Fluctuation and Noise Letters
Vol. 15, No. 1 (2016) 1650009 (26 pages)

DOI: 10.1142/S0219477516500097

Singular Spectrum Analysis Based on L_1-Norm

Mahdi Kalantari*, Masoud Yarmohammadi* and Hossein Hassani†,‡

*Department of Statistics
Payame Noor University, 19395–4697, Tehran, Iran

†Institute for International Energy Studies
Tehran 1967743 711, Iran

‡hassani.stat@gmail.com

Received 2 August 2015
Accepted 5 January 2016
Published 18 March 2016

Communicated by Andras Der

In recent years, the singular spectrum analysis (SSA) technique has been further developed and increasingly applied to solve many practical problems. The aim of this research is to introduce a new version of SSA based on L_1-norm. The performance of the proposed approach is assessed by applying it to various real and simulated time series, especially with outliers. The results are compared with those obtained using the basic version of SSA which is based on the *Frobenius* norm or L_2-norm. Different criteria are also examined including reconstruction errors and forecasting performances. The theoretical and empirical results confirm that SSA based on L_1-norm can provide better reconstruction and forecasts in comparison to basic SSA when faced with time series which are polluted by outliers.

Keywords: Singular spectrum analysis; forecasting; reconstruction; outliers; L_1-norm.

1. Introduction

Econometric methods have been widely used to forecast the evolution of quarterly and yearly national account data. However, the standard time series forecasting models may fail to accurately predict economic and financial time series. This may be due to technological advances, changes in government policies and consumer preferences which create structural breaks and cause the data to be non-stationary in either mean or variance [1]. Furthermore, for standard forecasting methods such as ARIMA, assuming stationarity for the data, linearity for the model and normality for the residuals can provide only an approximation to the true situation [2]. Therefore, a method which does not depend on these assumptions could be very useful for modeling and forecasting economic data. The singular spectrum analysis (SSA) technique is a very good example of such a method [3].

Recently, the SSA technique has witnessed an increased application for solving real world problems (see for example [2, 4–10, 13]). What is interesting is that these applications cover a wide range of topics from economics and finance to medicine and audiology and so on. In brief, the aim of the SSA technique is to decompose and filter a time series via signal extraction which enables a forecasting from a less noisy time series. At present, the most basic version of SSA is based on the *Frobenius* norm which is sometimes referred to as the L_2-norm. The aim of this paper is to introduce a novel version of SSA which is based on the L_1-norm. In assessing the performance of the proposed approach, we consider different criteria in relation to reconstruction and forecasting via both simulated and real data, and give consideration to outliers.

We believe it is important to briefly comment on the impact of outliers on time series analysis and forecasting. The presence of outliers and shocks could make a time series non-stationary [14] and easily mislead time series analysis, forecasting results and regression analysis (see, for example [15–18]). It is widely believed that parametric methods are more sensitive than non-parametric methods to the presence of outliers. For example, outliers have negative impacts on the structure of autocorrelation in time series and that ARMA models are affected through the bias caused in autocorrelation function (ACF) and partial autocorrelation function (PACF) [19]. More relevant examples on the impact of outliers on parametric models can be found in *inter-alia* [20, 21]. In terms of the impact of outliers on SSA, in Ref. 14 it was found that outliers have a significant impact on SSA reconstruction and forecasts. As such, it is most certainly important to give due consideration to outliers and how it impacts the proposed SSA based on L_1-norm.

It should be mentioned that the L_1-norm based least absolute deviation (LAD) regression (basically that was used for robust SSA in Ref. 16) has breakdown point of 0%, because of the leverage points. However, the breakdown point can be as large as 25% using optimal design [22]. There exists some other robust regression methods (see for example [23]). A robust SSA based on robust PCA has been developed in Ref. 24.

The importance of this study can be further evidenced as follows. In doing so, we seek to explain its importance in simple terms so as to enable better understanding. A more technical explanation is presented later on. Firstly, we know that the basic SSA based on the L_2-norm is sensitive to the presence of outliers as was found in [14]. This was partly expected as the L_2-norm is concerned with seeking an average or mean which is by definition affected by outliers. In contrast, the proposed L_1-norm based SSA suggests looking for and exploiting the median which in turn should result in little or no sensitiveness to outliers in time series. If that is the case, this new version of SSA would be able to provide a comparatively improved reconstruction and forecast for time series with outliers which would further enhance the applicability and relevance of SSA for solving real world problems in the field of time series analysis and forecasting. Secondly, positive results from this research would imply that practitioners can save time and ensure there is no loss of information as there will then be a possibility of modeling any given time series as it is without having to remove outliers from the data.

The remainder of this paper is structured as follows. The new SSA version and related theoretical background are presented in Sec. 2. The performance of the new version and currently used SSA technique are evaluated in Sec. 3. Finally, Sec. 4 presents a summary of the study and some concluding remarks.

2. SSA Based on L_1-Norm

2.1. *Decomposition and reconstruction*

2.1.1. *Stage* 1: *Decomposition*

Step 1. Embedding and the Trajectory Matrix

In explaining the SSA process we mainly follow [25]. Consider a univariate stochastic process $\{Y_t\}_{t\in\mathbb{Z}}$ and suppose that a realization of size N from this process is available $Y_N = \{y_1, y_2, \ldots, y_N\}$. The single SSA choice relevant to the embedding step is the *window length* L, an integer such that $2 \le L \le N$. Embedding can be regarded as a mapping operation that transfers a one-dimensional time series Y_N into a multidimensional series $X_1, \ldots, X_K$ with vectors

$$X_i = (y_i, y_{i+1}, y_{i+2}, \ldots, y_{i+L-1})^T, \tag{1}$$

for $i = 1, 2, \ldots, K$ where $K = N - L + 1$. The result of this step is the trajectory matrix

$$\mathbf{X} = [X_1, \ldots, X_K] = (x_{ij})_{i,j=1}^{L,K} = \begin{pmatrix} y_1 & y_2 & y_3 & \cdots & y_K \\ y_2 & y_3 & y_4 & \cdots & y_{K+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ y_L & y_{L+1} & y_{L+2} & \cdots & y_N \end{pmatrix}. \tag{2}$$

It is noteworthy that the trajectory matrix $\mathbf{X}$ is a Hankel matrix, which in turn means that all the elements along the diagonal $i + j = \text{const.}$ are equal.

Step 2. Singular Value Decomposition (SVD)

The SVD of the trajectory matrix $\mathbf{X}$ can be written as:

$$\mathbf{X} = \mathbf{U\Sigma V}^T,$$

where $\mathbf{U} = [U_1 : \cdots : U_L] \in \mathbb{R}^{L\times L}$, $\mathbf{V} = [V_1 : \cdots : V_L] \in \mathbb{R}^{K\times L}$ and $\Sigma = \text{diag}(\sqrt{\lambda_1} \ge \sqrt{\lambda_2} \ge \cdots \ge \sqrt{\lambda_L}) \in \mathbb{R}^{L\times L}$. Here, $\lambda_1, \ldots, \lambda_L$ are denoted as the eigenvalues of $\mathbf{S} = \mathbf{XX}^T$ in decreasing order of magnitude $(\lambda_1 \ge \cdots \lambda_L \ge 0)$ and $U_1, \ldots, U_L$ are the corresponding eigenvectors. Set $d = \max(i, \text{such that}\, \lambda_i > 0) = \text{rank}\,\mathbf{X}$. If we denote $V_i = \mathbf{X}^T U_i/\sqrt{\lambda_i}$, then the SVD of the trajectory matrix can be written as:

$$\mathbf{X} = \mathbf{X}_1 + \cdots + \mathbf{X}_d, \tag{3}$$

where $\mathbf{X}_i = \sqrt{\lambda_i} U_i V_i^T$ $(i = 1, \ldots, d)$. The matrices $\mathbf{X}_i$ have rank 1.

Note that SVD (3) is optimal in the sense that among all the matrices of rank $r < d$, the matrix $\mathbf{X}^{(r)} = \sum_{i=1}^{r} \mathbf{X}_i$ provides the best approximation to the trajectory matrix $\mathbf{X}$, so that $\|\mathbf{X} - \mathbf{X}^{(r)}\|_F$ is minimum, where $\|.\|_F$ is the *Frobenius* norm.

Here, we perform L_1-Decomposition of the trajectory matrix $\mathbf{X}$ with obtaining the SVD of the trajectory matrix $\mathbf{X}$ and representing it as a sum of rank-one bi-orthogonal elementary matrices.

Suppose that the trajectory matrix $\mathbf{X}$ is rank deficient, i.e., $r = \text{rank}(\mathbf{X}) < L < K$. Define the diagonal matrix $\mathbf{W}$ as $\mathbf{W} = \text{diag}(\underbrace{w_1, w_2, \ldots, w_r}_{r}, \underbrace{0, 0, \ldots, 0}_{L-r}) \in \mathbb{R}^{L \times L}$.

Note that

$$\mathbf{W} = \left(\begin{array}{cccc|ccc} w_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & w_2 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_r & 0 & \cdots & 0 \\ \hline 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \end{array}\right) = \begin{pmatrix} \mathbf{W}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix},$$

where $\mathbf{W}_r = \text{diag}(w_1, w_2, \ldots, w_r) \in \mathbb{R}^{r \times r}$.

For L_1-norm approximation of the signal, one seeks to find the diagonal matrix $W_{L \times L}$ such that $\|\mathbf{X} - \mathbf{U}\mathbf{W}\boldsymbol{\Sigma}\mathbf{V}^T\|_{L_1}$ is minimized; more precisely, the following minimization problem must be solved:

$$\min_{\mathbf{W}} \|\mathbf{X} - \mathbf{U}\mathbf{W}\boldsymbol{\Sigma}\mathbf{V}^T\|_{L_1}.$$

Note that

$$\begin{aligned} \mathbf{U}\mathbf{W}\boldsymbol{\Sigma}\mathbf{V}^T &= [\mathbf{U}_1 \quad \mathbf{U}_2] \begin{bmatrix} \mathbf{W}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma}_1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_2 \end{bmatrix} \begin{bmatrix} \mathbf{V}_1^T \\ \mathbf{V}_2^T \end{bmatrix} \\ &= [\mathbf{U}_1\mathbf{W}_r \quad \mathbf{0}] \begin{bmatrix} \boldsymbol{\Sigma}_1\mathbf{V}_1^T \\ \boldsymbol{\Sigma}_2\mathbf{V}_2^T \end{bmatrix} \\ &= \mathbf{U}_1\mathbf{W}_r\boldsymbol{\Sigma}_1\mathbf{V}_1^T, \end{aligned} \tag{4}$$

where $\mathbf{U}_1 \in \mathbb{R}^{L \times r}$, $\mathbf{V}_1 \in \mathbb{R}^{K \times r}$ and $\boldsymbol{\Sigma}_1 \in \mathbb{R}^{r \times r}$. It can be easily shown that

$$\mathbf{U}_1\mathbf{W}_r\boldsymbol{\Sigma}_1\mathbf{V}_1^T = \sum_{i=1}^{r} \sqrt{\lambda_i} w_i U_i V_i^T. \tag{5}$$

By (4), we must solve the following minimization problem:

$$\min_{\mathbf{W}_r} \|\mathbf{X} - \mathbf{U}_1\mathbf{W}_r\boldsymbol{\Sigma}_1\mathbf{V}_1^T\|_{L_1}.$$

For this purpose, suppose that $\mathbf{A}_r = \mathbf{U}_1\mathbf{W}_r$ and $\mathbf{B}_1 = \boldsymbol{\Sigma}_1\mathbf{V}_1^T$. As matrices $\mathbf{B}_1$ and $\mathbf{U}_1$ are known, the goal is to find the matrix $\mathbf{A}_r$ such that $\|\mathbf{X} - \mathbf{A}_r\mathbf{B}_1\|_{L_1}$ is minimized. On the other hand, for signal approximation the following minimization problem must be solved:

$$\min_{\mathbf{A}_r} \|\mathbf{X} - \mathbf{A}_r\mathbf{B}_1\|_{L_1}. \tag{6}$$

For solving (6), let $\mathbf{S} = \mathbf{A}_r\mathbf{B}_1$. Then $\mathbf{S}^T = \mathbf{B}_1^T\mathbf{A}_r^T$ and

$$S_j^T = \mathbf{B}_1^T A_j^T, \tag{7}$$

where A_j^T and S_j^T are jth column of matrices $\mathbf{A}_r^T$ and $\mathbf{S}^T$, respectively.

By consideration of (10), (11), and (7) we have:

$$\begin{aligned}\|\mathbf{X} - \mathbf{S}\|_{L_1} &= \|(\mathbf{X} - \mathbf{S})^T\|_{L_1} \\ &= \|\mathbf{X}^T - \mathbf{S}^T\|_{L_1} \\ &= \sum_{j=1}^{L} \|X_j^T - S_j^T\|_{L_1} \\ &= \sum_{j=1}^{L} \|X_j^T - \mathbf{B}_1^T A_j^T\|_{L_1}, \end{aligned} \tag{8}$$

where X_j^T is jth column of matrix $\mathbf{X}^T$. The relationship in Eq. (8) shows that for minimization of $\|\mathbf{X} - \mathbf{S}\|_{L_1}$, it is sufficient to minimize $\|X_j^T - \mathbf{B}_1^T A_j^T\|_{L_1}$ with respect to A_j^T.

Fortunately, this minimization problem is a well known problem in LAD Regression Analysis (or L_1-Regression). If $X_j^T, \mathbf{B}_1^T$ and A_j^T are considered as a vector of a dependent variable, a matrix of observed independent variables and unknown vector of coefficients respectively, we can find vector A_j^T by using an iterative numerical method described in Refs. 26 and 27. By computing vector A_j^T, matrix $\mathbf{A}_r$ is achieved.

For LAD Regression Analysis, one can use the `l1fit` function in S-PLUS, which does not run in `R`. Hence, we use function `rq` from `quantreg` package with $\tau = 0.5$ in the `R` software which is freely accessible.

Recall that the L_1-norm of a matrix $\mathbf{A} = (a_{ij})_{i,j=1}^{m,n}$ is defined as follows:

$$\|\mathbf{A}\|_{L_1} = \sum_{i=1}^{m}\sum_{j=1}^{n} |a_{ij}|. \tag{9}$$

It is obvious that:

$$\|\mathbf{A}\|_{L_1} = \|\mathbf{A}^T\|_{L_1} \tag{10}$$

and

$$\|\mathbf{A}\|_{L_1} = \sum_{j=1}^{n} \|A_j\|_{L_1}, \tag{11}$$

where A_j is the jth column of $\mathbf{A}$, $\|A_j\|_{L_1}$ is the L_1-norm of the vector $A_j = (a_{1j}, a_{2j}, \ldots, a_{mj})^T$ defined as $\|A_j\|_{L_1} = \sum_{i=1}^{m} |a_{ij}|$.

Note that in basic SSA, the *Frobenius* norm is used for signal approximation and defined as follows:

$$\|\mathbf{A}\|_F^2 = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}^2.$$

Sometimes this norm is called L_2-norm.

2.1.2. *Stage* 2: *Reconstruction*

Step 1. Grouping

The grouping step corresponds to splitting the elementary matrices into several groups and summing the matrices within each group. Let $I = \{i_1, \ldots, i_p\}$ be a group of indices $i_1, \ldots, i_p$. Then the matrix $\mathbf{X}_I$ corresponding to the group I is defined as $\mathbf{X}_I = \mathbf{X}_{i_1} + \cdots + \mathbf{X}_{i_p}$. The spilt of the set of indices $\{1, \ldots, L\}$ into disjoint subsets $I_1, \ldots, I_m$ corresponds to the representation $\mathbf{X} = \mathbf{X}_{I_1} + \cdots + \mathbf{X}_{I_m}$. The procedure of choosing the sets $I_1, \ldots, I_m$ is called grouping. For a given group I, the contribution of the component $\mathbf{X}_I$ is measured by the share of the corresponding eigenvalues: $\sum_{i\in I} \lambda_i / \sum_{i=1}^{d} \lambda_i$.

Step 2. Diagonal averaging and Hankelization

In this step, we seek to transform each matrix $\mathbf{X}_{I_j}$ of the grouping step into a new series of length N. In basic SSA, Hankelization is obtained via diagonal averaging. Note that this Hankelization has an optimal property in the sense that the matrix $\mathcal{H}\mathbf{A}$ is the nearest to $\mathbf{A}$ (with respect to the *Frobenius* norm) among all Hankel matrices of the same dimension [3]. On the other hand, $\|\mathbf{A} - \mathcal{H}\mathbf{A}\|_F^2$ is minimum; so we denote this type of Hankelization that uses the diagonal averaging as L_2-Hankelization. In the following theorem, we propose a procedure for Hankelization based on L_1-norm.

Theorem (L_1-Hankelization). *Let* $\mathbf{A}$ *be a* $L \times K$ *matrix and* $s = i + j (2 \leq s \leq L + K)$, *then the element* $\widetilde{a}_{ij}$ *of the matrix* $\mathcal{H}\mathbf{A}$ *with respect to* L_1*-norm is*

$$\widetilde{a}_{ij} = \operatorname*{median}_{(l,k)\in A_s} a_{lk}, \tag{12}$$

where $A_s = \{(l, k) : l + k = s, 1 \leq l \leq L, 1 \leq k \leq K\}$.

Proof. By definition, a Hankel matrix $\mathbf{B} = \mathcal{H}\mathbf{A}$ with elements b_{ij} satisfies the conditions $b_{ij} = g_s$ for $i + j = s$ and some numbers g_s. By (9) we have:

$$\begin{aligned} \|\mathbf{A} - \mathcal{H}\mathbf{A}\|_{L_1} = \|\mathbf{A} - \mathbf{B}\|_{L_1} &= \sum_{i=1}^{L}\sum_{j=1}^{K} |a_{ij} - b_{ij}| \\ &= \sum_{s=2}^{L+K}\sum_{i+j=s} |a_{ij} - g_s|. \end{aligned} \tag{13}$$

It is well known that the numbers g_s that minimizes the right-hand side of the relation (13) are the median of A_s elements. □

It is obvious that L_1-Hankelization corresponds to compute median of the matrix elements over the "antidiagonal".

L_1-Hankelization (12) applied to a resultant matrix $\mathbf{X}_{I_j}$ of the grouping step, produces a *reconstructed series* $\widetilde{Y}_N^{(j)} = \{\widetilde{y}_1^{(j)}, \ldots, \widetilde{y}_N^{(j)}\}$. Therefore, the initial series $\{y_1, \ldots, y_N\}$ is decomposed into a sum of m reconstructed series:

$$y_t = \sum_{j=1}^{m} \widetilde{y}_t^{(j)}, \quad t = 1, 2, \ldots, N.$$

2.2. *Forecasting*

An important advantage of SSA is that it allows to produce forecasts for either the individual components of the series and/or the reconstructed series itself. In general,

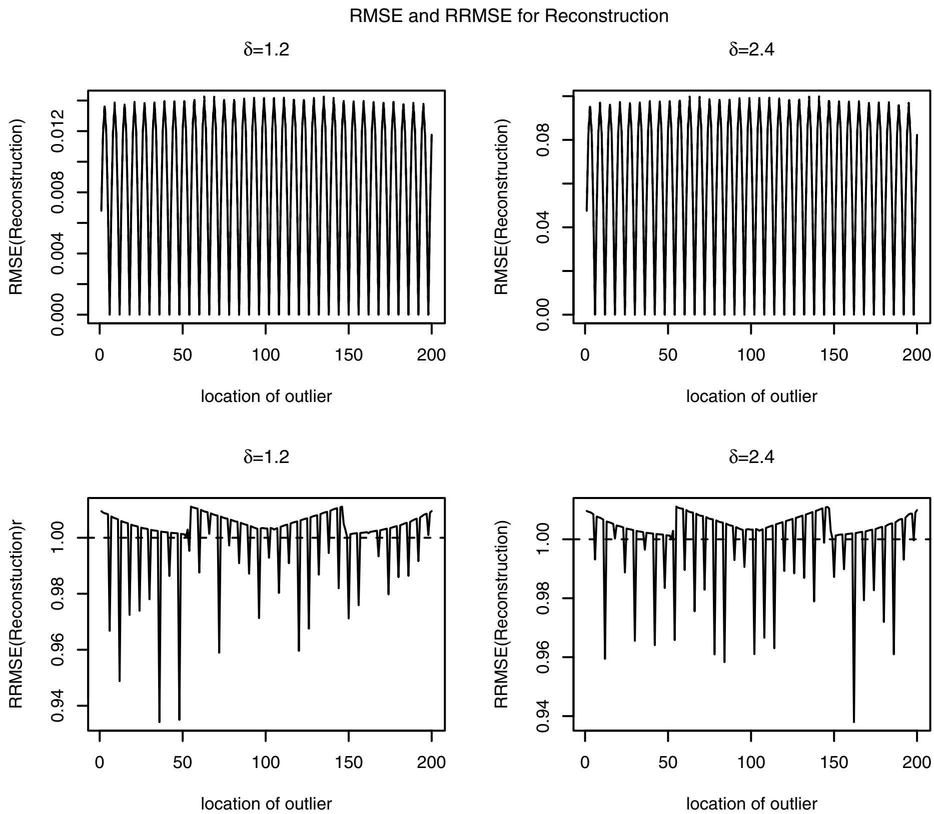

Fig. 1. Plots of RMSE and RRMSE for reconstruction (Example 3.1).

the SSA technique can be applied in forecasting any time series that approximately satisfies the linear recurrent formulae (LRF):

$$y_{i+d} = \sum_{k=1}^{d} \alpha_k y_{i+d-k}, \quad 1 \leq i \leq N - d, \tag{14}$$

of some dimension d with the coefficients $\alpha_1, \ldots, \alpha_d$. It should be noted that d=rank$\mathbf{X}$=rank $\mathbf{X}\mathbf{X}^T$ (the order of the SVD decomposition of the trajectory matrix $\mathbf{X}$). An important property of the SSA decomposition is that, if the original time series Y_N satisfies the LRF (14), then for any N and L there are at most d nonzero singular values in the SVD of the trajectory matrix $\mathbf{X}$; therefore, even if the window length L and $K = N - L + 1$ are larger than d, we only need at most d matrices $\mathbf{X}_i$ to reconstruct the series. Let us now formally describe the algorithm for the SSA forecasting method. The SSA forecasting algorithm, as proposed in Ref. 3, is as follows:

(1) Consider a time series $Y_N = \{y_1, \ldots, y_N\}$ with length N.

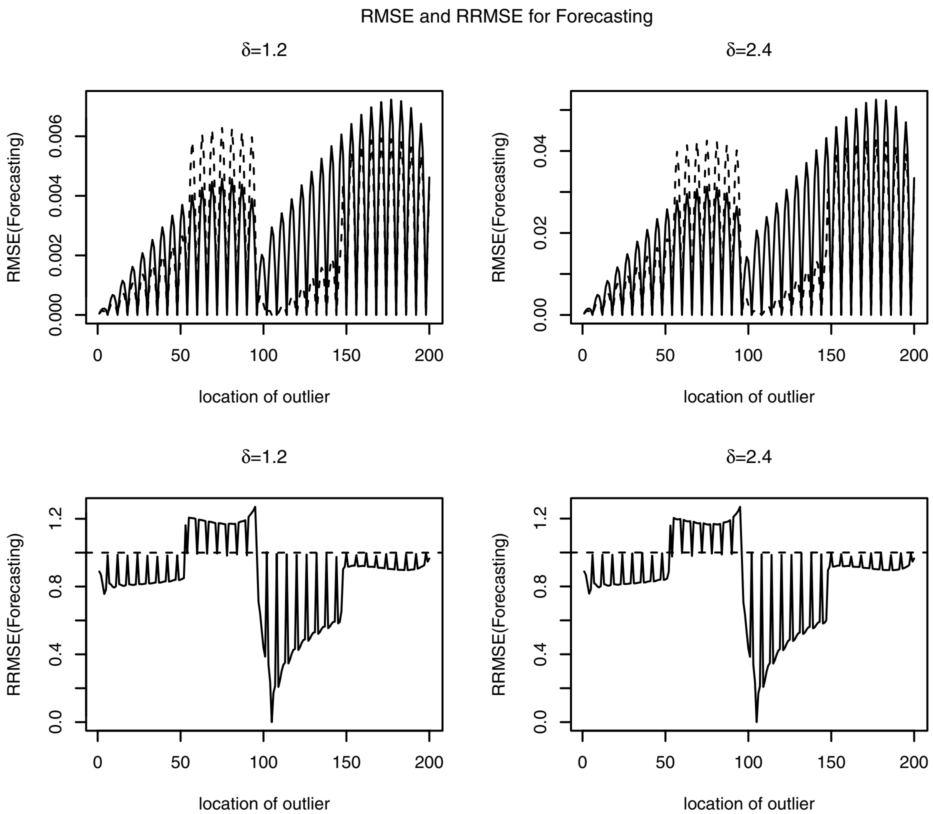

Fig. 2. Plots of RMSE and RRMSE for forecasting (Example 3.1).

(2) Fix the window length L.

(3) Consider the linear space $\mathfrak{L}_r \subset \mathbf{R}^L$ of dimension $r < L$. It is assumed that $e_L \notin \mathfrak{L}_r$, where $e_L = (0, 0, \ldots, 1)^T \in \mathbf{R}^L$.

(4) Construct the trajectory matrix $\mathbf{X} = [X_1, \ldots, X_K]$ of the time series Y_N.

(5) Construct the vectors U_i $(i = 1, \ldots, r)$ from the SVD of $\mathbf{X}$. Note that U_i is an orthonormal basis in $\mathfrak{L}_r$.

(6) Orthogonal projection step; compute matrix $\widehat{\mathbf{X}} = [\widehat{X}_1 : \cdots : \widehat{X}_K] = \sum_{i=1}^{r} U_i U_i^T \mathbf{X}$. The vector $\widehat{X}_i$ is the orthogonal projection of X_i onto the space $\mathfrak{L}_r$.

(7) Construct the matrix $\widetilde{\mathbf{X}} = \mathcal{H}\widehat{\mathbf{X}} = [\widetilde{X}_1 : \cdots : \widetilde{X}_K]$, which is called the Hankelization step.

(8) Set $v^2 = \pi_1^2 + \cdots + \pi_r^2$, where π_i is the last component of the vector U_i $(i = 1, \ldots, r)$. Moreover, assume that $e_L \notin \mathfrak{L}_r$. This implies that $\mathfrak{L}_r$ is not a vertical space. Therefore, $v^2 < 1$.

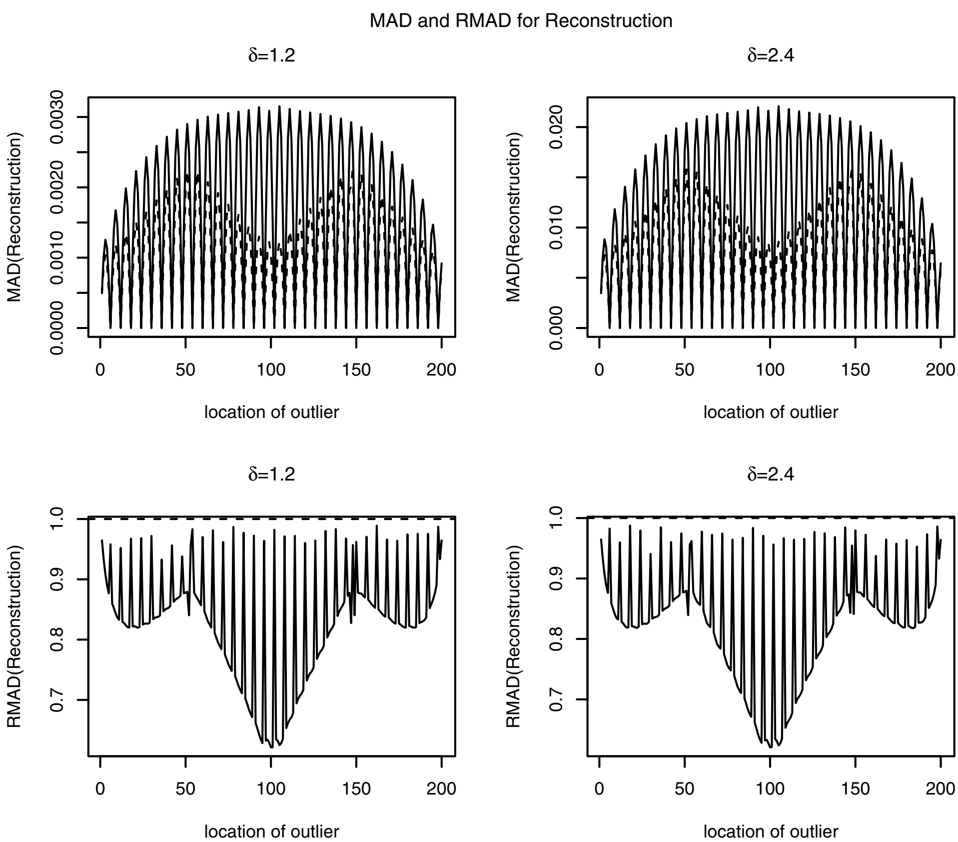

Fig. 3. Plots of MAD and RMAD for reconstruction (Example 3.1).

(9) Determine vector $A = (\alpha_{L-1}, \ldots, \alpha_1)^T$:

$$A = \frac{1}{1 - v^2} \sum_{i=1}^{r} \pi_i U_i^{\nabla},$$

where $U^{\nabla} \in \mathbf{R}^{L-1}$ is the vector consisting of the first $L-1$ components of the vector $U \in \mathbf{R}^L$. It can be proved that the last component y_L of any vector $Y = (y_1, \ldots, y_L)^T \in \mathfrak{L}_r$ is a linear combination of the first y_{L-1} components, i.e.,

$$y_L = \alpha_1 y_{L-1} + \cdots + \alpha_{L-1} y_1,$$

and this does not depend on the choice of a basis $U_1, \ldots, U_r$ in the linear space $\mathfrak{L}_r$.

(10) Define the time series $Y_{N+h} = \{y_1, \ldots, y_{N+h}\}$ by the formula

$$y_i = \begin{cases} \widetilde{y}_i & \text{for } i = 1, \ldots, N, \\ \sum_{j=1}^{L-1} \alpha_j y_{i-j} & \text{for } i = N+1, \ldots, N+h, \end{cases} \tag{15}$$

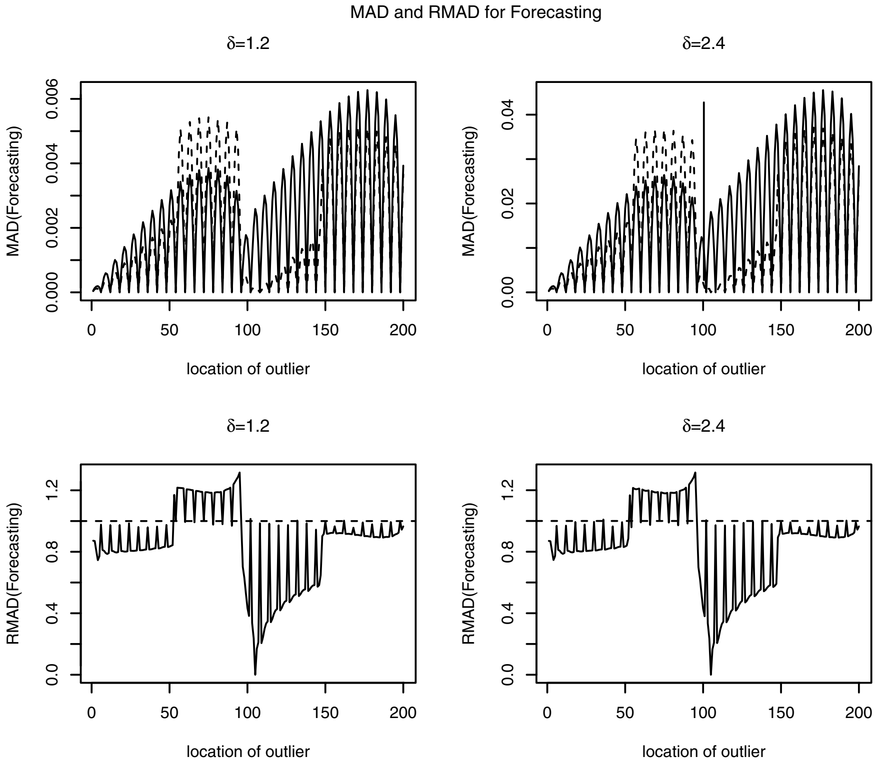

Fig. 4. Plots of MAD and RMAD for forecasting (Example 3.1).

where $\widetilde{y}_i (i = 1, \ldots, N)$ are the reconstructed series. Then, $y_{N+1}, \ldots, y_{N+h}$ are the h step ahead recurrent forecasts.

3. Empirical Results

In this section, the performance of L_1-SSA and basic SSA in reconstruction and forecasting are evaluated by applying them to various real and simulated time series which are contaminated with outlier(s). To measure the accuracy of forecasting results, we use the commonly adopted forecasting performance evaluation measures of root mean squared error (RMSE) and mean absolute deviation (MAD). These criterions have been extensively used in time series analysis and forecasting literature, see for example [2, 14, 28]. Moreover, the effect of an outlier on the forecasting and reconstruction for two procedures are assessed. The following ratios are used for comparing the established and newly proposed SSA methods:

$$\mathrm{RRMSE}^{(i)} = \frac{\mathrm{RMSE}^{(i)} \text{ based on } L_1\text{-SSA}}{\mathrm{RMSE}^{(i)} \text{ based on basic SSA}},$$

$$\mathrm{RMAD}^{(i)} = \frac{\mathrm{MAD}^{(i)} \text{ based on } L_1\text{-SSA}}{\mathrm{MAD}^{(i)} \text{ based on basic SSA}},$$

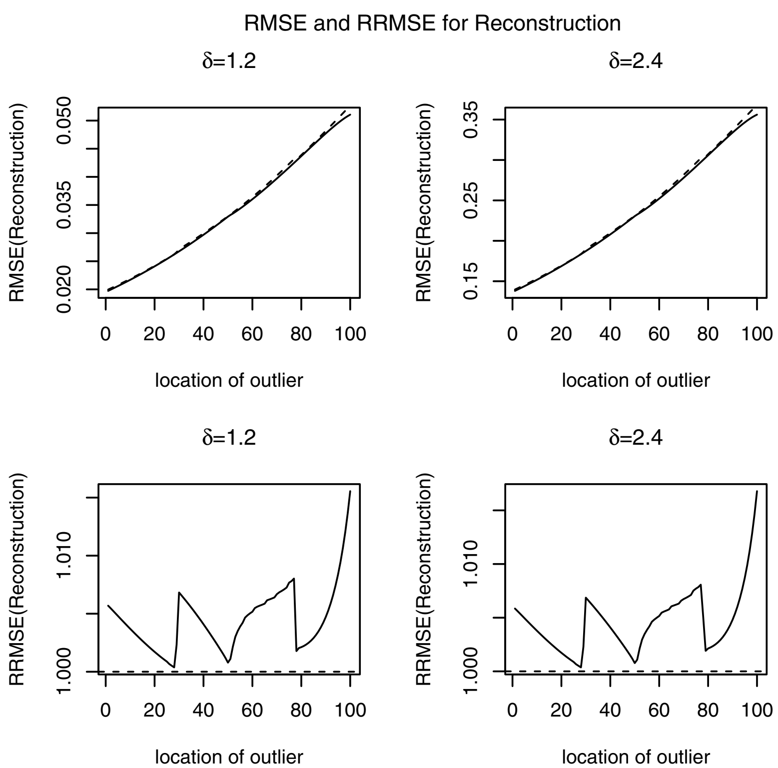

Fig. 5. Plots of RMSE and RRMSE for reconstruction (Example 3.2).

where $\text{RMSE}^{(i)}$ and $\text{MAD}^{(i)}$ denote the value of RMSE and MAD respectively, after replacing the ith observation with an outlier.

If $\text{RRMSE}^{(i)} < 1$ and $\text{RMAD}^{(i)} < 1$, then the L_1-SSA procedure outperforms basic SSA. Alternatively, when $\text{RRMSE}^{(i)} > 1$ and $\text{RMAD}^{(i)} > 1$, it would indicate that the performance of L_1-SSA procedure is worse than basic SSA.

Note that in all following figures of RMSE and MAD, the solid and dash lines correspond to basic SSA and L_1-SSA, respectively. For better comparison, the dashed horizontal line $y = 1$ is added to all figures of RRMSE and RMAD.

Let $Y_N^{(i)} = \{y_1, \ldots, \delta y_i, \ldots, y_N\}$ be the time series with length N that all elements are equal to the original time series $Y_N = \{y_1, \ldots, y_N\}$ except its ith observation which is δ times of the corresponding element of Y_N. It is clear that i is the location of an outlier. In all examples, $\delta = 1.2$ and 2.4 has been selected which is an accepted selection based on the results in Ref. 14. For reconstruction and forecasting of simulated series, the number of eigenvalues (r) has been selected according to the rank of the corresponding trajectory matrix.

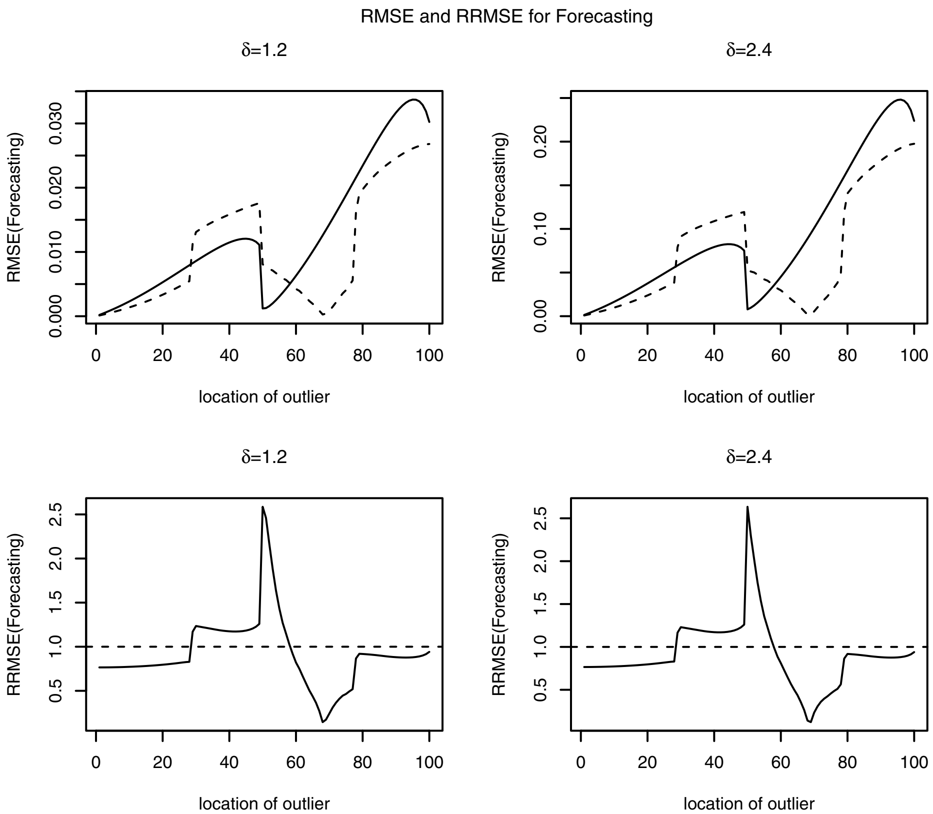

Fig. 6. Plots of RMSE and RRMSE for forecasting (Example 3.2).

3.1. *Simulated series*

Example 3.1. As the first example, we consider the simple Sine series:

$$y_t = \sin(2\pi t/12), \quad t = 1, 2, \ldots, 300.$$

The first 200 observations were considered as in-sample (reconstruction) and the rest as out-of-sample series (forecasting). Since the rank of trajectory matrix for this model is equal to 2 we choose $r = 2$. The window length $L = 96$ is chosen. For more details and useful recommendations about window length, see [25].

As it appears from Figs. 1–4, the general patterns of RMSE (reconstruction and forecasting) are similar for $\delta = 1.2$ and $\delta = 2.4$, and both plots show fluctuations but the magnitude differs in each case. Also, this is true for MAD.

In Fig. 1, the plots of RMSE and RRMSE for reconstruction are depicted. The plots of RRMSE show that sometimes L_1-SSA is better than basic SSA and vice versa (note that for some i, $\mathrm{RRMSE}^{(i)} < 1$).

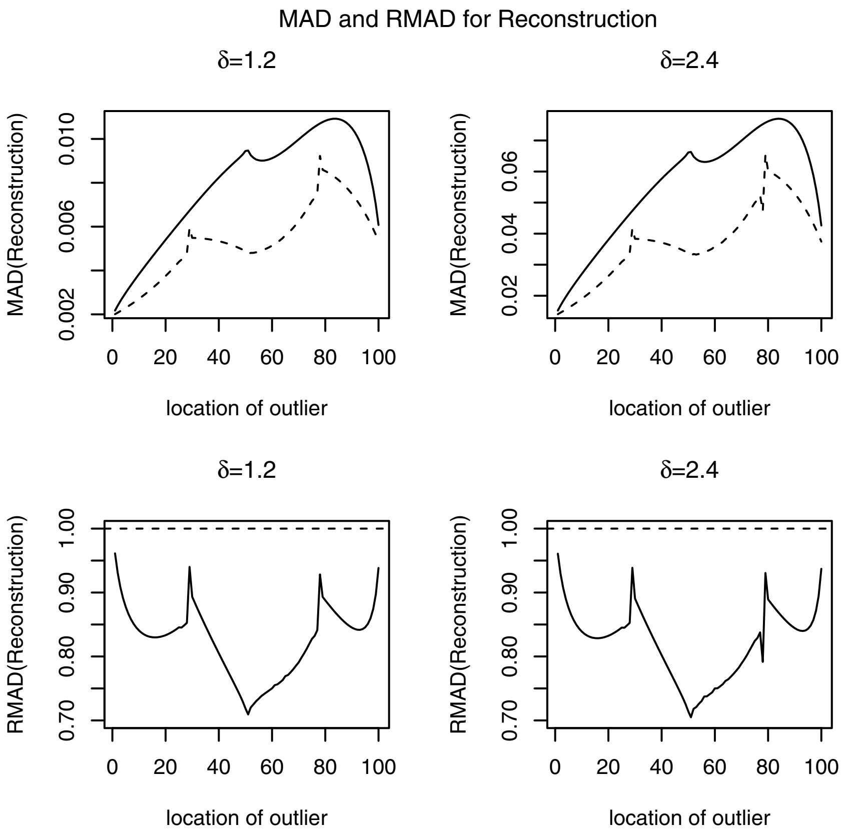

Fig. 7. Plots of MAD and RMAD for reconstruction (Example 3.2).

In Fig. 2, the plots of RMSE and RRMSE for forecasting are drawn. The plots of RRMSE show that almost always L_1-SSA outperforms basic SSA in forecasting, if there is no outlier between times 53 and 95.

Figure 3 depicts the plots of MAD and RMAD for reconstruction. The plots of RMAD show that L_1-SSA outperforms basic SSA at all times.

In Fig. 4, the plots of MAD and RMAD for forecasting are presented. This figure is similar to Fig. 2.

Example 3.2. In this example, we consider the Exponential model:

$$y_t = \exp(\alpha_0 + \alpha_1 t), \quad t = 1, 2, \ldots, 120.$$

The first 100 observations were considered as in-sample (reconstruction) and the rest as out-of-sample series (forecasting). The rank of the trajectory matrix for this model is one, $r = 1$, for different values of α_0 and α_1. For simplicity, we set $\alpha_0 = 0$ and $\alpha_1 = 0.01$. The window length $L = 50$ is chosen.

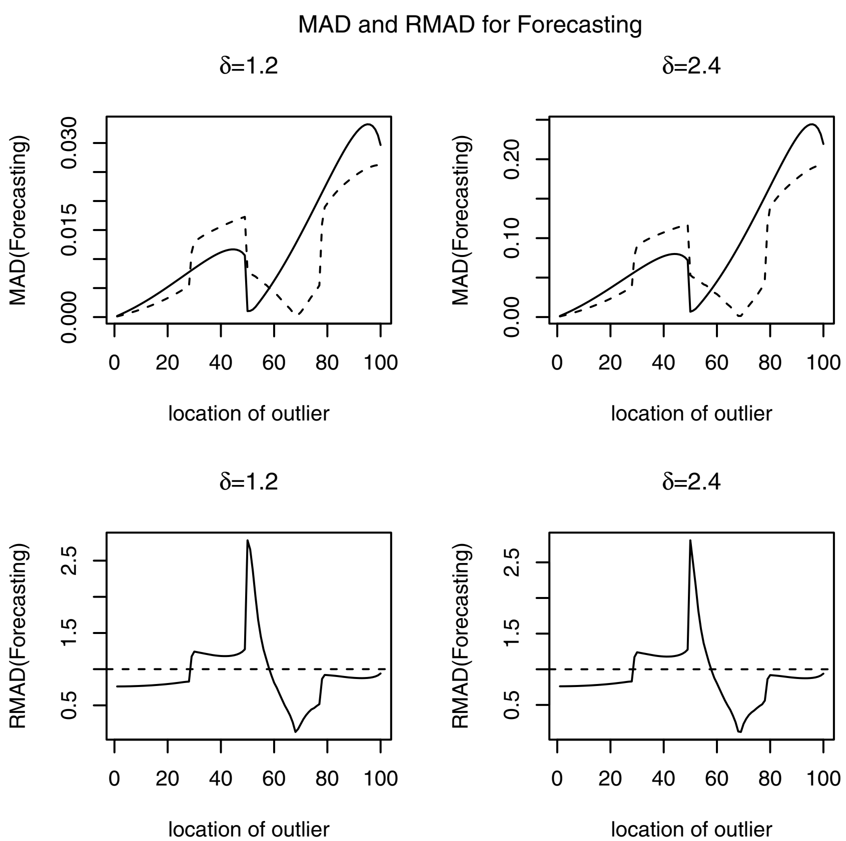

Fig. 8. Plots of MAD and RMAD for forecasting (Example 3.2).

As it appears from Figs. 5–8, the general pattern of RMSE (reconstruction and forecasting) is similar for $\delta = 1.2$ and $\delta = 2.4$ but the magnitude differs in each case. Also, this is true for MAD.

In Fig. 5, the plots of RMSE and RRMSE for reconstruction are depicted. The plots of RMSE show that whenever an outlier is closer to the end of the series, the RMSE increases. From plots of RRMSE, we conclude that basic SSA is better than L_1-SSA.

In Fig. 6, the plots of RMSE and RRMSE for forecasting are shown. It is clear through the RMSE plots that if an outlier is in the middle of the series, the RMSE decreases. The plots of RRMSE show that L_1-SSA outperforms the basic SSA if there is no outlier between times 29 and 58.

In Fig. 7 depicts the plots of MAD and RMAD for reconstruction. It is obvious that L_1-SSA outperforms the basic SSA at all times in this case.

In Fig. 8, the plots of MAD and RMAD for forecasting are drawn. Illustration for this figure is similar to Fig. 6.

Example 3.3. In this example, a simple linear model, $y_t = \alpha_0 + \alpha_1 t, t = 1, 2, \ldots, 300$, with a structural change is considered. The first 200 observations were

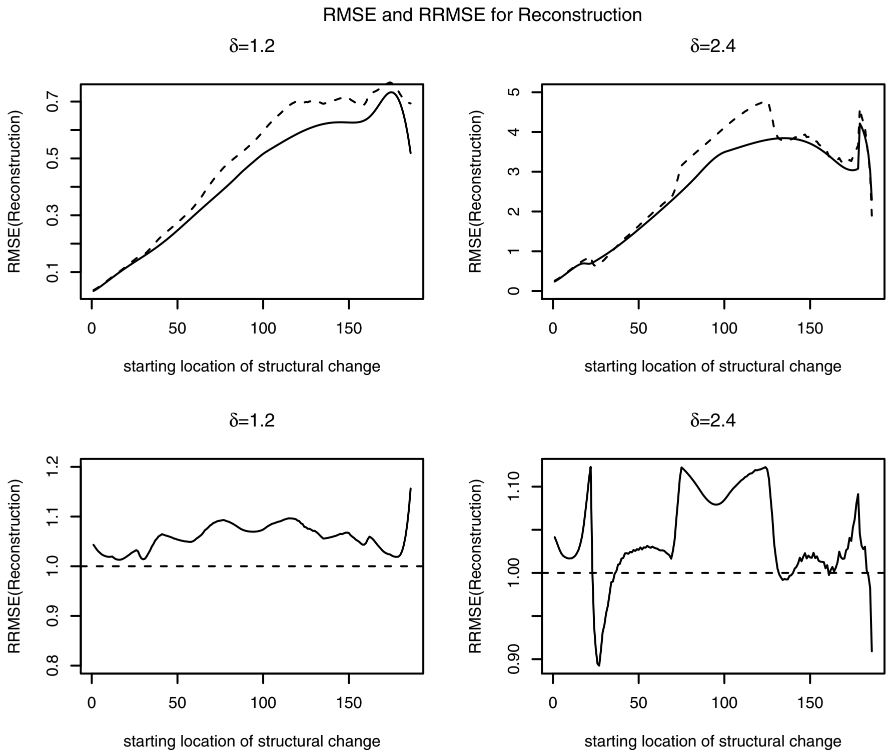

Fig. 9. Plots of RMSE and RRMSE for reconstruction (Example 3.3).

considered as in-sample (reconstruction), and the rest as out-of-sample (forecasting). Since the rank of trajectory matrix for this series is equal to 2 for different values of α_0 and α_1, we choose $r = 2$. Here, $\alpha_0 = 0$, $\alpha_1 = 0.1$ and $L = 100$ are considered. Let $Y_N^{(i)} = \{y_1, \ldots, \delta y_i, \ldots, \delta y_{i+14}, \ldots, y_N\}$ be the time series consisting structural change starting from observation y_i for 15 consequence observations. Thus, all elements of $Y_N^{(i)}$ are equal to the original time series $Y_N = \{y_1, \ldots, y_N\}$ except its 15 observations which are δ times of the corresponding elements of Y_N.

Figure 9 depicts the plots of RMSE and RRMSE for reconstruction stage. Different messages can be concluded for various values of $\delta = 1.2$. In general, there is no obvious preference of using either basic SSA or SSA based on L_1-norm. However, the plots of RRMSE for forecasting indicate that in most cases L_1-SSA outperforms basic SSA, specially if structural change starts from observation 100 onward (see Fig. 10).

Figure 11 depicts plots of MAD and RMAD for reconstruction. the results confirm that L_1-SSA outperforms the basic SSA for all cases. Moreover, Fig. 12 depicts plots of MAD and RMAD for forecasting cases. Similarly, the results confirm that

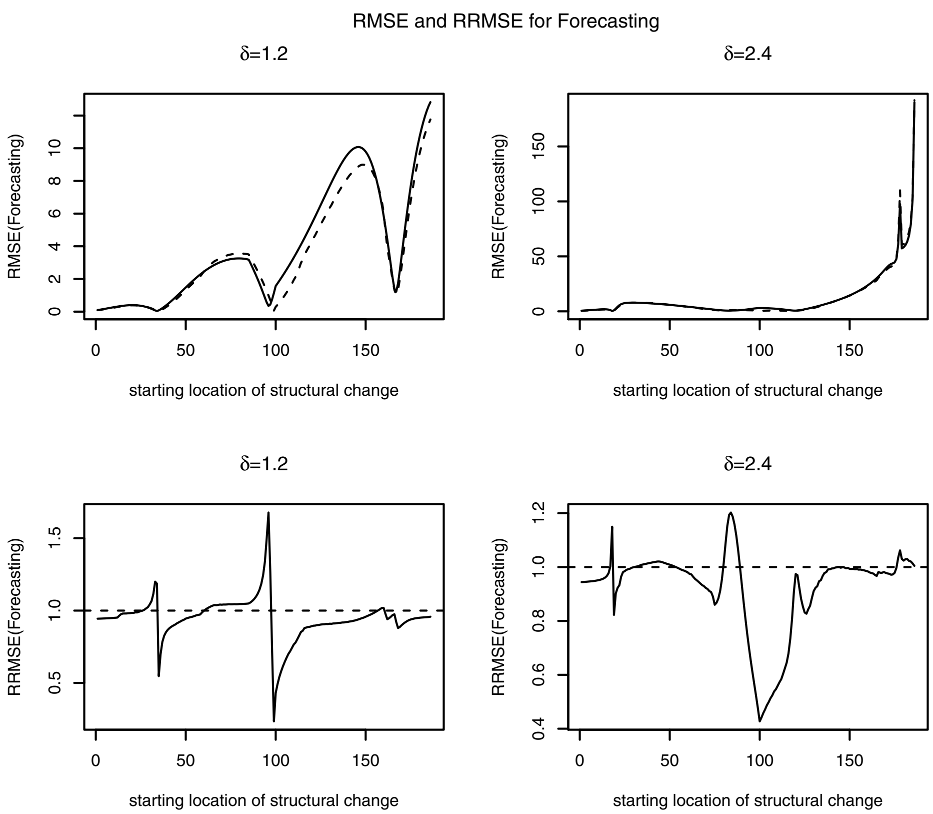

Fig. 10. Plots of RMSE and RRMSE for forecasting (Example 3.3).

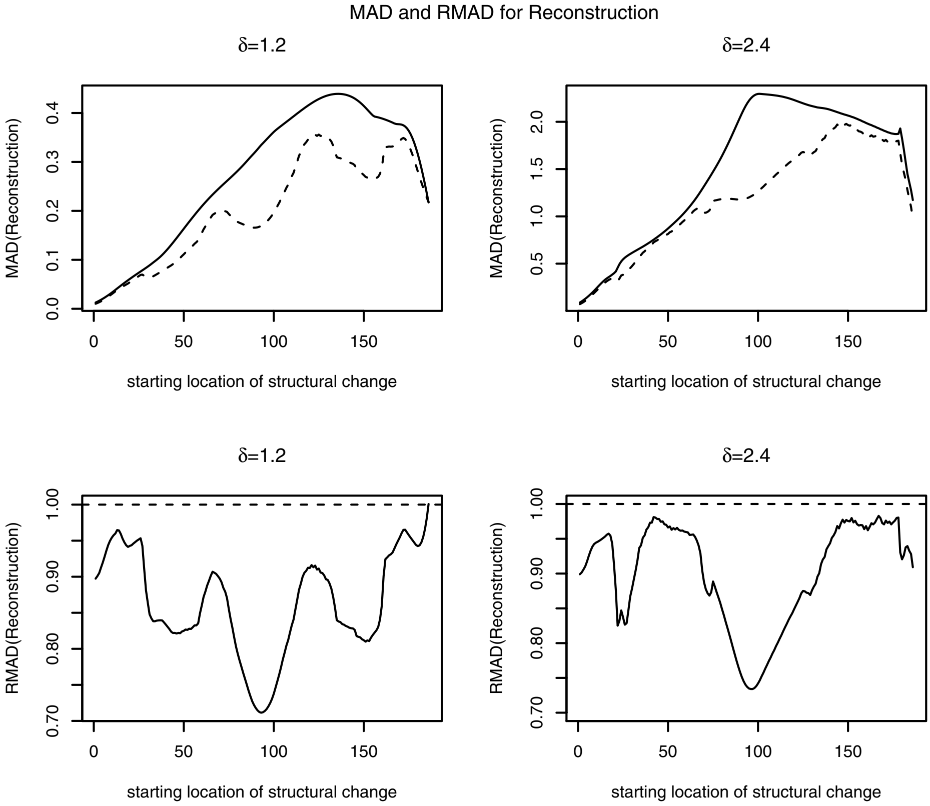

Fig. 11. Plots of MAD and RMAD for reconstruction (Example 3.3).

L_1-SSA outperforms the basic SSA for all cases for forecasting series with structural break.

3.2. *Real data*

Example 3.4. In this example, we consider the well-known time series data set, namely, monthly accidental Deaths in the USA between 1973 and 1978. This data has been used by many authors and can be found in many time series books (see for example [25, 29]) and in every `R` software installation. We use the Death series data that is stored in `R`. Figure 13 shows the time series plot of these data that contain 72 observations.

Similar to simulated series, an outlier has been added into different parts of the U.S. Death series. For the best reconstructing performance, $L = 24$ and $r = 20$ is recommended in Ref. 28 and for the best forecasting performance by using the recurrent method, $L = 24$ and $r = 13$ is offered in Ref. 14.

Brockwell and Davis [29] applied the traditional Box–Jenkins SARIMA models, the ARAR algorithm and the seasonal Holt–Winters algorithm on the Death series

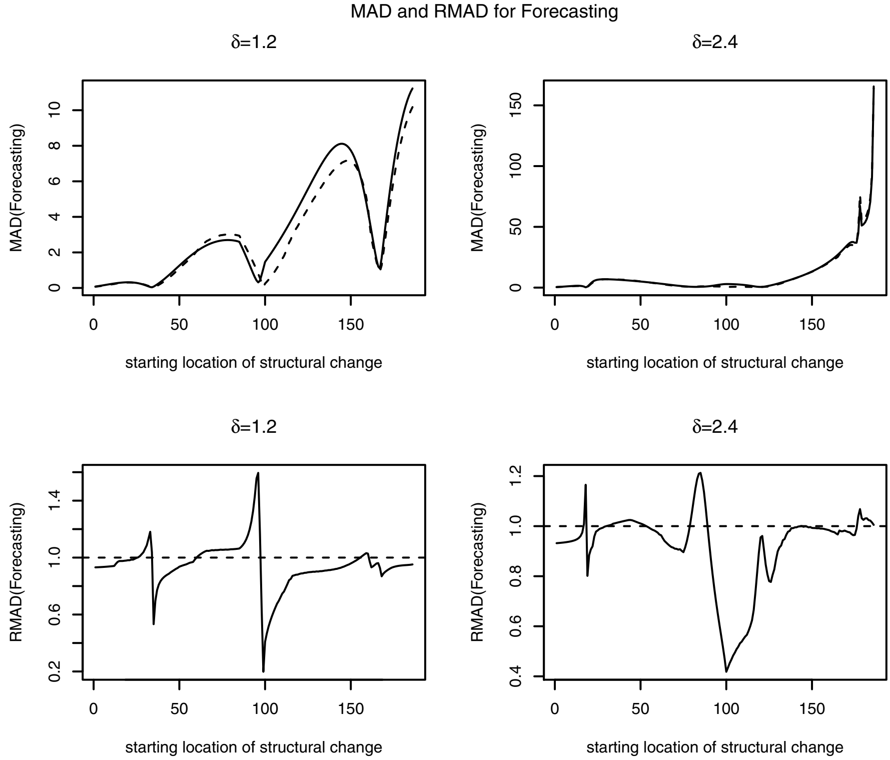

Fig. 12. Plots of MAD and RMAD for forecasting (Example 3.3).

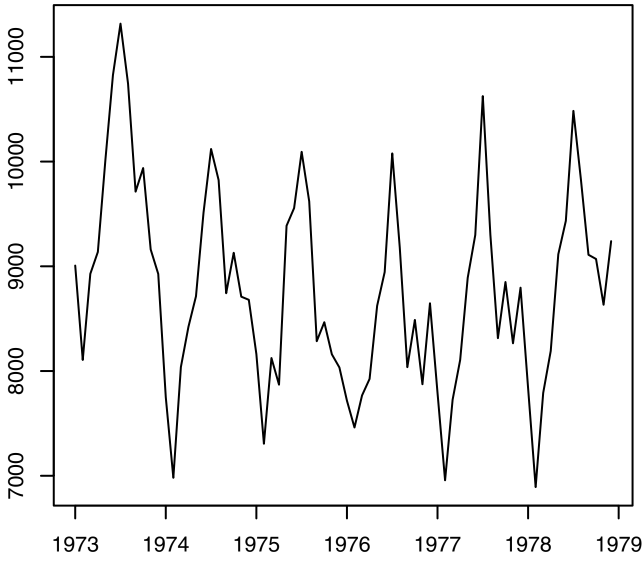

Fig. 13. Time series plot of Death series.

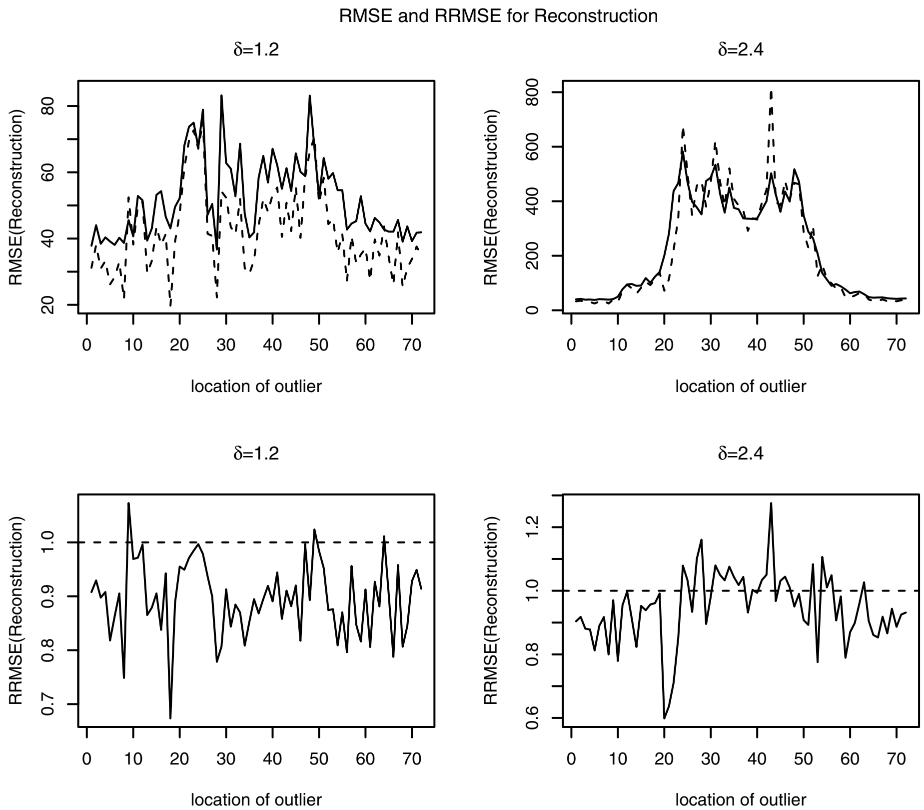

Fig. 14. Plots of RMSE and RRMSE for reconstruction of Death series.

to forecast six future data points. We use those six future data points and calculate RMSE and MAD for forecasting with SSA.

In Fig. 14, the plots of RMSE and RRMSE for reconstruction are drawn. From this figure, we can deduce that almost always L_1-SSA reports a better performance during reconstruction if $\delta = 1.2$.

In Fig. 15, the plots of RMSE and RRMSE for forecasting are presented. The plots of RMSE show that whenever an outlier appears closer to the end of the series, the RMSE increases. Also, from plots of RRMSE one can conclude that almost always L_1-SSA reports a better performance in forecasting.

In Fig. 16, the plots of MAD and RMAD for reconstruction are drawn. It is obvious that L_1-SSA outperforms basic SSA at all times in this case.

In Fig. 17, the plots of MAD and RMAD for forecasting are shown. Illustration for this figure is similar to Fig. 15.

Example 3.5. For comparing the performance of L_1-SSA and basic SSA in reconstruction and forecasting of time series that are polluted with outliers; in this

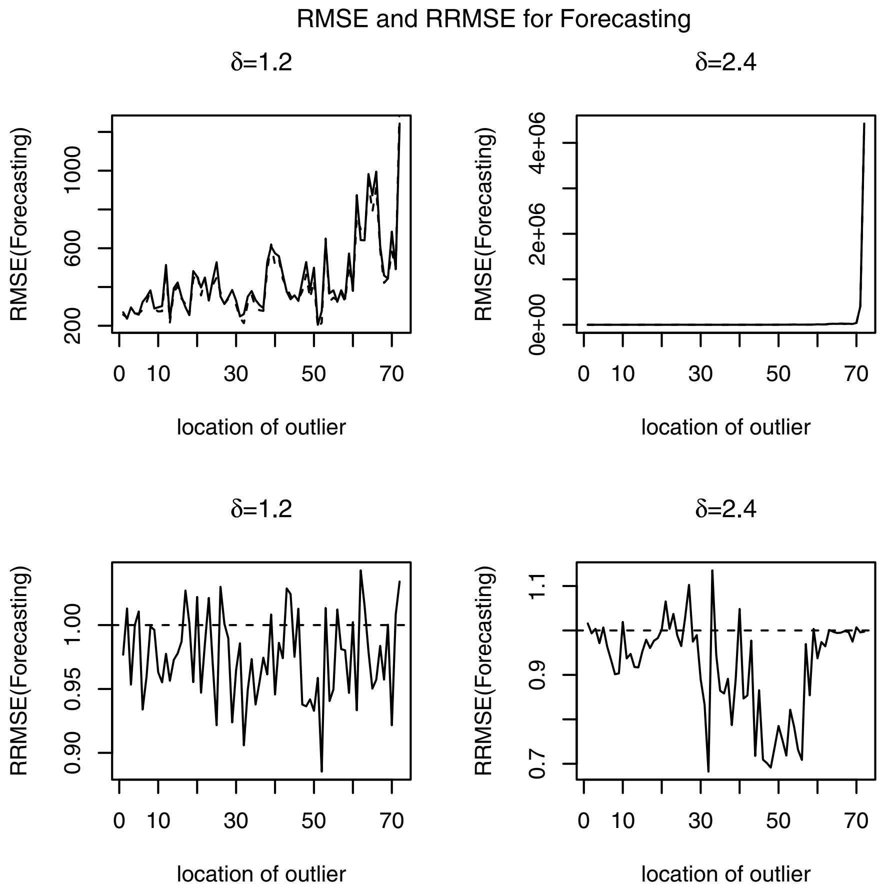

Fig. 15. Plots of RMSE and RRMSE for forecasting of Death series.

example, we consider the U.S. War series [30]. This data represents U.S. combat deaths in the Vietnam War, monthly from 1966 to 1971 and contain 72 observations. Figure 18 shows the time series plot of these data.

From this plot, it seems that there are two outliers in February and May 1968. The first 66 observations were considered as in-sample (reconstruction) and the rest as out-of-sample series (forecasting). The first eight singular values have been used in reconstructing and forecasting the series and other singular values have been considered as noise components.

Figure 19 shows the RMSE, RRMSE, MAD and RMAD of reconstructed series for different values of window length, L. As it appears, the L_1-SSA shows a better performance in reconstruction for every L with respect to MAD measure.

In Fig. 20, the plots of RMSE, RRMSE, MAD and RMAD of the forecasted series for different values of L are depicted. From this figure, we can conclude that almost always L_1-SSA reports a comparatively better forecasting performance.

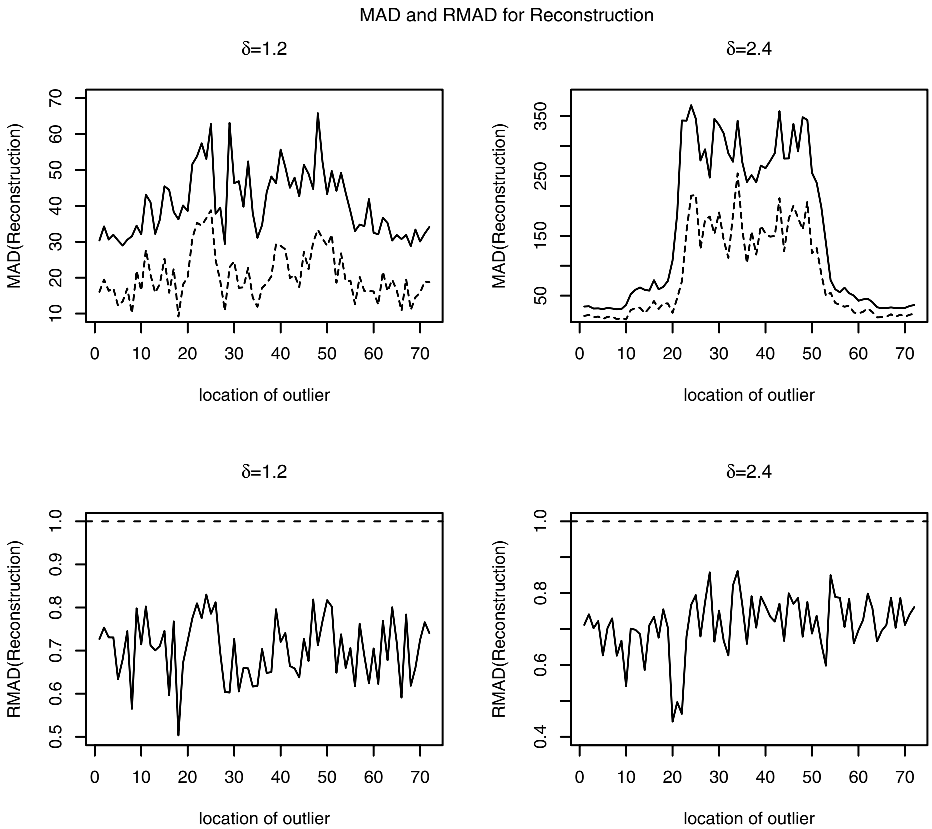

Fig. 16. Plots of MAD and RMAD for reconstruction of Death series.

4. Conclusion

This paper begins with a brief description on SSA, its importance and its applications before moving directly to the issue at hand, i.e., the introduction of SSA based on L_1-norm (L_1-SSA). The challenges imposed by the presence of outliers in time series for modelling and forecasting are discussed and a non-technical explanation is presented to support the idea underlying L_1-SSA. In particular, it was expected that L_1-SSA would enable better reconstruction and forecasting in comparison to basic SSA in the presence of outliers because the L_2-norm is definitely sensitive to outliers whilst L_1-norm is by definition insensitive to the presence of outliers.

Comparisons between L_1-SSA and basic SSA were carried out using both simulated and real world data to demonstrate SSA's capability at analyzing and forecasting time series with outliers. The study considered both RMSE and MAD criterions when arriving at conclusions. In brief, the comparison of forecasting results showed that L_1-SSA is more accurate than the basic SSA version, further confirming the results obtained via theoretical results. The results have also confirmed that the

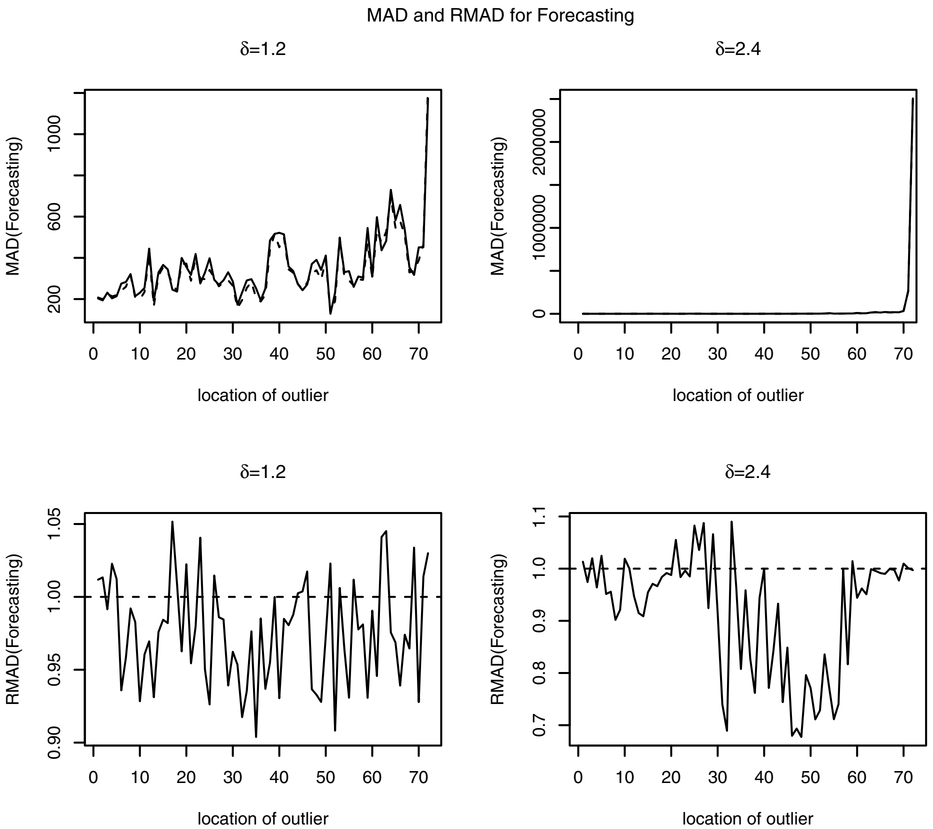

Fig. 17. Plots of MAD and RMAD for forecasting of Death series.

errors obtained via L_1-SSA are stochastically smaller than those obtained by basic SSA in the univariate case.

Applications to the U.S. Death series modified with outliers showed that based on the RMSE and RRMSE criterions, L_1-SSA reports a better reconstruction when $\delta = 1.2$, and better forecasting in comparison to basic SSA where an outlier appears closer to the end of the series. Based on the MAD and RMAD criterions it was evident that L_1-SSA outperforms basic SSA at all times during reconstruction. An application to the U.S. War series showed that when considering the MAD criterion, L_1-SSA outperforms basic SSA by achieving a better reconstruction for every L whilst almost always L_1-SSA was seen providing a better forecast than basic SSA.

It is interesting that the results in Ref. 14 suggested that the identification and removal of outliers could further enhance SSA reconstruction and forecasting capabilities (as SSA based on L_2-norm is more sensitive to outliers) whilst the results in this paper have shown that when faced with time series polluted by

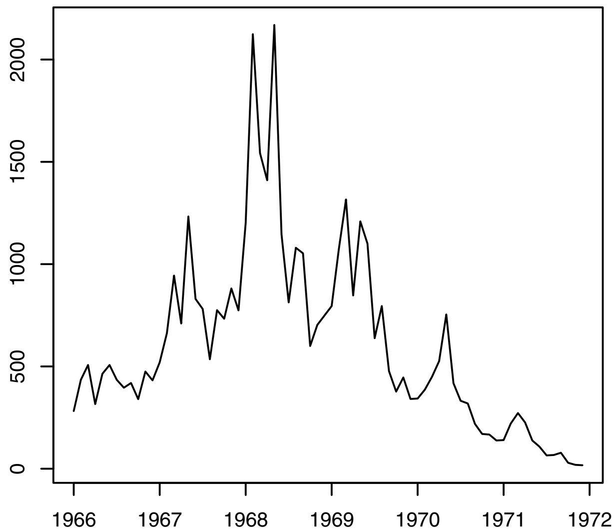

Fig. 18. Time series plot of War series.

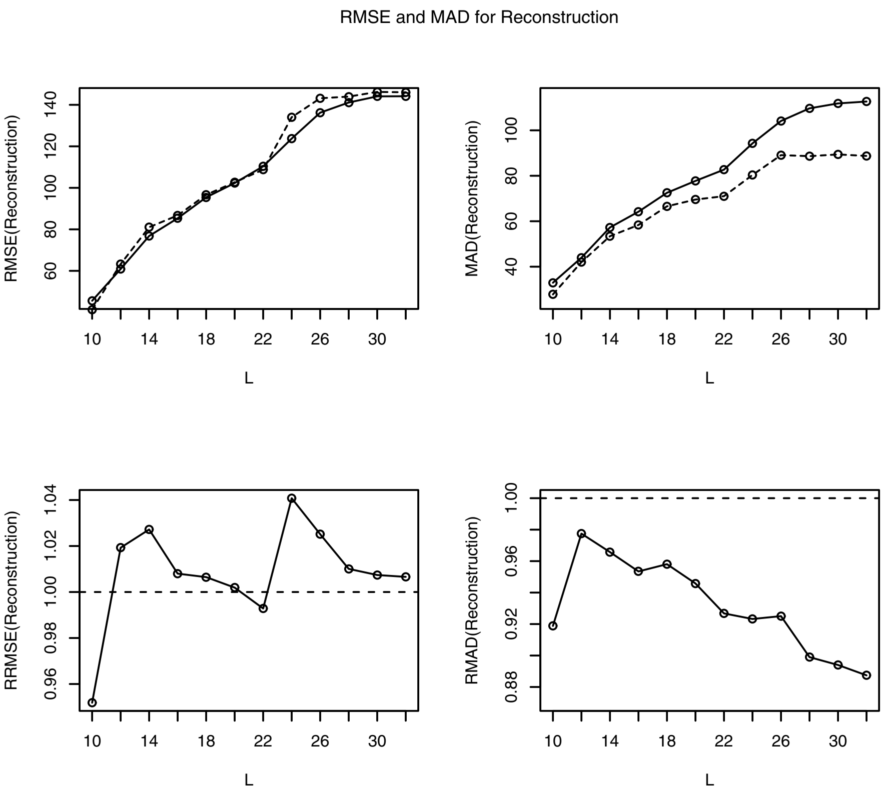

Fig. 19. Plots of RMSE, RRMSE, MAD and RMAD for reconstruction of War series.

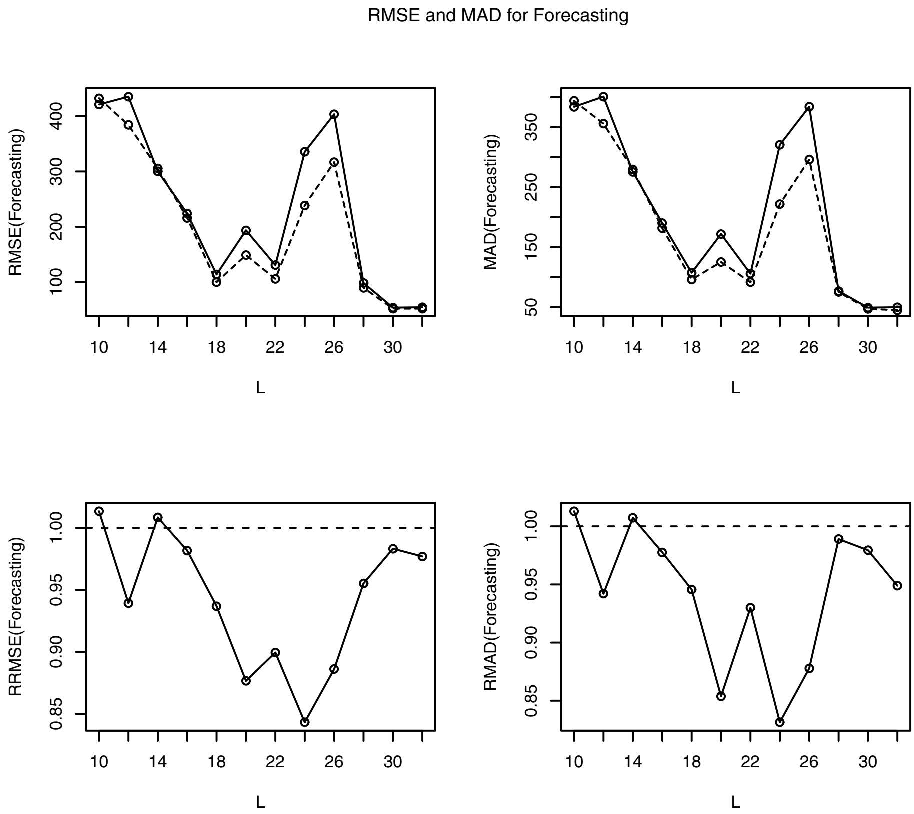

Fig. 20. Plots of RMSE, RRMSE, MAD and RMAD for forecasting of War series.

outliers practitioners can obtain better results by switching to L_1-SSA. This in turn means there is no requirement for data transformations which will ensure no information is lost. In terms of future research, it would be worthwhile to ascertain the performance of L_1-SSA when faced with different distributions and different noise levels in order to get a better understanding with regard to the robustness of this newly proposed approach for modeling and forecasting with SSA.

References

[1] N. Antonakakis and G. Filis, Oil prices and stock market correlation: A time-varying approach, *Int. J. Energy Stat.* **1**(1) (2013) 17–29.

[2] H. Hassani, A. Webster, E. S. Silva and S. Heravi, Forecasting U.S. Tourist arrivals using optimal singular spectrum analysis, *Tourism Manage.* **46** (2015) 322–335.

[3] N. Golyandina, V. Nekrutkin and A. Zhigljavsky, *Analysis of Time Series Structure: SSA and Related Techniques* (Chapman & Hall/CRC, NY, London, 2001).

[4] Z. Hou, G. Wen, P. Tang and G. Cheng, Periodicity of carbon element distribution along casting direction in continuous-casting billet by using singular spectrum analysis, *Metall. Mater. Trans. B* **45**(5) (2014) 1817–1826.

[5] K. Liu, S. S. Law, Y. Xia and X. Q. Zhu, Singular spectrum analysis for enhancing the sensitivity in structural damage detection, *J. Sound Vibration* **333**(2) (2014) 392–417.

[6] K. L. Bail, J. M. Gipson and D. S. MacMillan, Quantifying the correlation between the MEI and LOD variations by decomposing LOD with singular spectrum analysis, *Earth on the Edge: Science for a Sustainable Planet*, International Association of Geodesy Symposia, Springer, Vol. 139, pp. 473–477.

[7] Q. Chen, T. V. Dam, N. Sneeuw, X. Collilieux, M. Weigelt and P. Rebischung, Singular spectrum analysis for modeling seasonal signals from GPS time series, *J. Geodyn.* **72** (2013) 25–35.

[8] B. Muruganatham, M. A. Sanjith, B. Krishnakumar and S. A. V. Satya Murty, Roller element bearing fault diagnosis using singular spectrum analysis, *Mech. Syst. Signal Process.* **35**(1–2) (2013) 150–166.

[9] H.-S. Chao and C.-H. Loh, Application of singular spectrum analysis to structural monitoring and damage diagnosis of bridges, *Struct. Infrastruct. Eng. Maint. Manage. Life-Cycle Des. Perform.* **10**(6) (2014) 708–727.

[10] S. Aydin, H. M. Saraoglu and S. Kara, Singular spectrum analysis of sleep EEG in insomnia, *J. Med. Syst.* **35**(4) (2011) 457–461.

[11] H. Hassani, Z. Xu and A. Zhigljavsky, Singular spectrum analysis based on the perturbation theory, *Nonlinear Analysis: Real World Applications* **12** (2011) 2752–2766.

[12] Sanei, Saeid and Hossein Hassani. Singular spectrum analysis of biomedical signals. CRC Press, 2015.

[13] M. Ghodsi and M. Yarmohammadi, Exchange rate forecasting with optimum singular spectrum analysis. *J. Syst. Sci. Complex.* **27**(1) (2014) 47–55.

[14] H. Hassani, R. Mahmoudvand, H. N. Omer and E. S. Silva, A preliminary investigation into the effect of outlier(s) on singular spectrum analysis, *Fluct. Noise Lett.* **13**(4) (2014) 1450029.

[15] H. Hassani, R. Mahmoudvand, M. Zokaei and M. Ghodsi, On the separability between signal and noise in singular spectrum analysis, *Fluct. Noise Lett.* **11**(2) (2012) 1250014.

[16] H. Hassani, R. Mahmoudvand and M. Yarmohammadi, Filtering and denoising in the linear regression model, *Fluct. Noise Lett.* **9**(4) (2010) 343-358.

[17] R. S. Tsay, Outliers, level shifts, and variance changes in time series, *J. Forecast.* **7**(1) (1988) 1–20.

[18] D. Pena, *Outliers*, Influential observations and missing data, in *A Course in Time Series*, eds. D. Pena, G. C. Tiao and R. S. Tsay (John Wiley, New York, 2001).

[19] J. Tolvi, Outliers in time series: A review. University of Turku, Department of Economics, Research Reports No. 76 (1998), pp. 1–30.

[20] S. J. Deutsch, J. E. Richards and J. J. Swain, Effects of a single outlier on arma identification, *Commun. Stat. Theory Methods* **19**(6) (1990) 2207–2227.

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

[22] S. P. Ellis and S. Morgenthaler, Leverage and Breakdown in L1 Regression, *J. Am. Stat. Assoc.* **87**(417) (1992) 143–148.

[23] R. Rajko, Treatment of model error in calibration by robust and fuzzy procedures, *Anal. Lett.* **27**(1) (1994) 215–228.

[24] K. de Jacques. Automated outlier detection in singular spectrum analysis, www.statistics.gov.hk/wsc/CPS204-P35-S.pdf.

[25] H. Hassani, Singular spectrum analysis: Methodology and comparison, *J. Data Sci.* **5**(2) (2007) 239–257.

[26] D. Birkes and Y. Dodge, *Alternative Methods of Regression* (Wiley, New York, 1993).
[27] P. Bloomfield and W. Steiger, *Least Absolute Deviations: Theory, Applications and Algorithms* (Birkhäuser, Boston, 1983).
[28] H. Hassani and R. Mahmoudvand, Multivariate singular spectrum analysis: A general view and new vector forecasting approach, *Int. J. Energy Stat.* **1**(1) (2013) 55–83.
[29] P. J. Brockwell and R. A. Davis, *Introduction to Time Series and Forecasting*, 2nd edn. (Springer, New York, 2002).
[30] M. Janowitz and B. Schweizer, Ordinal and percentile clustring, *Math. Soc. Sci.* **18** (1989) 135–186.