowski, M., Schettlinger, K., Gather, U. (2009). Multivariate real time signal processing by a robust adaptive regression filter. *Communications in Statistics – Simulation and Computation*, **38**(2), 426-440.

Croux, C. and Rousseeuw, P.J., Hössjer, O. (1994). Generalized S-Estimators, *Journal of the American Statistical Association* **89**, 1271-1281.

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

Fried R. (2004). Robust Filtering of Time Series with Trends, *Journal of Nonparametric Statistics* **16**(3), 313-328.

Fried, R., Bernholt, T., Gather, U. (2006). Repeated Median and Hybrid Filters, *Computational Statistics and Data Analysis* **50**, Special Issue on 'Statistical Signal Extraction and Filtering', 2313-2338.

Fried, R., Einbeck, J., Gather, U. (2007). Weighted Repeated Median Smoothing and Filtering, *Journal of the American Statistical Association* **102**(480), 1300-1308.

Fried, R., Gather, U. (2002). Fast and Robust Filtering of Time Series with Trends, in: W. Härdle and B. Rönz (eds.), *Proceedings in Computational Statistics*, Physica, Heidelberg, 367-372.

Fried, R., Gather, U. (2007). On Rank Tests for Shift Detection in Time Series, *Computational Statistics and Data Analysis, Special Issue on Machine Learning and Robust Data Mining* **52**, 221-233.

Gather, U., Fried, R. (2004). Methods and Algorithms for Robust Filtering, in: J. Antoch (ed.), *Proceedings in Computational Statistics*, Physica, Heidelberg, 159-170.

Gather, U., Schettlinger, K., Fried, R. (2006). Online Signal Extraction by Robust Linear Regression, *Computational Statistics*, **21**(1), 33-51.

Hampel, F.R. (1975). Beyond Location Parameters: Robust Concepts and Methods, *Bulletin of the International Statistical Institute* **46**, 375-382.

Lanius, V., Gather, U. (2010), Robust online signal extraction from multivariate time series, *Computational Statistics and Data Analysis* **54**, 966-975.

Rousseeuw, P.J.(1983). Multivariate Estimation with High Breakdown Point, *Proceedings of the 4th Pannonian Symposium on Mathematical Statistics and Probability* **B**, Dordrecht, Reidel.

Rousseeuw, P.J. (1984). Least Median of Squares Regression, *Journal of the American Statistical Association* **79**, 871-880.

Rousseeuw, P.J., Hubert, M. (1999). Regression Depth, *Journal of the American Statistical Association* **94**, 388-402.

Schettlinger, K., Fried, R., Gather, U. (2006). Robust Filters for Intensive Care Monitoring - Beyond the Running Median, *Biomedizinische Technik* **51**(2), 49-56.

Schettlinger, K., Fried, R., Gather, U. (2010). Real time signal processing by adaptive repeated median filters. *International Journal of Adaptive Control and Signal Processing* **24**, 346-362.

Siegel, A.F. (1982). Robust Regression Using Repeated Medians, *Biometrika* **69**, 242-244.