Fast and Robust Filtering of Time Series with Trends

Roland Fried[1] and Ursula Gather[1]

[1] Department of Statistics, University of Dortmund, 44221 Dortmund, Germany

Abstract. Fast and robust methods are needed for denoising time series data measured with high sampling frequencies. In intensive care e.g. physiological variables like the heart rate are observed in short time intervals. Systematic changes have to be detected quickly and distinguished from clinically irrelevant short term fluctuations and artifacts. Median filtering works well if there is no substantial trend in the data but improvements are possible by approximating the data by a local linear trend.

Keywords. Signal extraction, linear regression, level shifts, outliers

1 Introduction

Tukey (1977) suggests median filtering as a simple approach to smooth noisy time series, clean the data from artifacts and extract the current level. Atypical observations (outliers) and sudden changes (level shifts) in the data can be identified in this way. However, median filtering does not consider trend periods although in view of high sampling frequencies most changes occur gradually. Fitting a local linear trend instead of a local level may improve the approximation and provide information on trends.

In a recent study, Davies, Fried and Gather (2002) compare the properties of L_1 regression, the repeated median and the least median of squares (LMS) for fitting a local linear trend to a time series. They find the LMS to resist a large percentage of outliers without showing a significant bias, while the repeated median performs somewhat better than L_1 regression. However, calculation of the LMS is time consuming as its computational complexity for fitting a line is $O(n^4)$ (Stromberg, 1993). Therefore, we cannot apply the LMS to time intervals. Instead we can use a faster method like the repeated median (complexity $O(n^2)$), or we can preprocess the data using e.g. the median (complexity $O(n)$) or the repeated median and apply the LMS to the more sparsely sampled outcome of this preprocessing.

In order to develop a procedure for non-parametric filtering of a time series with trends we compare the performance of methods for preprocessing. Thereafter we apply combinations of the LMS or the repeated median with such a preprocessing procedure to real and simulated time series and compare the results to the outcome of a "long" repeated median.

2 Preprocessing

Let $y_1, \ldots, y_N$, $N = n_1 n_2$, be a sequence of measurements, $n_1 = 2m + 1$. For preprocessing the data we first approximate the level μ_{jn_1+m+1} at the center of each time interval $jn_1 + 1, \ldots, (j + 1)n_1$, $j = 0, \ldots, n_2 - 1$. Thereafter we can apply a method for robust linear regression to the approximates

$\tilde{\mu}_{jn_1+m+1}, j = 0, \ldots, n_2 - 1$. For ease of notation we restrict the discussion to the first time interval centered about $m+1$ in the following.

2.1 Methods for preprocessing

A simple possibility for approximation of μ_{m+1} is to use the median $\tilde{\mu}_{Me}$ of $y_1, \ldots, y_{n_1}$. Alternatively, one can take the slope within the current time window into account and apply the repeated median T_{RM} (Siegel, 1982)

$$\tilde{\mu}_{RM} = med_i(y_i - \tilde{\beta}_{RM}(i-m-1)),$$
$$\tilde{\beta}_{RM} = med_i\left(med_{j\neq i}\frac{y_i - y_j}{i-j}\right)$$

or L_1 regression T_{L1} defined by

$$(\tilde{\mu}_{L1}, \tilde{\beta}_{L1}) = \text{argmin}\left\{(a,b) : \sum_{i=1}^{n_1} |y_i - a - b(i-m-1)|\right\}. \qquad (1)$$

Davies et al. (2002) find the small-sample performance of T_{RM} and T_{L1} to be rather similar, although T_{RM} has the larger breakdown point, that is $\lfloor n/2 \rfloor / n \approx 0.5$. The breakdown point of T_{L1} for design points on a lattice as in the situation considered here is approximately $1 - 1/\sqrt{2} \approx 0.293$.

Fast computation of T_{L1} is possible using a descent technique (Sposito, 1990). An one-step or a two-step approximation starting from the median for the level and zero slope might be interesting because of the robustness of the starting point and its computational speed. In the iterations, we alternate between calculation of the generalized median of the scaled deviations from the current level with weights $|j-m+1|$, and calculation of the median of the trend-corrected observations. However, we find the breakdown point of the k-step estimator, $k \geq 2$, to be the same as the breakdown point of T_{L1}. Therefore we do not consider such k-step approximations in the following.

2.2 Efficiency

For a comparison of the small-sample properties of the distinct methods we simulate data from the model

$$Y_t = \mu + \beta(t-m-1) + \epsilon_t, \quad t = 1, \ldots, 2m+1,$$

with $\mu = 0$ and several slopes β. The errors ϵ_t are always Gaussian white noise with mean zero and unit variance. The value of m is determined by requiring on the one hand a certain stability (m large) and on the other hand the demands made by the on-line nature of the application (m small). We consider small to moderate $m = 5, 10, 15, 30$ corresponding to $n_1 = 11, 21, 31, 61$.

As a first step we derive the relative efficiencies of $\tilde{\mu}_{Me}$, $\tilde{\mu}_{L1}$ and $\tilde{\mu}_{RM}$ with respect to the least squares estimator $\tilde{\mu}_{L2}$. Since for high-frequency measurements the standard deviation of the measurement error will typically be large in comparison to realistic slope sizes, we consider $\beta = 0, 0.01, 0.05, 0.1$ (Table 1). For the level, $\tilde{\mu}_{L1}$ is always more efficient than $\tilde{\mu}_{Me}$, which is slightly more efficient than $\tilde{\mu}_{RM}$ if $\beta = 0$. The relative efficiency of $\tilde{\mu}_{Me}$ decreases with increasing $|\beta|$ and increasing sample size. For $\beta = 0.1$ fixed, increasing the sample size further than $n_1 = 31$ does almost no longer decrease the MSE of $\tilde{\mu}_{Me}$. In general, the median is not even consistent if we extend the sample range, but it is consistent if we consider in-fill asymptotics corresponding to increasing the sampling frequency.

Table 1. Efficiencies relative to L_2 regression (in percent) measured by the simulated MSE for $\tilde{\mu}_{Me}$, $\tilde{\mu}_{L1}$ and $\tilde{\mu}_{RM}$ for $N(0,1)$ errors.

m	β	$\tilde{\mu}_{Me}$	$\tilde{\mu}_{L1}$	$\tilde{\mu}_{RM}$	m	β	$\tilde{\mu}_{Me}$	$\tilde{\mu}_{L1}$	$\tilde{\mu}_{RM}$
5	0.00	66.0	69.2	65.0	15	0.00	64.8	66.7	64.3
5	0.01	66.5	69.4	65.3	15	0.01	64.2	66.3	63.9
5	0.05	65.5	69.6	65.6	15	0.05	60.1	66.6	63.7
5	0.10	63.8	69.2	65.2	15	0.10	49.1	67.0	63.7
10	0.00	64.6	66.8	63.9	30	0.00	64.5	65.5	64.1
10	0.01	64.4	66.7	63.9	30	0.01	63.3	65.6	63.8
10	0.05	63.1	67.7	64.1	30	0.05	49.0	65.4	63.4
10	0.10	57.3	68.2	64.2	30	0.10	29.4	65.9	64.2

2.3 Computation times

The computation time needed is important if many variables are to be monitored simultaneously or other algorithms are run concurrently. Table 2 shows the mean times of applying the functionals to 1000 samples of sizes 11 to 31 using a self-written FORTRAN program on a Sun workstation ultra spark with 170 MHz and 320 MB Ram. While $\tilde{\mu}_{Me}$ needs considerably more time than the arithmetic mean $\tilde{\mu}_{AM}$, it is much cheaper than $\tilde{\mu}_{L1}$ and $\tilde{\mu}_{RM}$.

The computation time needed for the whole procedure can be calculated from these values as follows: If we apply method A in the first step to n_1 observations and then apply method B to n_2 subsequent outcomes of the first step the computation time will be approximately (with obvious notation) $n_2 C^A_{n_1} + C^B_{n_2}$ for the first $n_1 n_2$ observations but only $C^A_{n_1} + C^B_{n_2}$ thereafter.

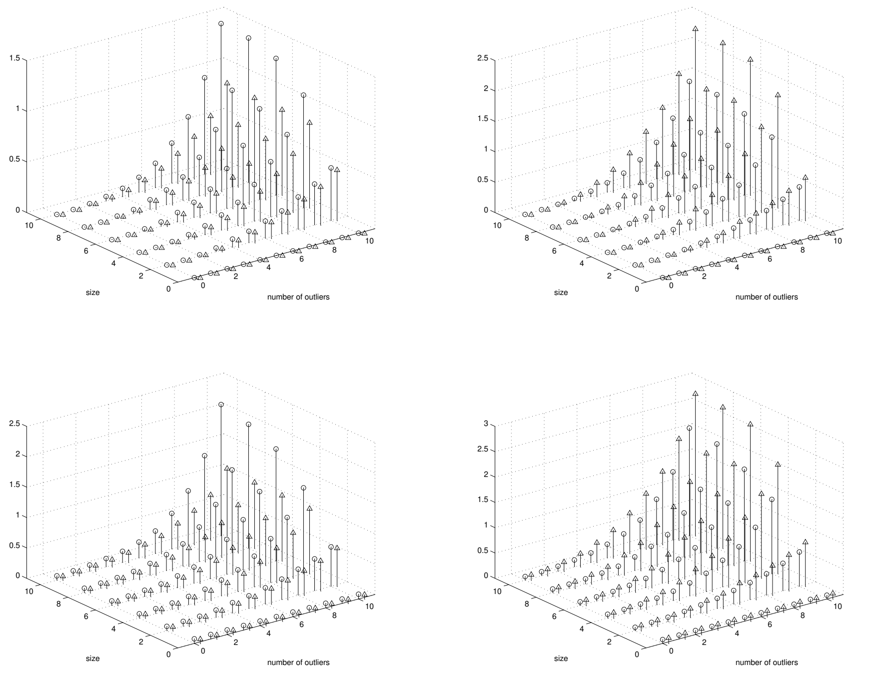
Table 2. Mean time C^A_n of applying method A to 1000 samples of size n (in seconds).

	C^{AM}_n	C^{Me}_n	C^{L1}_n	C^{RM}_n
$n = 11$	0.0075	0.042	0.55	0.89
$n = 21$	0.0096	0.047	2.41	2.63
$n = 31$	0.0120	0.177	4.72	4.38

2.4 Outliers

In applications like intensive care many outliers occur. For the sake of brevity we restrict attention to a sample of size $n_1 = 21$ and replace an increasing number of observations 0(1)10 by additive outliers of increasing size 0(2)10 at random time points in the window. We concentrate on one-sided positive outliers as these constitute a more difficult challenge than outliers with random sign. Each of the 121 cases is simulated 2000 times and the squared bias, variance and mean square error are calculated. In view of the results obtained above and in Davies et al. (2002) we only consider $\tilde{\mu}_{Me}$ and $\tilde{\mu}_{RM}$ for preprocessing further on, cf. Figure 1. Both methods are able to resist up to six outliers. For more outliers, the comparison depends on the slope: $\tilde{\mu}_{Me}$ resists outliers better in case of a small slope ($\beta = 0$), while $\tilde{\mu}_{RM}$ is better in case of a rather large slope ($\beta = 0.2$), the differences being due to bias.

Fig. 1. Steady state ($\beta = 0$, left) and trend ($\beta = 0.2$, right): Simulated squared bias (top) and MSE (bottom) for the level. Repeated median $\circ$ and median $\triangle$.



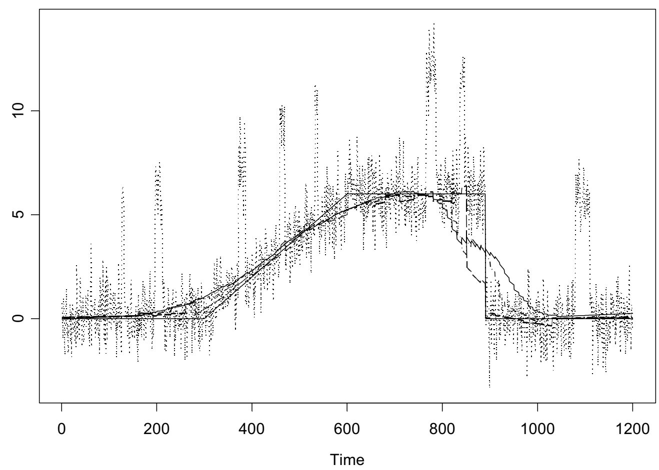
3 Combination of the procedures

Figure 2 shows a simulated time series with 5% of the observations replaced by isolated and 10% by patchy outliers (i.e. 15% outliers altogether) as well as some approximates of the level and the slope. The choice of the outliers is guided by our experience with real physiologic time series measured every second. We used $n_1 = 11$ observations for preprocessing and $n_2 = 31$ in the second step. This former is motivated by the previous findings that we should choose n_1 small when using the median, while the latter is due to the findings of Davies et al. (2002) who consider 31 observations to be sufficient to get a reliable slope approximate. The two-step approximate combining the repeated median and the median performs at least as good as the repeated median applied to the full time interval with $n_1 n_2$ observations and it is computationally much less expensive. This is also true w.r.t. the slope approximates not shown here. T_{RM} as well as the combination methods provide good approximates in the steady state as well as in the trend period.

Sudden changes in the slope or the level can better be detected using T_{LMS} instead of T_{RM}. However, T_{LMS} shows instabilities w.r.t. certain monotone (random) patterns in the data.

Similar results are obtained when applying the methods to the heart rate of a patient observed every second, cf. Figure 3. T_{LMS} bridges the decrease at the beginning but it still signals a trend when the level has already stabilized. T_{RM} and T_{RM} in combination with $\tilde{\mu}_{RM}$ perform almost identical, while T_{RM} in combination with $\tilde{\mu}_{AM}$ is strongly influenced by some negative outliers.

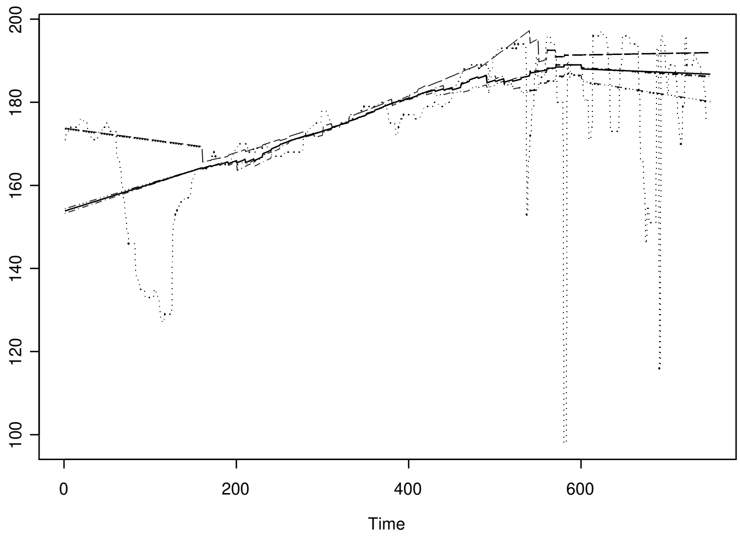
Fig. 2. Simulated time series (dotted), underlying level (bold solid) and level estimates: T_{RM} with $\tilde{\mu}_{Me}$ (dashed-dotted), T_{LMS} with $\tilde{\mu}_{Me}$ (dashed), T_{RM} (solid).



4 Conclusion

The suitable choice of a filtering procedure for fitting of local linear trends strongly depends on the application. The suitable length of the time windows should be guided by the expected length of outlier patches and the minimal length of relevant structural changes. If many variables are to be monitored simultaneously, application of T_{LMS} or T_{RM} to a very long time window may not be possible. E.g., in our simulation example calculation of the T_{RM} from the long time window takes more than 200 times longer than using a combination of two repeated medians. In theory, using such a combination reduces the breakdown point as it is asymptotically 1/4 instead of 1/2 for each of the approximates. However, in practice more outliers can be resisted as outliers will often occur in long patches if the sampling frequency is high. If the

Fig. 3. Heart rate (dotted) and level estimates: T_{RM} with $\tilde{\mu}_{Me}$ (dashed-dotted), T_{RM} with $\tilde{\mu}_{AM}$ (dashed-dotted-dotted), T_{LMS} with $\tilde{\mu}_{Me}$ (dashed), T_{RM} (solid). T_{RM} with $\tilde{\mu}_{Me}$ and T_{RM} are almost identical.



sampling frequency is high in comparison to realistic slope sizes, application of $\tilde{\mu}_{Me}$ with a small window width seems sufficient for preprocessing as it is rather efficient, fast and can withstand several outliers. Since its relative performance in comparison to $\tilde{\mu}_{RM}$ worsens with increasing window width and increasing slope sizes, $\tilde{\mu}_{RM}$ may sometimes be a better choice.

References

Davies, P.L., Fried, R. and Gather, U. (2002). Robust signal extraction for on-line monitoring data. Technical Report 1/2002, SFB 475, University of Dortmund, Germany.

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

Sposito, V.A. (1990). Some properties of L_p-Estimators. In: K.D. Lawrence and J.L. Arthur (eds.). *Robust Regression. Analysis and Applications,* 23–58. New York: Marcel Dekker.

Stromberg, A.J. (1993). Computing the exact least median of squares estimate and stability diagnostics in multiple linear regression. *SIAM J. Sci. Comput.*, **14**, 1289–1299.

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