REVSTAT – Statistical Journal
Volume 5, Number 1, March 2007, 41–61

ESTIMATING SPECTRAL DENSITY FUNCTIONS ROBUSTLY

Authors: Bernhard Spangl
– Group of Applied Statistics and Computing,
University of Natural Resources and Applied Life Sciences,
Vienna, Austria
bernhard.spangl@boku.ac.at

Rudolf Dutter
– Department of Statistics and Probability Theory,
Vienna University of Technology,
Austria
R.Dutter@tuwien.ac.at

Abstract:

- We consider in the following the problem of robust spectral density estimation.

  Unfortunately, conventional spectral density estimators are not robust in the presence of additive outliers (cf. [18]). In order to get a robust estimate of the spectral density function, it turned out that cleaning the time series in a robust way first and calculating the spectral density function afterwards leads to encouraging results. To meet these needs of cleaning the data we use a robust version of the Kalman filter which was proposed by Ruckdeschel ([26]). Similar ideas were proposed by Martin and Thomson ([18]).

  Both methods were implemented in R (cf. [23]) and compared by extensive simulation experiments. The competitive method is also applied to real data. As a special practical application we focus on actual heart rate variability measurements of diabetes patients.

Key-Words:

- *robustness; spectral density function; AO-model.*

AMS Subject Classification:

- 62F35, 62M15, 60G35.

1. INTRODUCTION

Our research has been motivated by the frequency-domain analysis of short-term heart rate variability (HRV) measurements. This is a non-invasive method which has been increasingly used in medicine (cf. [8, 22]). To analyze biosignals or, generally speaking, time series, the spectral density function is commonly used in many application areas. Further areas of applications besides medicine are signal processing (cf. [31]) and geophysics (cf. [4, 9]).

The additive outlier model (AO model) which was introduced by Fox ([6]) is a commonly used model for outliers in time series. The AO model consists of a stationary core process, x_t, to which occasional outliers are added. The observed process $\{y_t,\, t=1,...,n\}$ is said to have additive outliers if it is defined by

$$y_t \;=\; x_t + v_t \tag{1.1}$$

where the contaminations v_t are independent and identically distributed. For the methods presented in this paper, it is convenient to model the univariate distribution of v_t by a contaminated normal distribution with degenerated central component, i.e.,

$$\mathcal{CN}(\gamma, 0, \sigma^2) \;=\; (1-\gamma)\,\mathcal{N}(0,0) + \gamma\,\mathcal{N}(0,\sigma^2)\;. \tag{1.2}$$

Hence, the core process x_t is observed with probability $1-\gamma$ whereas the core process plus a disturbance v_t is observed with probability γ. We shall also assume that x_t and v_t are independent.

The AO model seems to be an appropriate model when analyzing heart rate variability data. To access the variability of heart rate in the frequency domain the spectral density function of the tachogram is estimated. The tachogram is the series of time intervals between consecutive heart beats, the so called R-R-intervals (e.g. Figure 1). The R-R-interval denotes the period between an R-peak and the next R-peak in an electrocardiogram.

Non-sinus ectopic beats and other artifacts can cause outlying observations in the tachogram. If, during the recording and sampling, an R-peak is missed in the electrocardiogram (ECG) this will result in a very large value in the tachogram. Or, if an ectopic beat occurs, i.e., if there is an extra heart beat between two regular beats, the amplitude in the ECG of the heart beat following the ectopic beat will be very low and therefore this beat will usually be missed. This results in a lower tachogram value followed by a higher one.

The aim of accessing the heart rate variability is accomplished by estimating the spectral density function of the tachogram robustly in order to be insensitive against outlying tachogram values caused by ectopic beats and other artifacts.

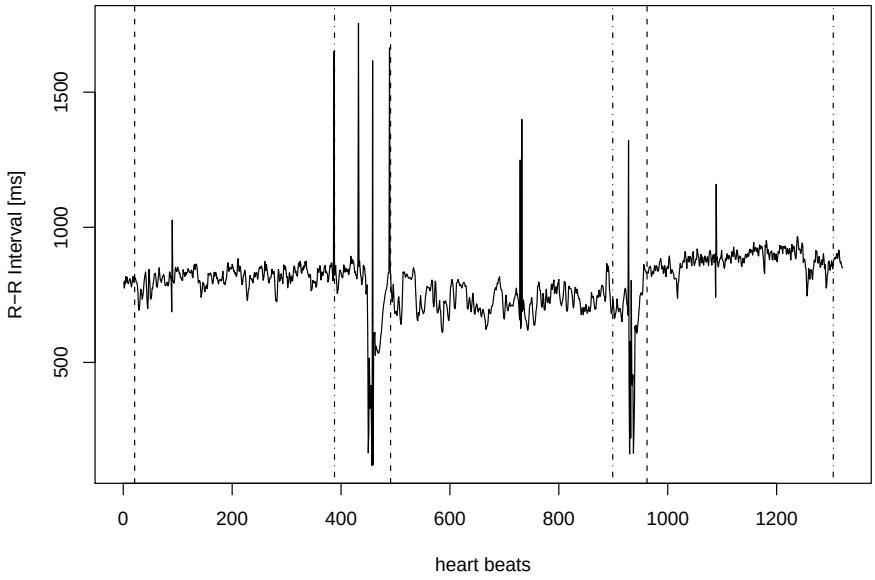

Figure 1: Tachogram of 1321 consecutive heart beats.

We do not compute the spectral density function of the entire tachogram series, but estimate several within overlapping windows to assure stationarity in each window (cf. also [28]). Each slice in Figure 2 represents the spectral density estimate of the corresponding time interval.

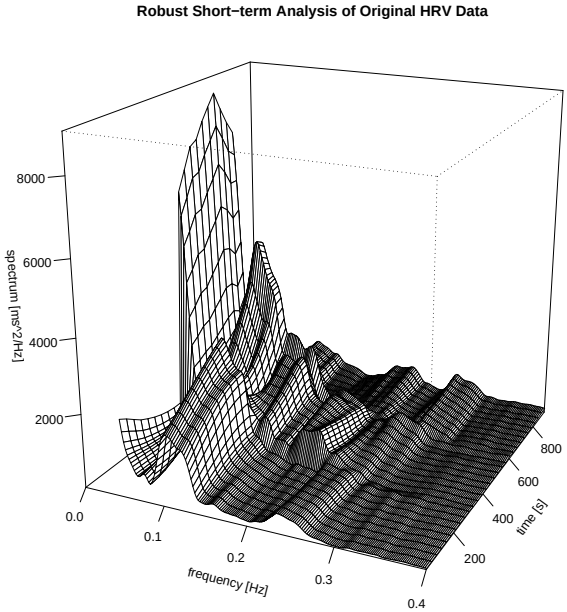

Figure 2: Robust dynamic Fourier analysis of the original short-term HRV data.

Although we do not use the entire tachogram series but several overlapping windows to access the heart rate variability we only focus on an analysis in the frequency domain. We are not interested in modeling the heart rate in the time domain nor in forecasting as this is often the aim in the context of online-monitoring.

In the present paper we consider the problem of estimating the spectral density function robustly. Unfortunately, conventional spectral density estimators are not robust in the presence of additive outliers. See [12] or [18] for details. To obtain a robust estimate of the spectral density function we present two different multi-step procedures. The first procedure was proposed by Martin and Thomson ([18]) and incorporates an important robust filtering operation which is accomplished by an approximate conditional-mean (ACM) type filter. For the second multi-step procedure we suggest to replace the ACM type filter and use the rLS filter proposed by Ruckdeschel ([26]) instead. Both filters are robustified versions of the Kalman filter. In order to compare both approaches we implement them in R.

In the next section we state the definitions of the state-space model and the classical Kalman filter which is the basis of the robustifying approaches proposed by Martin and Thomson ([18]) and Ruckdeschel ([26]). Both methods are described in Section 3. In Section 4 we give an outline of our simulation study and the results are presented in Section 5. Some remarks are given in Section 6.

2. PRELIMINARIES

2.1. State-space models

Let us assume we observe a q-dimensional, vector-valued process $\boldsymbol{y}_t$, $t=1,...,n$, which is only a linear transformation of an unobserved p-dimensional signal $\boldsymbol{x}_t$ with some noise added. Then the *state-space model* can be defined as follows:

$$\begin{aligned} \boldsymbol{x}_t &= \boldsymbol{\Phi}\,\boldsymbol{x}_{t-1} + \boldsymbol{\varepsilon}_t\ , \\ \boldsymbol{y}_t &= \boldsymbol{H}\boldsymbol{x}_t + \boldsymbol{v}_t\ , \end{aligned} \tag{2.1}$$

where $\boldsymbol{x}_t$ is the unobserved p-dimensional vector called the *state vector*. The first equation in (2.1) is called *state equation* and the second is called the *observation equation*. It is assumed that $\boldsymbol{\varepsilon}_t$ has dimension p, $\boldsymbol{\Phi}$ is a $p\times p$ matrix and $\boldsymbol{H}$ is a $q\times p$ matrix. We further assume that $\boldsymbol{x}_t$ is independent of future $\boldsymbol{\varepsilon}_t$, and that $\boldsymbol{\varepsilon}_t$ and $\boldsymbol{v}_t$ are individually zero mean independent and identically distributed (iid) sequences which also are mutually independent but could be non-Gaussian.

A more general definition of state-space models considering correlated errors as well as more complex models including exogenous variables or selection matrices can be found in [27] and [5].

2.2. The classical Kalman filter

The primary aim of any analysis using state-space models as defined by (2.1) is to produce estimators of the underlying unobserved signal $\boldsymbol{x}_t$, given the data $\boldsymbol{Y}_s = \{\boldsymbol{y}_1, \boldsymbol{y}_2, ..., \boldsymbol{y}_s\}$, up to time s. If $s < t$, $s = t$ or $s > t$, the problem is called *prediction*, *filtering* or *smoothing*, respectively.

In addition, we want to get estimators $T_t(\boldsymbol{Y}_s)$ of $\boldsymbol{x}_t$ which are best in the sense of the minimum mean-squared error, i.e.,

$$E\big(\|\boldsymbol{x}_t - T_t(\boldsymbol{Y}_s)\|^2\big) = \min_{T_t}! \; . \tag{2.2}$$

The solution is the conditional mean of $\boldsymbol{x}_t$ given $\boldsymbol{Y}_s$, i.e.,

$$T_t(\boldsymbol{Y}_s) = E(\boldsymbol{x}_t \,|\, \boldsymbol{Y}_s) \; , \tag{2.3}$$

and will further on be denoted by $\boldsymbol{x}_{t|s}$.

However, in general the conditional mean is hard to calculate and therefore we restrict ourselves to the class of linear estimators. Then the solution to these problems is accomplished via the *Kalman filter* and *smoother* (cf. [10, 11]). The estimators we obtain are the minimum mean-squared error estimates within the class of linear estimators.

In the following we will just focus on the Kalman filter. Its advantage is that it specifies how to update the filter values from $\boldsymbol{x}_{t-1|t-1}$ to $\boldsymbol{x}_{t|t}$ once a new observation $\boldsymbol{y}_t$ is obtained, without having to reprocess the entire data set $\boldsymbol{y}_1, \boldsymbol{y}_2, ..., \boldsymbol{y}_t$. The Kalman filter recursions can be split into three steps:

(**i**) Initialization $(t = 0)$:

$$\boldsymbol{x}_{0|0} = \boldsymbol{\mu}_0 \; , \qquad \boldsymbol{P}_0 = \boldsymbol{\Sigma}_0 \; , \tag{2.4}$$

where $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$ are the unconditional mean and $p \times p$ covariance matrix of $\boldsymbol{x}_0$;

(**ii**) Prediction $(t \geq 1)$:

$$\begin{aligned} \boldsymbol{x}_{t|t-1} &= \boldsymbol{\Phi}\, \boldsymbol{x}_{t-1|t-1} \; , \\ \boldsymbol{M}_t &= \boldsymbol{\Phi}\, \boldsymbol{P}_{t-1} \boldsymbol{\Phi}^\top + \boldsymbol{Q} \; ; \end{aligned} \tag{2.5}$$

(**iii**) Correction $(t \geq 1)$:

$$\begin{aligned} \boldsymbol{x}_{t|t} &= \boldsymbol{x}_{t|t-1} + \boldsymbol{K}_t(\boldsymbol{y}_t - \boldsymbol{H}\boldsymbol{x}_{t|t-1}) \ , \\ \boldsymbol{P}_t &= \boldsymbol{M}_t - \boldsymbol{K}_t \boldsymbol{H} \boldsymbol{M}_t \ , \\ \text{with} \qquad \boldsymbol{K}_t &= \boldsymbol{M}_t \boldsymbol{H}^\top (\boldsymbol{H}\boldsymbol{M}_t\boldsymbol{H}^\top + \boldsymbol{R})^{-1} \ . \end{aligned} \tag{2.6}$$

The $p \times q$ matrix $\boldsymbol{K}_t$ is called the *Kalman gain*. The $p \times p$ matrix $\boldsymbol{M}_t$ is the conditional prediction error covariance matrix,

$$\boldsymbol{M}_t = E\Big((\boldsymbol{x}_t - \boldsymbol{x}_{t|t-1}) (\boldsymbol{x}_t - \boldsymbol{x}_{t|t-1})^\top \mid \boldsymbol{Y}_{t-1} \Big) \ , \tag{2.7}$$

and the conditional filtering error covariance matrix $\boldsymbol{P}_t$ is given by

$$\boldsymbol{P}_t = E\Big((\boldsymbol{x}_t - \boldsymbol{x}_{t|t}) (\boldsymbol{x}_t - \boldsymbol{x}_{t|t})^\top \mid \boldsymbol{Y}_t \Big) \ . \tag{2.8}$$

Moreover, the $p \times p$ matrix $\boldsymbol{Q}$ and the $q \times q$ matrix $\boldsymbol{R}$ denote the covariance matrices of $\boldsymbol{\varepsilon}_t$ and $\boldsymbol{v}_t$, respectively.

3. ROBUST SPECTRAL DENSITY ESTIMATION

In order to obtain a robust estimate of the spectral density function, we clean the data in a robust way first and compute the spectral density function afterwards using a prewhitened spectral density estimate. This approach was proposed by Martin and Thomson ([18]) and leads to encouraging results. The data-cleaning operation wherein the robustness is introduced is accomplished by a robustified version of the Kalman filter.

Martin and Thomson ([18]), based on the work of Martin ([15]), propose to modify the calculation of the filter estimate as well as of the conditional filtering error covariance matrix in the correction step (2.6). In [15] Martin, motivated by Masreliez's result ([20]), only considers autoregressive models. This limitation to univariate signals and several approximations lead to a simplification of the correction step that enables a robust estimation of the filter estimate as well as of the conditional filtering error covariance matrix.

Another approach, proposed by Ruckdeschel ([26]), preserves the general concept of the Kalman filter, that allows for multivariate signals, and modifies only the updating of the filter estimate.

3.1. Robust prewhitening

Let $\{y_t,\, t = 1, ..., n\}$ again denote the observed process which is assumed to be second-order stationary and to have mean zero. The cleaning operator C maps the original data y_t into the cleaned data Cy_t. In the context of the AO model (1.1), we want the Cy_t to reconstruct the core process x_t, and so we will use the labeling $Cy_t = \widehat{x}_{t|t}$, where $\widehat{x}_{t|t}$ denotes an estimate of x_t at time t. The second index of $\widehat{x}_{t|t}$ should indicate that the kind of data cleaning procedure we have in mind here is a robust filtering procedure which uses the past and present data values $y_1, ..., y_t$ to produce a cleaned filter estimate $\widehat{x}_{t|t}$ of x_t, $t=1, ..., n$. For AO models with a fraction of contamination γ not too large, it turns out that the data cleaner has the property that $Cy_t = y_t$ most of the time, that is about $(1-\gamma)\times 100$ percent of the time.

The filter-cleaner procedure involves a robust estimation of an autoregressive approximation to the core process x_t of order p, with estimated coefficients $\widehat{\phi}_1, ..., \widehat{\phi}_p$. Now, the residual process

$$r_t \;=\; Cy_t - \sum_{i=1}^{p} \widehat{\phi}_i \, Cy_{t-i} \;, \qquad t = p+1, ..., n \;, \tag{3.1}$$

can easily be formed. Since cleaned data are used to obtain these residuals, and the $\widehat{\phi}_i$ are robust estimates, the transformation (3.1) is called a robust prewhitening operation. The benefit in the use of prewhitening in the context of spectral density estimation is to reduce the bias, i.e., the transfer of power from one frequency region of the spectral density function to another, known as leakage (cf. [3]).

The robust spectral density estimate is based on the above robust prewhitening as follows. Let

$$\widehat{H}_p(f) \;=\; 1 - \sum_{j=1}^{p} \widehat{\phi}_j \, e^{-i\,2\pi j f} \tag{3.2}$$

be the transfer function of the prewhitening operator (3.1) at frequency f, and let $\widehat{S}_r^{(lw)}(f)$ denote a lag window spectral estimate based on the residual process r_t. Then the spectral density estimate is

$$\widehat{S}(f) \;=\; \frac{\widehat{S}_r^{(lw)}(f)}{\left|\widehat{H}_p(f)\right|^2} \;, \tag{3.3}$$

where $\widehat{S}(f)$ is evaluated at the Fourier frequencies $f_k = k/n,\; k = 0, ..., [n/2]$.

3.2. The robust filter-cleaner algorithm

The robust filter-cleaner proposed by Martin and Thomson ([18]) is an approximate conditional-mean (ACM) type filter motivated by Masreliez's result ([20]).

3.2.1. The robust filter-cleaner

The filter-cleaner algorithm as presented in the paper of Martin and Thomson ([18]) relies on the p-th order autoregressive approximation of the underlying process x_t, which can be represented in state-space form (2.1) as follows. Assuming that x_t satisfies

$$x_t \;=\; \phi_1\, x_{t-1} + \phi_2\, x_{t-2} + \cdots + \phi_p\, x_{t-p} + \varepsilon_t$$

the state space model can be written as

$$\begin{aligned} \boldsymbol{x}_t \;&=\; \boldsymbol{\Phi}\, \boldsymbol{x}_{t-1} + \boldsymbol{\varepsilon}_t \;, \\ y_t \;&=\; x_t + v_t \;, \end{aligned} \tag{3.4}$$

with

$$\boldsymbol{x}_t \;=\; (x_t, x_{t-1}, ..., x_{t-p+1})^\top \;, \tag{3.5}$$

$$\boldsymbol{\varepsilon}_t \;=\; (\varepsilon_t, 0, ..., 0)^\top \tag{3.6}$$

and

$$\boldsymbol{\Phi} \;=\; \begin{pmatrix} \phi_1 & \cdots & \phi_{p-1} & \phi_p \\ 1 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 \end{pmatrix} \;. \tag{3.7}$$

Additionally, we set

$$\operatorname{cov}(\boldsymbol{\varepsilon}_t) \;=\; \boldsymbol{Q} \;=\; \begin{pmatrix} \sigma_\varepsilon^2 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix} \tag{3.8}$$

and

$$\operatorname{var}(v_t) \;=\; \boldsymbol{R} \;=\; \sigma_0^2 \;. \tag{3.9}$$

The algorithm computes robust estimates $\widehat{\boldsymbol{x}}_{t|t}$ of the unobservable $\boldsymbol{x}_t$ according to the following recursion:

$$\widehat{\boldsymbol{x}}_{t|t} \;=\; \boldsymbol{\Phi}\, \widehat{\boldsymbol{x}}_{t-1|t-1} + \frac{\boldsymbol{m}_{.1,t}}{s_t^2}\; s_t\; \psi\!\left(\frac{y_t - \widehat{y}_{t|t-1}}{s_t} \right) \tag{3.10}$$

with $\boldsymbol{m}_{.1,t}$ being the first column of $\boldsymbol{M}_t$ which is computed recursively as

$$\boldsymbol{M}_{t+1} = \boldsymbol{\Phi}\,\boldsymbol{P}_t\,\boldsymbol{\Phi}^\top + \boldsymbol{Q}\ , \tag{3.11}$$

$$\boldsymbol{P}_t = \boldsymbol{M}_t - w\left(\frac{y_t - \widehat{y}_{t|t-1}}{s_t}\right)\frac{\boldsymbol{m}_{.1,t}\,\boldsymbol{m}_{.1,t}^\top}{s_t^2}\ . \tag{3.12}$$

The weight function w is defined by

$$w(r) = \frac{\psi(r)}{r}\ , \tag{3.13}$$

where ψ stands for some psi-function described below. The scale s_t is set to

$$s_t^2 = m_{11,t} \tag{3.14}$$

and $\widehat{y}_{t|t-1}$ denotes a robust one-step-ahead prediction of y_t based on $\boldsymbol{Y}_{t-1} = \{y_1, ..., y_{t-1}\}$, and is given by

$$\widehat{y}_{t|t-1} = (\boldsymbol{\Phi}\,\widehat{\boldsymbol{x}}_{t-1|t-1})_1\ . \tag{3.15}$$

Finally, the cleaned process at time t results in

$$\widehat{x}_{t|t} = (\widehat{\boldsymbol{x}}_{t|t})_1\ . \tag{3.16}$$

It should be noted that if ψ is the identity function, which gives $w \equiv 1$, and (3.14) is replaced by $s_t^2 = m_{11,t} + \sigma_0^2$ with $\sigma_0^2 = \mathrm{var}(v_t)$ in the AO model, the above recursions are those of the Kalman filter. The use of $\sigma_0^2 = 0$ in (3.14) corresponds to the assumptions that $v_t = 0$ a large fraction of time and that a contaminated normal distribution with degenerated central component (1.2) provides a reasonable model. Correspondingly, $\boldsymbol{M}_t$ and $\boldsymbol{P}_t$ are the prediction and filtering error-covariance matrices as described in the previous section (Section 2). Again, in order to agree with the definition of the classical Kalman filter recursions, we specify the initial conditions for the above recursions by setting $\widehat{\boldsymbol{x}}_{0|0} = 0$ and $\boldsymbol{P}_0 = \widehat{\boldsymbol{C}}_{\boldsymbol{x}}$ where $\widehat{\boldsymbol{C}}_{\boldsymbol{x}}$ is an estimate of the $p \times p$ covariance matrix of the state process. We note that there also exists another way to specify those initial conditions (see [17]).

The psi-function ψ and the weight function w which are essential to obtain robustness should be bounded and continuous. Additionally, it is highly desirable that both have zero values outside a bounded, symmetric interval around the origin. Furthermore, $\psi(s)$ is odd and should look like the identity function for small values of s (see [15]). Boundedness assures that no single observation has an arbitrarily large effect on the filter-cleaner. Continuity assures that small variations, e.g., due to rounding, will not have a major effect. Compact support results in the following behavior which is desirable for a filter-cleaner: if an observation y_t deviates from its prediction $\widehat{y}_{t|t-1}$ by a sufficiently large amount, then $\widehat{\boldsymbol{x}}_{t|t}$ will be the pure prediction $\widehat{\boldsymbol{x}}_{t|t} = \boldsymbol{\Phi}\,\widehat{\boldsymbol{x}}_{t-1|t-1}$ and the filtering error covariance

$\boldsymbol{P}_t$ is set equal to the prediction error covariance $\boldsymbol{M}_t$. Martin and Thomson ([18]) proposed to use for ψ a special form of *Hampel's three-part redescending psi-function* ([7]),

$$\psi_{HA}(s) \;=\; \begin{cases} s & \text{if} \quad |s| \le a\,, \\ a\,\mathrm{sgn}(s) & \text{if} \quad a < |s| \le b\,, \\ \dfrac{a}{b-c}\big(s - c\,\mathrm{sgn}(s)\big) & \text{if} \quad b < |s| \le c\,, \\ 0 & \text{if} \quad c < |s|\;, \end{cases} \tag{3.17}$$

namely, Hampel's two-part redescending psi-function, with $b = a$, which has all the desirable properties.

3.2.2. An approximate optimality result

There is an approximate optimality result for the filter described above if we replace (3.14) by

$$s_t^2 \;=\; m_{11,t} + \sigma_0^2\;, \tag{3.18}$$

and ψ and w in (3.10) and (3.13), respectively, by

$$w(r) \;=\; \psi'(r) \;=\; \frac{\partial}{\partial r}\psi(r)\;. \tag{3.19}$$

Namely, under the assumption that the state prediction density $f_{\boldsymbol{x}_t}(\,.\,|\,\boldsymbol{Y}_{t-1})$ is Gaussian and that $\psi(r) = -(\partial/\partial r)\log g(r)$, where g is an approximation of the observation prediction density $f_{y_t}(\,.\,|\,\boldsymbol{Y}_{t-1})$, the filter is the conditional-mean filter proposed by Masreliez ([20]). The preceding assumption will never hold exactly under an AO model where v_t is non-Gaussian (see [15], Sec. 5). However, there is some evidence that $f_{\boldsymbol{x}_t}(\,.\,|\,\boldsymbol{Y}_{t-1})$ is nearly Gaussian and that the filter is a good approximation to the exact conditional-mean filter. Therefore the filter is referred to as an approximate conditional-mean (ACM) type filter. More details can be found in [15]. The results therein suggest that the use of Hampel's two-part redescending psi-function is reasonable when the observation noise v_t has a contaminated normal distribution. However, the weight function w given by (3.19) is discontinuous if using Hampel's two-part redescending psi-function, and therefore Martin and Thomson ([18]) prefer to specify w by (3.13).

3.2.3. Fixed-lag smoother-cleaners

As mentioned in [15], if one uses the last coordinate of the filter estimate $\widehat{\boldsymbol{x}}_{t|t}$ to produce cleaned data, then one has that $\widehat{x}_{t-p+1} = (\widehat{\boldsymbol{x}}_{t|t})_p$ is an estimate of x_{t-p+1} based on the observations $\boldsymbol{Y}_t$ up to time t. Such an estimate is usually called a fixed-lag smoother, with lag $p-1$ in this case.

3.2.4. Estimation of hyper parameters

To use the filter-cleaner algorithm we need robust estimates $\widehat{\boldsymbol{\phi}}$, $\hat{\sigma}_\varepsilon$ and $\widehat{\boldsymbol{C}}_{\boldsymbol{x}}$ of the AR(p) parameter vector $\boldsymbol{\phi} = (\phi_1, ..., \phi_p)^\top$, the innovations scale σ_ε and the $p \times p$ covariance matrix of the state process, respectively. Martin and Thomson ([18]) proposed to get initial estimates using bounded-influence autoregression (BIAR) via the *iteratively reweighted least squares* (IWLS) *algorithm*. Details about BIAR may be found in [19], [16] or [29].

3.2.5. Selection of order p

Martin and Thomson ([18]) propose the following procedure to select the order p of the autoregressive approximation. For increasing orders p BIAR estimates are computed and the estimated innovation scale estimates $\hat{\sigma}_\varepsilon(p)$ are examined for each order. The final order is selected as that value of p for which $\hat{\sigma}_\varepsilon(p+1)$ is not much smaller than $\hat{\sigma}_\varepsilon(p)$, e.g., less than a 10-percent decrement as suggested by Martin and Thomson ([18]).

Another robust order-selection rule based on BIAR estimates and motivated by Akaike's minimization criterion ([1]) was proposed by Martin ([16]).

3.3. The robust Least Squares (rLS) filter algorithm

In the following we describe a robustified version of the Kalman filter which was proposed by Ruckdeschel ([26]).

3.3.1. Robustified optimization problem

The idea is to reduce in the correction step (2.6) of the classical Kalman filter the influence of an observation $\boldsymbol{y}_t$ that is affected by an additive outlier. Instead of $\boldsymbol{K}_t \Delta\boldsymbol{y}_t$ with $\Delta\boldsymbol{y}_t = \boldsymbol{y}_t - \boldsymbol{H}\boldsymbol{x}_{t|t-1}$ we use a huberized version of it, i.e.,

$$H_{b_t}(\boldsymbol{K}_t \Delta\boldsymbol{y}_t) \;=\; \boldsymbol{K}_t \Delta\boldsymbol{y}_t \, \min\left\{1, \, \frac{b_t}{\|\boldsymbol{K}_t \Delta\boldsymbol{y}_t\|}\right\}, \tag{3.20}$$

so that the obtained result will be equal to the one of the classical Kalman filter, if $\|\boldsymbol{K}_t \Delta\boldsymbol{y}_t\|$ is not too large, whereas if $\|\boldsymbol{K}_t \Delta\boldsymbol{y}_t\|$ is too large, the direction

will remain unchanged and it will be projected on the q-dimensional ball with radius b_t.

This leads to a robustified optimization problem given by

$$E\big(\|\Delta \boldsymbol{x}_t - H_{b_t}(\boldsymbol{K}_t\,\Delta \boldsymbol{y}_t)\|^2\big) \;=\; \min_{\boldsymbol{K}_t}!\;, \tag{3.21}$$

where $\Delta \boldsymbol{x}_t = \boldsymbol{x}_t - \boldsymbol{x}_{t|t-1}$ denotes the prediction error. The above optimization problem is equivalent to the optimization problem (2.2) of the classical Kalman filter and its solution is named $\boldsymbol{K}_t^{\mathrm{rLS}}$.

3.3.2. The rLS filter

Hence, this gives us the following filter recursions:

(**i**) Initialization $(t=0)$:

$$\boldsymbol{x}_{0|0}^{\mathrm{rLS}} \;=\; \boldsymbol{\mu}_0\;; \tag{3.22}$$

(**ii**) Prediction $(t\geq 1)$:

$$\boldsymbol{x}_{t|t-1}^{\mathrm{rLS}} \;=\; \boldsymbol{\Phi}\,\boldsymbol{x}_{t-1|t-1}^{\mathrm{rLS}}\;; \tag{3.23}$$

(**iii**) Correction $(t\geq 1)$:

$$\boldsymbol{x}_{t|t}^{\mathrm{rLS}} \;=\; \boldsymbol{x}_{t|t-1}^{\mathrm{rLS}} + H_{b_t}\big(\boldsymbol{K}_t^{\mathrm{rLS}}(\boldsymbol{y}_t - \boldsymbol{H}\boldsymbol{x}_{t|t-1}^{\mathrm{rLS}})\big)\;. \tag{3.24}$$

The above filter recursions will be named *robust least squares* (rLS) *filter*.

Because the calculation of $\boldsymbol{K}_t^{\mathrm{rLS}}$ is computationally extensive Ruckdeschel ([26]) proposes to use $\boldsymbol{K}_t^{\mathrm{KK}}$ instead where $\boldsymbol{K}_t^{\mathrm{KK}}$ denotes the Kalman gain obtained by the classical Kalman filter recursions. Simulation studies therein have shown that the worsening, in sense of a larger mean-squared error, is only small if using $\boldsymbol{K}_t^{\mathrm{KK}}$ instead of $\boldsymbol{K}_t^{\mathrm{rLS}}$. Hence, this simplifying modification almost yields the classical Kalman filter recursions with the only exception of replacing the first line of the correction step (2.6) by

$$\boldsymbol{x}_{t|t} \;=\; \boldsymbol{x}_{t|t-1} + H_{b_t}\big(\boldsymbol{K}_t^{\mathrm{KK}}(\boldsymbol{y}_t - \boldsymbol{H}\boldsymbol{x}_{t|t-1})\big)\;. \tag{3.25}$$

From now on, if speaking of the rLS filter, we will only consider this modified version.

Moreover, Ruckdeschel ([26]) proved that the rLS filter is SO-optimal under certain side conditions. SO stands for substitutive outlier and means that, instead

of disturbing v_t, contamination effects y_t directly, replacing it by an arbitrarily distributed variable y_t' with some low probability. For further details we refer the reader to [26].

Still, the open problem of fixing the clipping height b_t remains.

3.3.3. Fixing the clipping height b_t

In order to properly choose b_t Ruckdeschel ([26]) proposes an assurance criterion: How much efficiency in the ideal model relative to the optimal procedure, i.e., the Kalman filter, am I ready to pay in order to get robustness under deviations from the ideal model? This loss of efficiency, which we will obtain if we use a robust version instead of the classical Kalman filter, is quantified as the relative worsening of the mean-squared error in the ideal model. Hence, for a given relative worsening $\delta > 0$ we solve

$$E\Big(\big\|\Delta\boldsymbol{x}_t - H_{b_t}(\boldsymbol{K}_t^{\mathrm{rLS}}\Delta\boldsymbol{y}_t)\big\|^2\Big) \stackrel{!}{=} (1+\delta)\, E\Big(\big\|\Delta\boldsymbol{x}_t - \boldsymbol{K}_t^{\mathrm{KK}}\Delta\boldsymbol{y}_t\big\|^2\Big)\ . \tag{3.26}$$

The symbol $\stackrel{!}{=}$ means that b_t is chosen in a way to achieve equality.

Again, we use the simplifying modifications just mentioned and replace $\boldsymbol{K}_t^{\mathrm{rLS}}$ by $\boldsymbol{K}_t^{\mathrm{KK}}$. Moreover, in most time-invariant situations, the sequence of $\boldsymbol{M}_t$ (and hence also of $\boldsymbol{P}_t$ and $\boldsymbol{K}_t^{\mathrm{KK}}$) stabilizes due to asymptotic stationarity. Thus, once $\boldsymbol{M}_t$ does not change for more than a given tolerance level, we can stop calibration and use the last calculated b_t for all subsequent times s, $s > t$. The Kalman gain and filtering error covariance matrix used in this last calibration step will be denoted by $\boldsymbol{K}_\infty^{\mathrm{KK}}$ and $\boldsymbol{P}_\infty$, respectively. For details we refer to [2] and [21]. Further we make another simplifying modification and assume that for all t

$$\Delta\boldsymbol{x}_t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{M}_t) \qquad \text{and} \qquad \boldsymbol{v}_t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{R})\ . \tag{3.27}$$

Thus, we may solve

$$\begin{aligned} E\Big(\big\|\Delta\boldsymbol{x} - H_b(\boldsymbol{K}_\infty^{\mathrm{KK}}\Delta\boldsymbol{y})\big\|^2\Big) &\stackrel{!}{=} (1+\delta)\, E\Big(\big\|\Delta\boldsymbol{x} - \boldsymbol{K}_\infty^{\mathrm{KK}}\Delta\boldsymbol{y}\big\|^2\Big) \\ &= (1+\delta)\ \operatorname{tr}\boldsymbol{P}_\infty\ , \end{aligned} \tag{3.28}$$

in b, uniquely for a given loss of efficiency δ, where $\operatorname{tr}\boldsymbol{P}_\infty$ denotes the trace of the conditional filtering error covariance matrix. We note that the relative time-expensive calibration, i.e., finding b to a given δ, can be done beforehand. Additional details may be found in [26] and [25].

4. SIMULATION STUDY

The outline of our simulation study is as follows: First we simulate a core process x_t of length $n = 100$. x_t is chosen to be an autoregressive process of order 2 given by

$$x_t = x_{t-1} - 0.9\, x_{t-2} + \varepsilon_t \ , \tag{4.1}$$

with $\varepsilon_t \sim \mathcal{N}(0,1)$. The variance of the core process x_t, i.e., the value of the autocovariance function at lag zero can be calculated by numerical integration and is given approximately by $\mathrm{var}(x_t) \approx 7.27$. Additionally, the additive outliers are simulated from a contaminated normal distribution with degenerate central component (1.2) with $\sigma^2 = 10^2$. The contamination γ is varied from 0% to 20% by steps of 5%. That means that with probability γ, v_t is an additive outlier with $v_t \neq 0$. To obtain the contaminated process y_t, the v_t's are added to the core process x_t. For each level of contamination this was done 400 times.

For each of the contaminated series, estimates of the hyper parameter, i.e., the innovations scale $\widehat{\sigma}_\varepsilon$, the autoregressive parameters $\widehat{\phi}_1, ..., \widehat{\phi}_p$ and the $p \times p$ covariance matrix $\widehat{\boldsymbol{C}}_{\boldsymbol{x}}$ of the state process $\boldsymbol{x}_t$, are computed via bounded-influence autoregression. The order p of the autoregressive approximation is chosen according to the order-selection criterion proposed by Martin and Thomson ([18]), which yields values of p from 2 to 3 subject to the contamination level. In order to be able to compare the results we choose an equal order p for all levels of contamination and fix it equal to 3. Using an order $p = 2$ in cases of lower contamination levels, where this is appropriate, we obtain almost perfect fits for both filtering algorithms. But, although the simulated core process is of order 2, the estimated BIAR parameters we obtain setting p equal to 3 are similar to the ones of the original core process, i.e., the first two AR parameters are close to the original ones and the third AR parameter is almost zero, as one would expect.

Then each process is cleaned using the ACM-type filter and the rLS filter proposed by Martin and Thomson ([18]) and Ruckdeschel ([26]), respectively. Afterwards, the hyper parameters of the filtered series are estimated again.

Those re-estimated hyper parameters are used to calculate a prewhitened spectral density estimate for each process. Last, the deviation of each estimated spectral density function from the true spectral density function is measured in the sense of the squared L_2-norm, i.e.,

$$err^2_{\widehat{S}(f)} := \left\| \widehat{S}(f) - S(f) \right\|^2 = \int \left(\widehat{S}(f) - S(f) \right)^2 df \ , \tag{4.2}$$

where $\widehat{S}(f)$ and $S(f)$ denote the estimated and true spectral density functions.

5. RESULTS

Regarding the computation time the rLS filter performs better than the ACM-type filter as we expected. This is due to the fact that additional weights have to be computed within the correction step of the ACM-type filter.

Figure 3 tries to visualize the results of our simulation study. For both methods and contamination levels 0%, 10% and 20% seven curves are plotted on a logarithmic scale. The thick line represents the true spectral density function, whereas the thin line is the spectral density estimate of one realization out of 400. Moreover, we may calculate the minimum and maximum, at each frequency, the first and third quartile and median value of all spectral density estimates. Connecting all median values we obtain the grey line, to which we will refer hereafter as median spectral density function. In the same sense we refer to all minimum values as minimum spectral density function, and so on. Hence, the lower and upper dotted lines are the minimum and maximum spectral density functions, whereas the lower and upper dashed lines represent the first and third quartile spectral density functions. The results obtained by using the ACM-type filter are plotted in the left column, whereas the results of the rLS filter are displayed in the right column.

As expected, for both methods the dispersion of the spectral density estimates becomes greater the higher the contamination. However, this effect is more visible, especially at higher frequencies, when using the ACM-type filter.

Next, we try to visualize the squared errors of the estimated spectral density functions. First, the logarithm of the squared errors is taken. For both methods Figure 4 shows boxplots of the squared errors in eight equally-sized frequency bands as well as the total squared errors (bottom right) for all different levels of contamination. Again, the squared errors become greater the higher the contamination, especially at higher frequencies. And, this effect again is greater, when using the ACM-type filter. However, these errors are very small and, looking at the total squared errors for different contamination levels, we see that the ACM-type filter performs better than the rLS filter. The greatest contribution to the total squared error is the amount of the frequency band where the spectral density function has its peak. There the squared errors using the rLS filter are higher than the ones using the ACM-type filter. Moreover, we see that all squared errors are in the same range for all contamination levels.

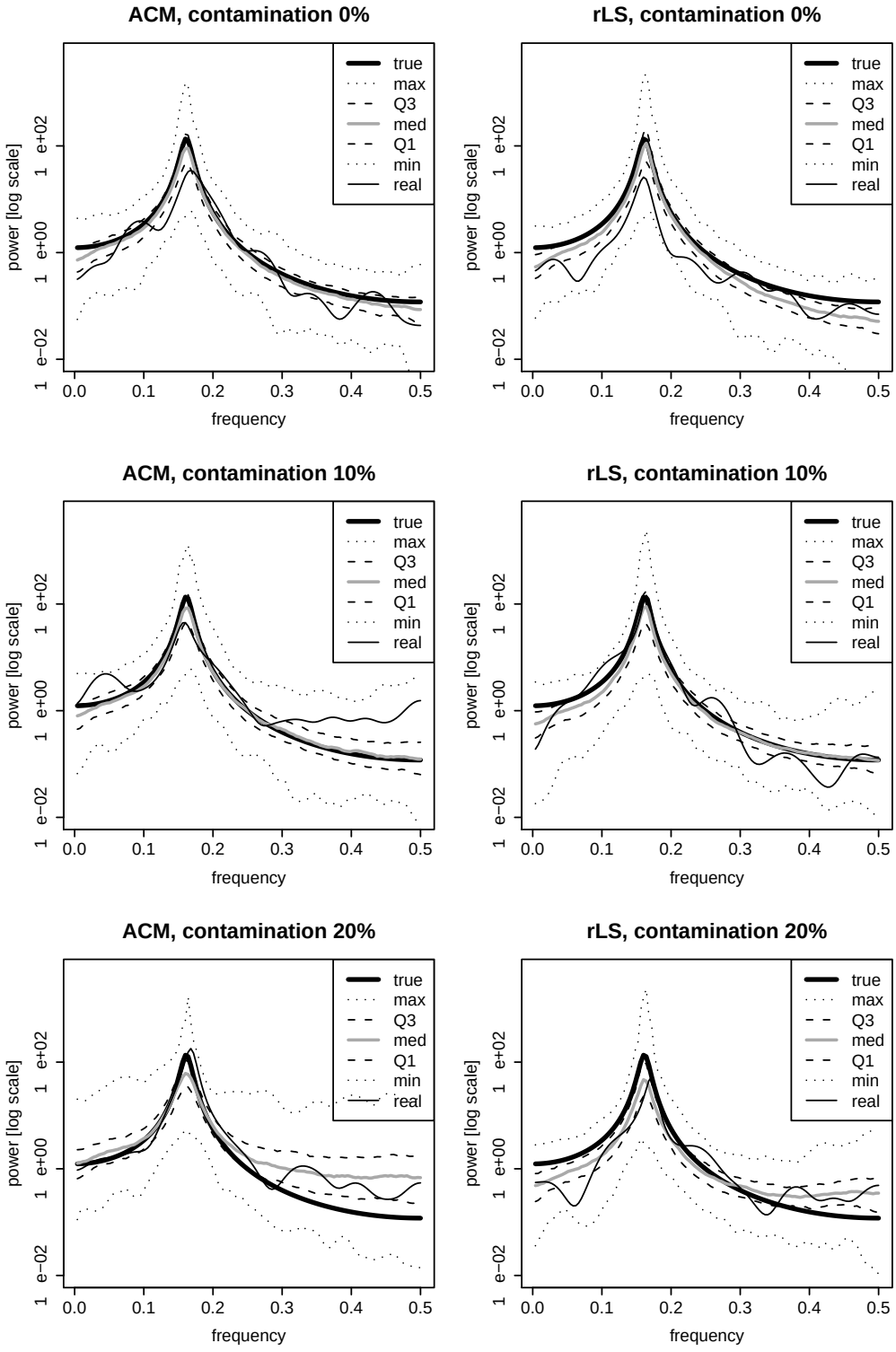

Figure 3: Robust spectral density estimates of the simulated data, left column 'ACM', right 'rLS'.

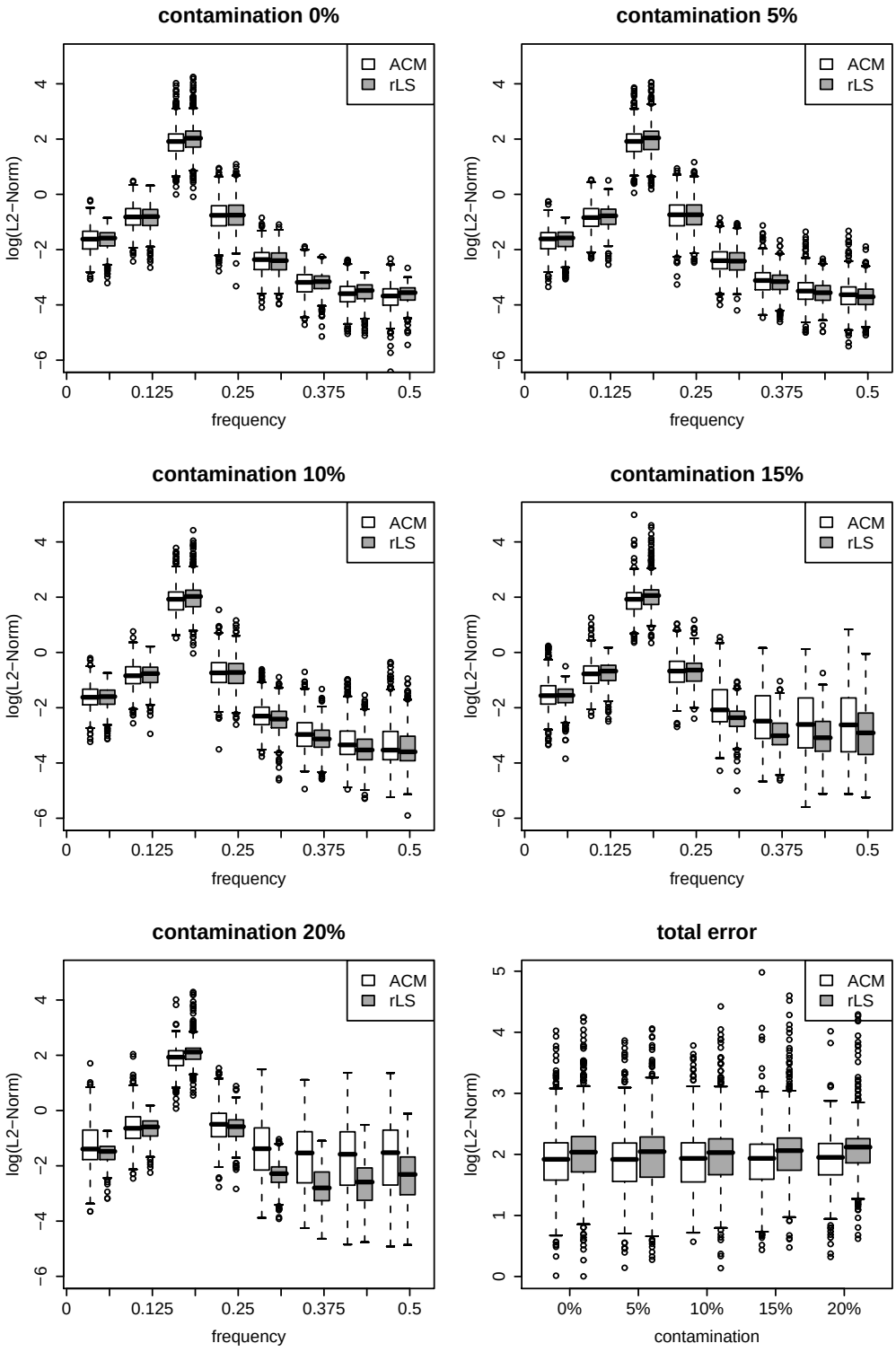

Figure 4: Boxplots of the errors.

6. DISCUSSION

In order to get a robust estimate of the spectral density function, it turns out that cleaning the series in a robust way first and calculating a prewhitened spectral density estimate afterwards leads to encouraging results. This data-cleaning operation wherein the robustness is introduced, is solved by two different robustified versions of the Kalman filter. Although, as far as we know, there exist no theoretical results on the statistical properties of both proposed multi-step procedures, the empirical results based on simulations and real data sets promise those procedures to be of high quality. The results of the simulation study suggest that the ACM-type filter algorithm performs slightly better than the rLS filter algorithm. Hence, the ACM-type filter algorithm was used to compute the robust spectral density estimates shown in Figure 2.

In [28] we compare the ACM-type filter approach with another approach proposed by Tatum and Hurvich ([30]). This procedure, called biweight filter-cleaner, also yields good results, but tends to underestimate the core process slightly. Moreover it is computational intensive.

The problem of estimating the hyper parameters was accomplished by bounded-influence autoregression. An alternative way would be to use a highly robust autocovariance function estimator (cf. [13]) and calculate estimates of the hyper parameters via the Yule-Walker equations. Hyper parameters may also be obtained by computing a robust covariance matrix via the MCD algorithm (cf. [24]) and estimate the parameters again using the Yule-Walker equations. Recently, Maronna et al. ([14]) propose to use τ-estimates. Our experience by now is that all these different approaches (except the last one, which we have not tried yet, although it seems worthwhile) leads to similar results.

The simulation study was only done for one specific autoregressive model of order 2. Other models seem worth trying. Further research and additional simulation studies have already been done, but, as well as the applications to the motivating real data, are not published here.

ACKNOWLEDGMENTS

The computation and all graphics were done with R (cf. [23]). The rLS filter algorithm was implemented by Peter Ruckdeschel. Moreover, an R-package, containing the BIAR algorithm, the ACM-type filter algorithm and the rLS filter algorithm, exists and is available at `www.uni-bayreuth.de/departments/math/org/mathe7/robKalman/`. We also like to thank two referees for their valuable suggestions.

REFERENCES

[1] Akaike, H. (1974). A new look at the statistical model identification, *IEEE Transactions on Automatic Control*, **AC-19**, 716–722.

[2] Anderson, B. and Moore, J. (1979). *Optimal Filtering*, Prentice Hall, Englewood Cliffs, NJ.

[3] Blackman, R. and Tukey, J. (1958). *The Measurement of Power Spectra*, Dover, New York.

[4] Chave, A.; Thomson, D. and Ander, M. (1987). On the robust estimation of power spectra, coherences, and transfer functions, *J. Geophys. Res.*, **92**(B1), 633–648.

[5] Durbin, J. and Koopman, S. (2001). *Time Series Analysis by State Space Methods*, Oxford University Press, New York.

[6] Fox, A. (1972). Outliers in time series, *J. Royal Statist. Soc.*, **34**(3), 350–363.

[7] Hampel, F. (1968). *Contributions to the Theory of Robust Estimation*, PhD thesis, University of California, Berkeley.

[8] Hartikainen, J.; Tahvanainen, K. and Kuusela, T. (1998). *Short-term measurement of heart rate variability.* In "Clinical Guide to Cardiac Autonomic Tests" (Malik, Ed.), Kluwer, Dordrecht, 149–176.

[9] Jones, A. and Hollinger, K. (1997). Spectral analysis of the KTB sonic and density logs using robust nonparametric methods, *J. Geophys. Res.*, **102**(B8), 18391–18403.

[10] Kalman, R. (1960). A new approach to linear filtering and prediction problems, *J. Basic Eng. – Trans. ASME*, **82**, 35–45.

[11] Kalman, R. and Bucy, R. (1961). New results in filtering and prediction theory, *J. Basic Eng. – Trans. ASME*, **83**, 95–108.

[12] Kleiner, R.; Martin, R. and Thomson, D. (1979). Robust estimation of power spectra, *J. Royal Statist. Soc. B*, **41**(3), 313–351.

[13] Ma, Y. and Genton, M. (2000). Highly robust estimation of the autocovariance function, *J. Time Series Analysis*, **21**(6), 663–684.

[14] Maronna, R.; Martin, R. and Yohai, V. (2006). *Robust Statistics: Theory and Methods*, John Wiley, New York.

[15] Martin, R. (1979). *Approximate conditional-mean type smoothers and interpolators.* In "Smoothing Techniques for Curve Estimation" (Gasser and Rosenblatt, Eds.), Springer, New York.

[16] Martin, R. (1980). *Robust estimation of autoregressive models.* In "Directions in Time Series" (D. Brillinger and G. Tiao, Eds.), Inst. Math. Statist. Publications, Haywood, CA, 228–254.

[17] Martin, R. (1981). *Robust methods for time series.* In "Applied Time Series II" (Findley, Ed.), Academic Press, New York.

[18] Martin, R. and Thomson, D. (1982). Robust-resistant spectrum estimation, *Proceedings of the IEEE*, **70**, 1097–1115.

[19] Martin, R. and Zeh, J. (1978). *Generalized M-estimates for autoregressions, including small-sample efficiency robustness*, Technical Report 214, Dept. of Electrical Engineering, Univ. Washington, Seattle.

[20] Masreliez, C. (1975). Approximate non-Gaussian filtering with linear state and observation relations, *IEEE Transactions on Automatic Control*, **AC-20**, 107–110.

[21] Moore, J. and Anderson, B. (1980). Coping with singular transition matrices in estimation and control stability theory, *Int. J. Control*, **31**, 571–586.

[22] Pumprla, J.; Howorka, K.; Groves, D.; Chester, M. and Nolan, J. (2002). Functional assessment of heart rate variability: physiological basis and practical applications, *Int. J. Cardiology*, **84**, 1–14.

[23] R Development Core Team (2005). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.

[24] Rousseeuw, P. and Van Driessen, K. (1999). A fast algorithm for the minimum covariance determinant estimator, *Technometrics*, **41**, 212–223.

[25] Ruckdeschel, P. (2000). *Robust kalman filtering.* In "XploRe. Application Guide" (Härdle, Hlávka and Klinke, Eds.), Springer, New York, Chapter 18, 483–516.

[26] Ruckdeschel, P. (2001). *Ansätze zur Robustifizierung des Kalman-Filters*, Volume 64 of *Bayreuther Mathematische Schriften*, Mathematisches Institut, Universität Bayreuth, Bayreuth, PhD thesis.

[27] Shumway, R. and Stoffer, D. (2000). *Time Series Analysis and Its Applications*, Springer, New York.

[28] Spangl, B. and Dutter, R. (2005). On robust estimation of power spectra, *Austrian Journal of Statistics*, **34**(2), 199–210.

[29] Stockinger, N. and Dutter, R. (1987). Robust time series analysis: A survey, *Kybernetika*, Supplement **23**, 1–90.

[30] Tatum, L. and Hurvich, C. (1993). *A frequency domain approach to robust time series analysis.* In "New Directions in Statistical Data Analysis and Robustness" (Morgenthaler, Ronchetti and Stahel, Eds.), Birkhäuser-Verlag, Basel.

[31] Thomson, D. (1994). An overview of multiple-window and quadratic-inverse spectrum estimation methods, *Proceedings of the IEEE ICASSP*, **6**, 185–194.