A robust approach for outlier imputation: Singular Spectrum Decomposition

Maryam Movahedifar*, Hossein Hassani†,
Masoud Yarmohammadi*, Mahdi Kalantari* and Rangan Gupta‡

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

†Institute for International Energy Studies
Tehran 1967743 711, Iran

‡University of Pretoria, Department of Economics, South Africa

†hassani.stat@gmail.com

February 23, 2021

Abstract

Singular Spectrum Analysis (SSA) is a nonparametric method for separating time series data into a sum of small numbers of interpretable components (signal+noise). One of the steps of the SSA method, which is referenced to *Embedding*, is extremely sensitive to contamination of outliers which are often founded in time series analysis. To reduce the effect of outliers, SSA based on Singular Spectrum Decomposition (SSD) method is proposed. In this paper the ability of SSA based on SSD and basic SSA are compared in time series reconstruction in the presence of outliers. A wide empirical study on both simulated and real data verifies the efficiency of basic SSA based on SSD in reconstructing the time series where polluted by outliers.

Keywords: Outlier; trajectory matrix; signal extraction; singular spectrum analysis and singular spectrum decomposition.

1 Introduction

Outliers in a real data set could be occurred by various reasons such as implementation of new regulation, major changes in political or economic policy, or occurrence of a disaster. Outliers may be regarded as an observation, which does not come from the target population. Detecting outliers is important because they have an impact on the selection of the model, the estimation of parameters and, consequently, on forecasts. Therefore, the presence of outliers may lead to wrong conclusions in the analysis of time series such as biased parameter estimation, poor forecasts and inappropriate decomposition of the series. In addition, the presence of outliers and shocks could make a time series non-stationary (see [1]). Furthermore, the parametric methods are more sensitive to the presence of outliers than non-parametric methods. For instance, outliers have negative impacts on the structure of autocorrelation in time series and that ARMA models are effected through the bias caused in autocorrelation function (ACF) and partial autocorrelation function (PACF)(see [2]). The impact of contaminated data in estimating the parameters of ARIMA models have been studied in [3–5]. In addition, the presence of contaminated data and shocks could easily mislead time series analysis, forecasting results and regression analysis (see, for example, [6, 7]). There is an extensive literature covering the impacts of outliers on parametric models, and interested readers are referred to [8, 9] for example.

In this paper, *Additive Outliers (AO)* is considered(see for more details, [10] and [11]). This type of outlier affect a time series in a variety of ways. For linear models, additive outliers, which influence a single observation were introduced in [10]. The author in [12] also showed how linear models could be applied for detection of AO in stationary time series data. In terms of the impact of outliers on SSA, in Ref. [1] it was deduced that outliers have a considerable impact on SSA reconstruction and forecasts. Furthermore, basic SSA and robust SSA approaches have been compared together in the presence of noisy/contaminated time series in Ref [11].

The aim of this paper is to compare the ability of basic SSA, L_1-SSA, SSD and L_1-SSD methods in the presence of additive outliers. In order to do this, in all the synthetic dataset and real data, *Additive Outliers (AO)* is considered as follows:
Additive Outliers (AO):
Each time series observation y_i is individually selected and replaced by $3 \times y_i$ and $5 \times y_i$ for two first simulated data series and by $15 \times y_i$ and $20 \times y_i$ for the second two simulated time series, that is., the magnitude of y_i is changed by a factor of $3, 5, 15$ and 20, respectively.
The rest of this paper is organized as follows. Sec. 2 presents a short theoretical background of basic SSA. Sec. 3 is devoted to introduce the new SSD method. Sec. 4 covers some numerical examples including simulation studies. In Sec. 5, real data analysis have been reported. Finally, Sec. 6 presents a summary of the study and some concluding remarks.

2 Singular Spectrum Analysis

The main purpose of SSA is to decompose the original time series into a sum of series, so that each component in this sum can be identified as either a trend, periodic or quasi-periodic (perhaps, amplitude-modulated), or noise. This is followed by a reconstruction of the original series. The most common version of SSA is known as Basic SSA. It is noteworthy that the matrix norm used in Basic SSA is the *Frobenius* norm or L_2-norm. There is a newer version of SSA that is based on L_1-norm and called L_1-SSA. It was confirmed that L_1-SSA is robust against outliers [13]. In the following subsections, a concise description of the methodology of Basic SSA and L_1-SSA are presented. For more details on Basic SSA, see [14–16]. More detailed information on L_1-SSA can be found in [13].

2.1 A Brief Description of Basic SSA

Let $Y_N = (y_1, ..., y_N)$ be a time series of length N. Fix an integer L, which is called window length, such that $2 \leq L \leq N/2$. Basic SSA consists of four steps as follows:

Step 1) Embedding: This step transfers one-dimensional time series $Y_N = (y_1, ..., y_N)$ of length N into the multi-dimensional series $X_1, \ldots, X_K$ with vectors $X_i = (y_i, \ldots, y_{i+L-1}) \in \mathbb{R}^L$, where $K = N - L + 1$. Vectors X_i are called L-lagged vectors (or, simply, lagged vectors). The single parameter of the embedding is the window length L. The result of this step is the trajectory matrix $\mathbf{X} = [X_1, \ldots, X_K]$.

Step 2) Singular Value Decomposition (SVD): In this step, the trajectory matrix $\mathbf{X}$ is decomposed into a sum of rank-one elementary matrices. The eigenvalues of $\mathbf{X}\mathbf{X}^T$ are denoted by $\lambda_1, \ldots, \lambda_L$ in decreasing order of magnitude ($\lambda_1 \geq \cdots \geq \lambda_L \geq 0$) and by $U_1, \ldots, U_L$, the eigenvectors of the matrix $\mathbf{X}\mathbf{X}^T$ corresponding to these eigenvalues. If $d = \max\{i, \text{such that } \lambda_i > 0\} = rank(\mathbf{X})$ then the SVD of the trajectory matrix can be written as $\mathbf{X} = \mathbf{X}_1 + \cdots + \mathbf{X}_d$, where $\mathbf{X}_i = \sqrt{\lambda_i} U_i V_i^T$ and $V_i = \mathbf{X}^T U_i / \sqrt{\lambda_i}$ $(i = 1, \ldots, d)$.

Step 3) Grouping: In this step, the set of indices $\{1, \ldots, d\}$ are partitioned into m disjoint subsets $I_1, \ldots, I_m$. Let $I = \{i_1, \ldots, i_p\}$. Then the matrix $\mathbf{X}_I$ corresponding to the group I is defined as $\mathbf{X}_I = \mathbf{X}_{i_1} + \cdots + \mathbf{X}_{i_p}$. For example, if $I = \{2, 3, 5\}$ then $\mathbf{X}_I = \mathbf{X}_2 + \mathbf{X}_3 + \mathbf{X}_5$. Having the SVD of $\mathbf{X}$, the split of the set of indices $\{1, \ldots, d\}$ into the disjoint subsets $I_1, \ldots, I_m$ corresponds to this representation: $\mathbf{X} = \mathbf{X}_{I_1} + \cdots + \mathbf{X}_{I_m}$.

Step 4) Diagonal Averaging: In this step, we seek to transform each matrix $\mathbf{X}_{I_j}$ of the grouping step into a Hankel matrix so that these can subsequently be converted into a time series. It is well known that in basic SSA, Hankelization is obtained via diagonal averaging of the matrix elements over the anti-diagonals [14].

2.2 A Brief Description of L_1-SSA

Similar to basic SSA, the L_1-SSA consists of four steps. However, except embedding and grouping steps, there are fundamental differences between basic SSA and L_1-SSA at the second and fourth steps. In the following, the four steps of L_1-SSA are outlined.

Step 1) Embedding: This step is similar to the embedding step of basic SSA and therefore we do not reproduce it here.

Step 2) L_1-Decomposition: In this step, we perform L_1-Decomposition of the trajectory matrix $\mathbf{X}$. The L_1-Decomposition of $\mathbf{X}$ can be written as $\mathbf{X} = \mathbf{X}_1 + \cdots + \mathbf{X}_d$, where $\mathbf{X}_i = \sqrt{\lambda_i} w_i U_i V_i^T$ and λ_i, U_i, V_i are those defined in the SVD step of basic SSA $(i = 1, \ldots, d)$. The matrices $\mathbf{X}_i$ have rank 1. The

w_i is the weight of *singular value* $\sqrt{\lambda_i}$. These weights are diagonal elements of diagonal weight matrix $\mathbf{W} = diag(\underbrace{w_1, w_2, \ldots, w_d}_{d}, \underbrace{0, 0, \ldots, 0}_{L-d})$ and are computed such that $\left\|\mathbf{X} - \mathbf{UW\Sigma V}^T\right\|_{L_1}$ be minimized; where $\mathbf{U} = [U_1 : \cdots : U_L]$, $\mathbf{V} = [V_1 : \cdots : V_L]$, $\mathbf{\Sigma} = diag(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \ldots, \sqrt{\lambda_L})$ and $\|.\|_{L_1}$ is the L_1 norm of a matrix. For more information, see [13].

Step 3) Grouping: The grouping step is once again identical to that of basic SSA and is therefore not reproduced here.

Step 4) L_1-Hankelization: Let $\mathcal{H}\mathbf{A}$ be the result of the Hankelization of matrix $\mathbf{A}$. In L_1-SSA, Hankelization corresponds to computing the median of the matrix elements over the "antidiagonal". This type of Hankelization is an optimal procedure in the sense that the matrix $\mathcal{H}\mathbf{A}$ is closest to $\mathbf{A}$ (with respect to the L_1 norm) among all Hankel matrices of the same dimension , that is, $\|\mathbf{A} - \mathcal{H}\mathbf{A}\|_{L_1}$ is minimum [13].

3 Singular Spectrum Decomposition

Here we consider another method for reconstructing time series called Singular Spectrum Decomposition (SSD). The major difference between basic SSA and SSD is that in the SSD method the choice of the embedding dimension and the selection of the eigentriples for reconstructing a specific component series has been made fully data-driven. Such as basic SSA, SSD method consists of four steps as follows:

Step 1) Embedding: In the first embedding step, for a given time series Y_N of length N, $Y_N = (y_1, ..., y_N)$, and an embedding dimension L, an $L \times N$ matrix $\mathbf{X}_{SSD}$, is generated such that its i-th row is obtained as: $X_i = (y_i, \ldots, y_N, y_1, \ldots, y_{i-1})^T$, with $i = 1, \ldots, N$, hence $\mathbf{X}_{SSD} = [X_1, X_2, \ldots, X_L]^T = [\mathbf{X}|\mathbf{A}]$, where the matrix $\mathbf{X}$ is the trajectory matrix of Y_N based on basic SSA and matrix $\mathbf{A}$ is as follows:

$$\mathbf{A} = \begin{pmatrix} y_{K+1} & \cdots & y_N \\ y_{K+2} & \cdots & y_1 \\ \vdots & \ddots & \vdots \\ y_1 & \cdots & y_{L-1} \end{pmatrix} \tag{1}$$

The matrix $\mathbf{A}$ is additional matrix which makes trajectory matrix $\mathbf{X}_{SSD}$ possible to use all the initial series Y_N in all matrix rows of $\mathbf{X}_{SSD}$. In fact in each row of the new trajectory matrix $\mathbf{X}_{SSD}$, one of the permutations of the given time series is embedded. So, the new trajectory matrix $\mathbf{X}_{SSD}$ contains L different permutations of the given time series, which are embedded in L rows, whilst in the basic SSA, the frequency of observations that are used in the trajectory matrix $\mathbf{X}$ is different.

Step 2) Singular Value Decomposition (SVD): This step is similar to basic SSA method described in Sec. 2.

Step 3) Grouping: The grouping step is once again identical to that of basic SSA and is therefore not reproduced here.

Step 4) Diagonal Averaging: For forth step, diagonal averaging, it is necessary to modify the diagonal averaging process for this new definition of the trajectory matrix, $\mathbf{X}_{SSD}$. Now if z_{ij} stands for an element of a matrix $\mathbf{Z}$, then the k-th term of the resulting series is obtained by averaging z_{ij} over all i and j such that:

$$i + j = \begin{cases} (k+1) \text{ and } (k+1) + N & \text{when } i + j < \text{N} \\ (k+1) & \text{when } i + j \geq \text{N} \end{cases} \tag{2}$$

The major point about this new diagonal averaging method is that, the sample size of averaging for every component of reconstructed series is equal to each other. For more details about SSD method see [17, 18].

3.1 Singular Spectrum Decomposition based on L_1-norm

Here we introduce a new version of SSD that is based on L_1-norm, denoted by L_1-SSD. Similar to SSD, the L_1-SSD includes four steps that are as follows:

Step 1) Embedding: This step is similar to the embedding step of SSD.

Step 2) L_1-Decomposition: In this step, the L_1-Decomposition of the trajectory matrix $\mathbf{X}_{SSD}$ is performed as $\mathbf{X}_{SSD} = \mathbf{X}_1 + \cdots + \mathbf{X}_d$, where $\mathbf{X}_i = \sqrt{\lambda_i} w_i U_i V_i^T$ and λ_i, U_i, V_i are those defined in the SVD step of SSD ($i = 1, \ldots, d$). The weight matrix $\mathbf{W} = diag(\underbrace{w_1, w_2, \ldots, w_d}_{d}, \underbrace{0, 0, \ldots, 0}_{L-d})$ is computed such that $\left\| \mathbf{X}_{SSD} - \mathbf{UW\Sigma V}^T \right\|_{L_1}$ be minimized; where $\mathbf{U} = [U_1 : \cdots : U_L]$, $\mathbf{V} = [V_1 : \cdots : V_L]$, $\mathbf{\Sigma} = diag(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \ldots, \sqrt{\lambda_L})$.

Step 3) Grouping: The grouping step is identical to that of SSD.

Step 4) L_1-Hankelization: This step is similar to the diagonal averaging step of SSD, but there is subtle difference between them. It is notable that in this step, we should calculate the median of z_{ij}s, instead of averaging them, over all i and j satisfying in Relation (2).

4 Numerical Examples

In this section, we compare basic SSA, SSD, $L_1 - SSA$ and $L_1 - SSD$ methods in terms of the quality of model fit. This is done by considering the effects of additive outliers on the decomposition and reconstructing results, evaluated using four different simulated series and two real dataset. The purpose of presenting these numerical examples is to help readers better and more easily understand the difference between the effects of outliers on SSA, SSD, $L_1 - SSA$ and $L_1 - SSD$ decomposition results. For the synthetic dataset and real dataset, we consider the following contamination scenario: *Additive Outliers* An additive outlier corresponds to an exogenous shift of a single observation of the time series and is usually associated with measurement errors. In two first simulated examples, $\delta = 3$ and $\delta = 5$ are selected based on the results in Ref. [1] and for two second simulated examples, $\delta = 15$ and $\delta = 20$ are selected. To measure the accuracy of reconstruction results, we use the commonly adopted reconstructing performance evaluation measures of Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE). They are defined as follow:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} e_i^2}$$

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^{N} |e_i|$$

where $e_i = y_i - \hat{y}_i$ is the estimated error and $\hat{y}_i$ is the estimated value of y.

4.1 Simulated Series

Example 4.1 As the first example, let us consider the simple model $y_t = sin(2\pi t/12) + \varepsilon_t$ for $t = 1, ..., 100$, where ε_t is the noise generated from the Normal distribution such that $SNR = 5$ (signal to noise variance ratio). For illustrative purposes, a simulation length of $M = 1000$ runs is considered. Since the rank of trajectory matrix for this model is equal to 2 we consider the first two eigentriples. Figures 1 and 2 show RMSE and MAE plots for different window lengths according to *Additive Outliers* scenario, where every time point y_i is replaced by $3 \times y_i$ and by $5 \times y_i$, that is, each y_i is changed by a factor of 3 and 5.

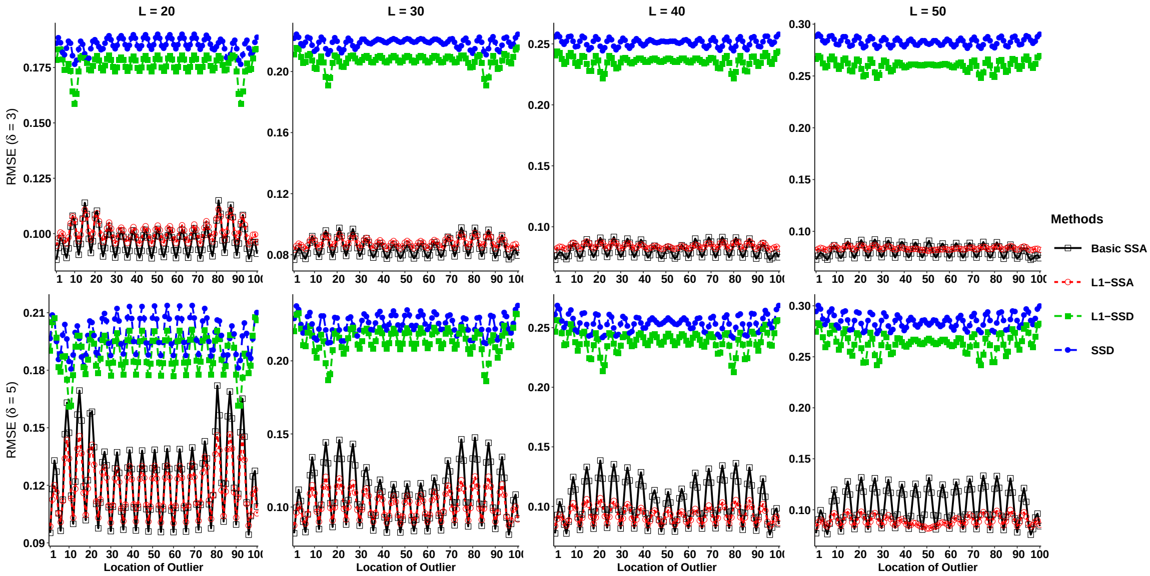

Figure 1: Plots of RMSE for $\delta = 3$ and $\delta = 5$ (Example 4.1)

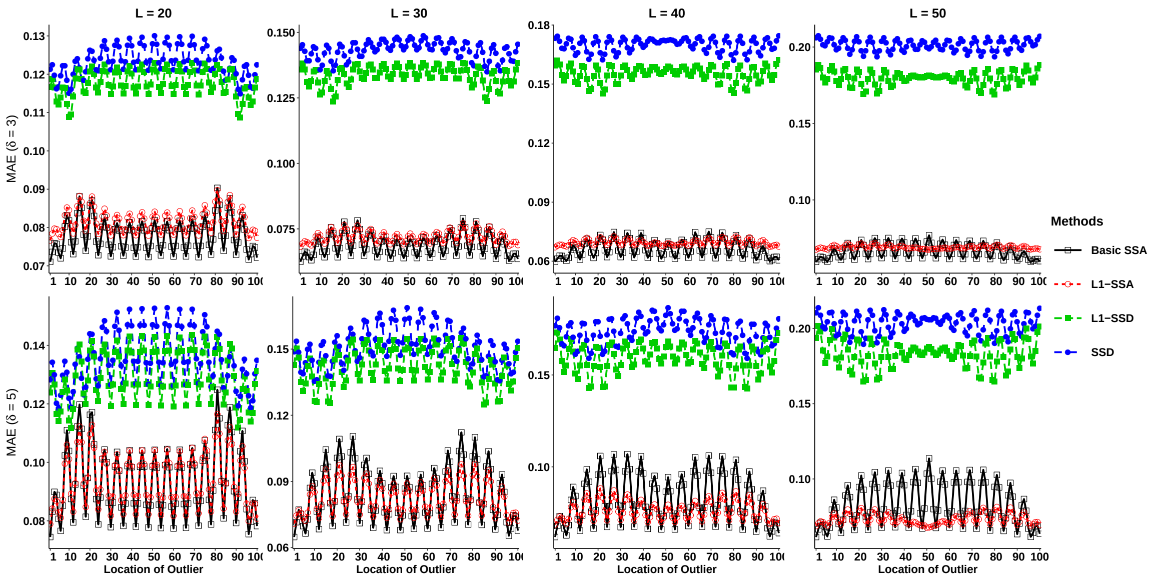

Figure 2: Plots of MAE for $\delta = 3$ and $\delta = 5$ (Example 4.1)

Undoubtedly, it can be inferred from Figures 1 and 2 that for all location of outliers in the series, RMSE and MAE of the basic SSA and $L_1 - SSA$ methods are less than the SSD and $L_1 - SSD$ methods. At a glance, by increasing the length of the window, both basic SSA and $L_1 - SSA$ methods overcome the SSD and $L_1 - SSD$ methods with more difference about 0.2.

Example 4.2 As our second example, let us consider the series $y_t = exp(\alpha_0 + \alpha_1)t + \varepsilon_t$, $(t = 1, \cdots, 100)$ where ε_t is the noise generated from the Normal distribution such that $SNR = 5$. As before, for illustrative purposes, a simulation length of $M = 1000$ runs is considered . Note that the rank of trajectory matrix is one for different values of α_0 and α_1. Therefore, the first eigentriple is considered for the reconstructing purpose. In this example, $\alpha_0 = 0$ and $\alpha_1 = 0.01$ are used. Figures 3 and 4 show RMSE and MAE plots for different window lengths according to *Additive Outliers* where every time point y_i is replaced by $3 \times y_i$ and by $5 \times y_i$, that is, each y_i is changed by a factor of 3 and 5.

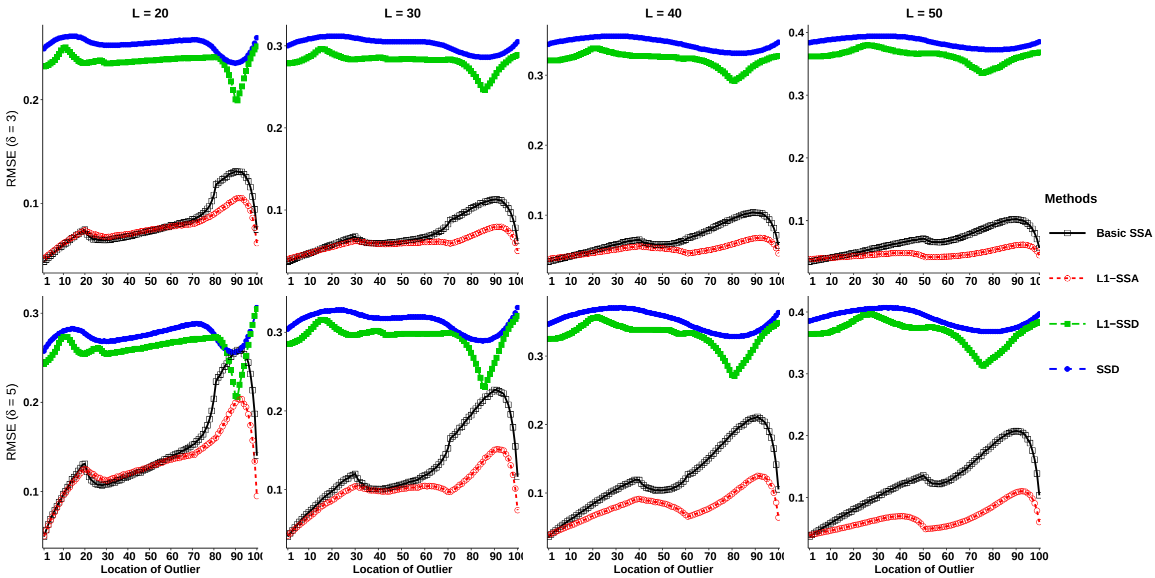

Figure 3: Plots of RMSE for $\delta = 3$ and $\delta = 5$ (Example 4.2)

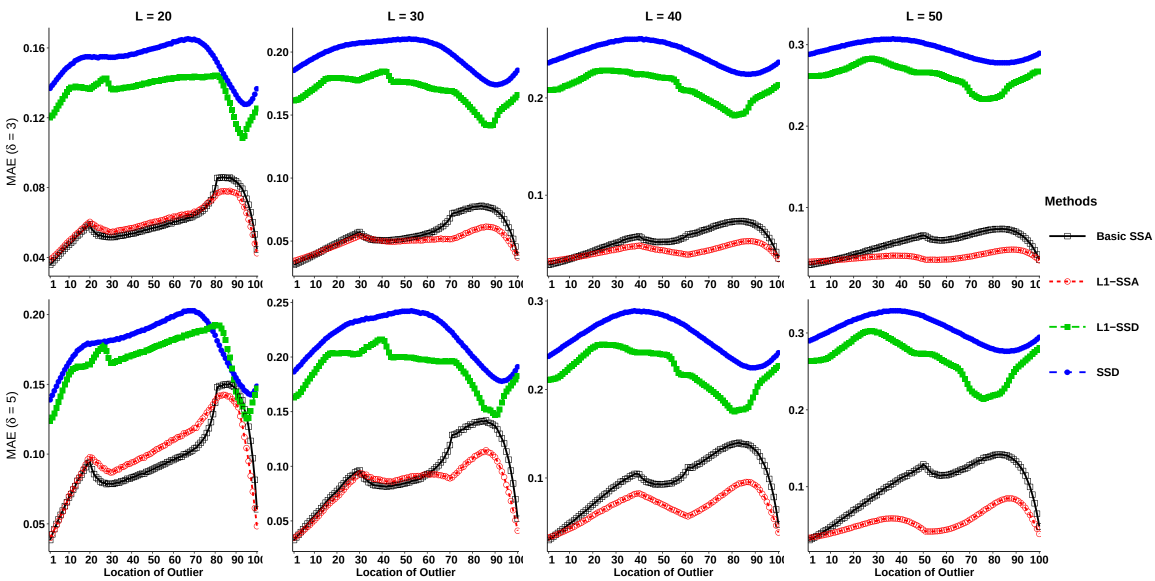

Figure 4: Plots of MAE for $\delta = 3$ and $\delta = 5$ (Example 4.2)

Having looked at Figures 3 and 4, it can be stated that, $L_1 - SSA$, basic SSA, $L_1 - SSD$ and SSD methods, respectively, are more powerful in reconstructing the *Additive Outliers* of the series.

Example 4.3 The third example is devoted to random values from the function $y_t = exp(\alpha_0 + \alpha_1)t \times t \times sin(2\pi t/12) + \varepsilon_t$, $(t = 1, \cdots, 100)$, where ε_t is the noise generated from the Normal distribution such that $SNR = 5$. As before, for illustrative purposes, a simulation length of $M = 1000$ runs is considered. Based on trajectory matrix, this function can be reconstructed with 4 eigentriples. Figures 5 and 6 show RMSE and MAE plots for different window lengths according to *Additive Outliers* where every time point y_i is replaced by $15 \times y_i$ and by $20 \times y_i$, that is, each y_i is changed by a factor of 15 and 20.

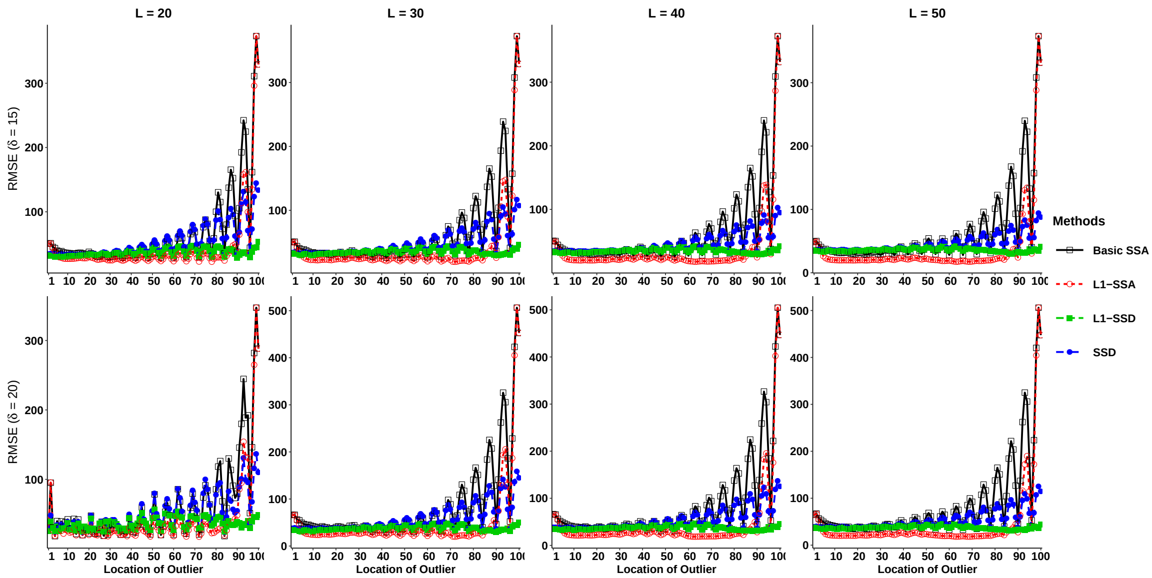

Figure 5: Plots of RMSE for $\delta = 15$ and $\delta = 20$ (Example 4.3)

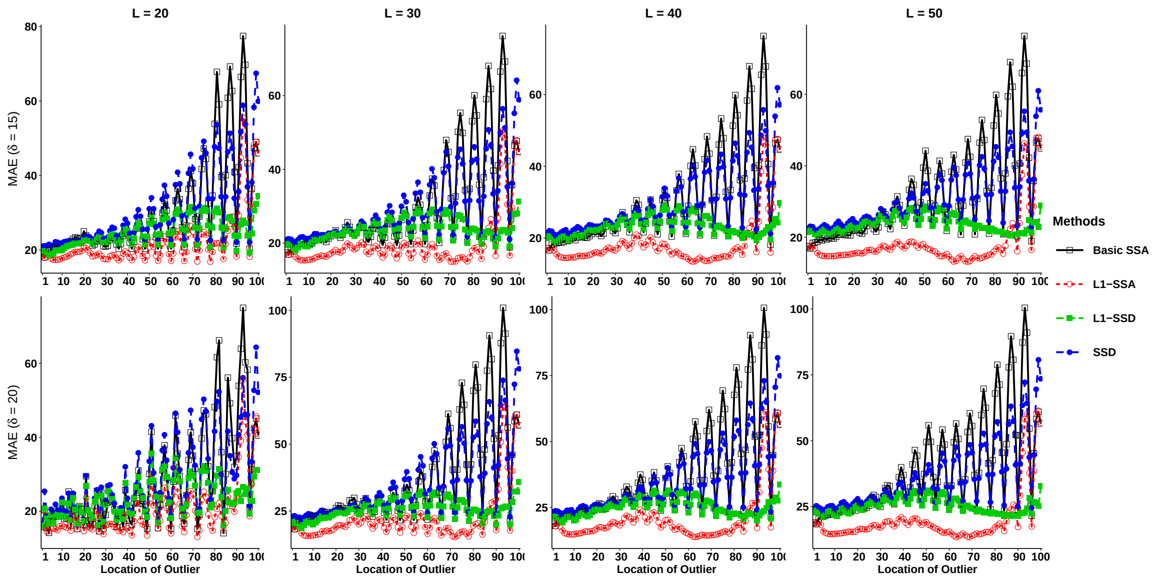

Figure 6: Plots of MAE for $\delta = 15$ and $\delta = 20$ (Example 4.3)

As it can be seen from Figures 5 and 6, it can be deduced that for all types of window lengths and δs, $L_1 - SSA$ has less errors rather than the other three methods. However, the $L_1 - SSD$ method overcomes the other three methods at the beginning and at the end of the series. Also it is clear from the Figure Figures 5 and 6, in all cases, the SSD method and $L_1 - SSD$ method are more powerful than basic SSA method.

Example 4.4 The forth example is devoted to random values from the function $y_t = exp(\alpha_0 + \alpha_1)t + t + sin(2\pi t/12) + \varepsilon_t$, $(t = 1, \cdots, 100)$, where ε_t is the noise generated from the Normal distribution such that $SNR = 5$. As before, for illustrative purposes, a simulation length of $M = 1000$ runs is considered. Based on trajectory matrix, this function can be reconstructed with 5 eigentriples. Figures 7 and 8 show RMSE and MAE plots for different window lengths according to *Additive Outliers* where every time point y_i is replaced by $15 \times y_i$ and by $20 \times y_i$, that is, each y_i is changed by a factor of 15 and 20.

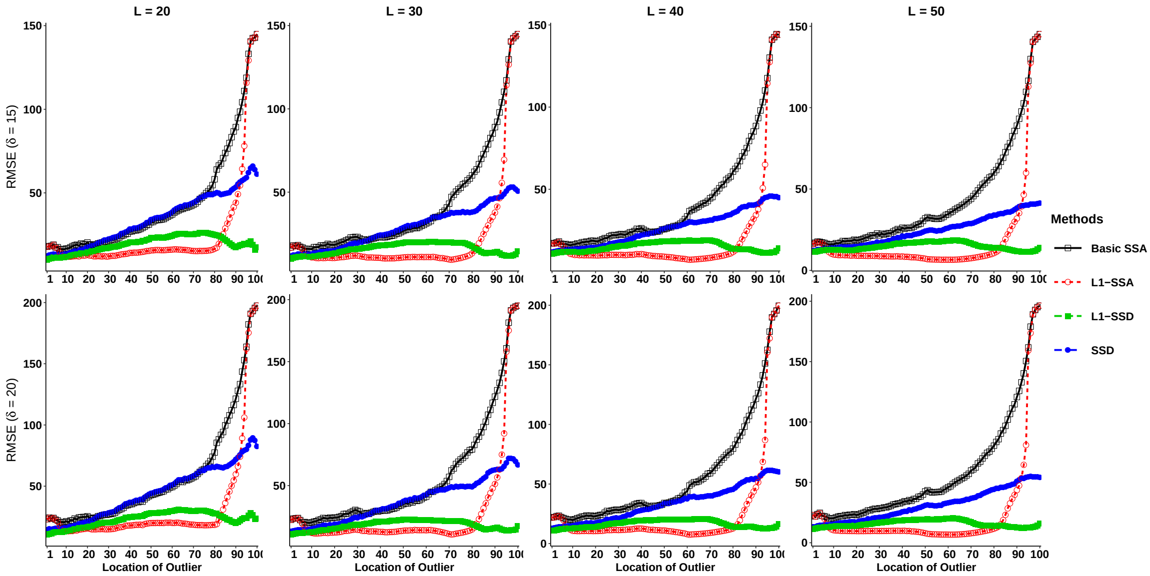

Figure 7: Plots of RMSE for $\delta = 15$ and $\delta = 20$ (Example 4.4)

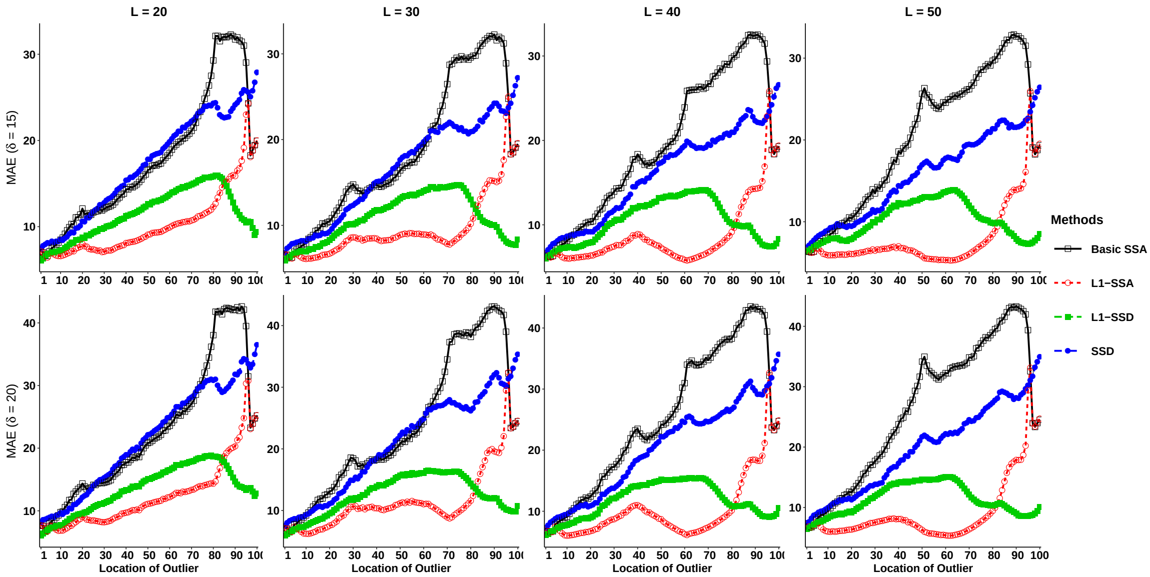

Figure 8: Plots of MAE for $\delta = 15$ and $\delta = 20$ (Example 4.4)

As it can be seen from Figures 7 and 8, it can be deduced that for all types of window lengths and δs, $L_1 - SSA$ has less errors rather than the other three methods. However, the $L_1 - SSD$ method overcomes the other three methods at the beginning and at the end of the series, especially when window length is greater than 80. Also it is clear from the Figures 7 and 8, in all cases, the SSD method and $L_1 - SSD$ method are more powerful than basic SSA method. Furthermore, the results imply that $L_1 - SSD$ is more robust than SSD method in reconstructing the *Additive Outliers* of the series.

5 Real Data Sets

In this section two different real data are considered to evaluate the efficiency of the SSD method compared to the basic SSA method in reconstructing the outlier points. The first data is related to the real interest rate in Japan, which is the lending interest rate adjusted for inflation as measured by the GDP deflator from 1961 to 2013. The second real data example is devoted to the mortality rate of children under one year out of a hundred children born in France. These two datasets are provided in [19].

Example 5.1 In this example, we consider the real interest rate in Japan, from 1961 to 2013. A real interest rate is an interest rate that has been adjusted to remove the effects of inflation to reflect the real cost of funds to the borrower and the real yield to the lender or to an investor. The real interest rate reflects the rate of time-preference for current goods with respect to future goods. Using $L = 10$, $r = 2$ and function `outlier` in R package `TSA`, it is concluded that the observations 10, 13, 14 and 54 are *Additive Outliers*, see [19]. Figure 9 shows the plot of time series and outlier points which are marked by red colour.

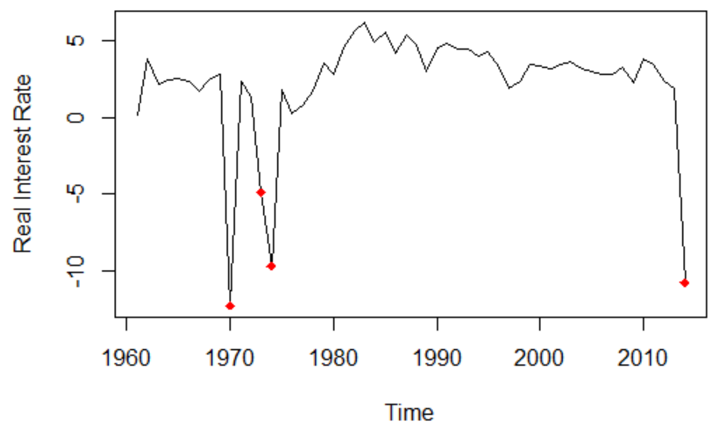

Figure 9: Time series plot of the real interest rate in Japan along with outlier points

In the following, the efficiency of the reconstruction of the Japan series which is included the entire outliers, is evaluated in terms of RMSE and MAE by using SSD, basic SSA, $L_1 - SSA$ and $L_1 - SSD$ methods. Table 1 shows the results of RMSE and MAE for the relevant series. Furthermore, it is noteworthy that the RMSE and MAE obtained for this series has its minimum values for SSD and $L_1 - SSD$ methods rather than basic SSA and $L_1 - SSA$. So it can be deduced that the SSD method overcomes the other three methods in reconstructing the series which is contaminated with outliers.

Method	$BasicSSA$	SSD	$L_1 - SSA$	$L_1 - SSD$
RMSE	2.51	2.44	2.6	2.5
MAE	1.37	1.3	1.35	1.3

Table 1: RMSE and MAE for reconstructing of Japan series

Example 5.2 This example describes dead less than 1 year mortality rate for one hundred children who born alive in France from 1977 to 1982. By applying $L = 20$ and the first two eigentriples with R function `outlier` it is deduced that the 2th, 24th and 38th observations are potentially outliers, see [19]. Figure 10 shows the plot of time series and outliers point which are marked by red colour.

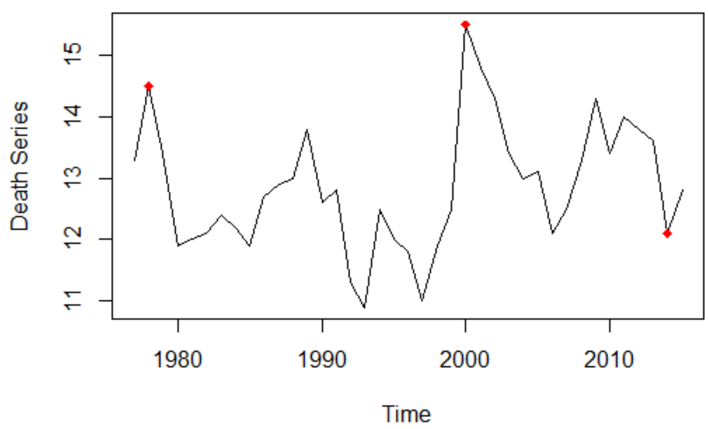

Figure 10: Time series plot of death series with outlier point

In the following, the efficiency of the reconstruction of the death series is evaluated in terms of RMSE and MAE using basic SSA, $L_1 - SSA$, SSD and $L_1 - SSA$ methods. Table 2 shows the results of RMSE and MAE for the death series. It is clear from the Table 2 that SSD method, $L_1 - SSD$, SSA and $L_1 - SSA$ have the minimum values of RMSE and MAE, respectively.

Method	$BasicSSA$	SSD	$L_1 - SSA$	$L_1 - SSD$
RMSE	0.83	0.79	0.86	0.84
MAE	0.62	0.61	0.62	0.65

Table 2: RMSE and MAE for reconstructing of death series

6 Conclusion

In this research a new approach for reconstruction was proposed in the presence of additive outliers based on Singular Spectrum Decomposition (SSD) method. In fact, in the SSD method the trajectory matrix of basic SSA is modified. By applying this improvement, the accuracy of reconstruction of time series increases, in the presence of additive outliers at the beginning and at the end of time series data. Both SSA and SSD methods were compared to reconstruct a time series, where both synthetic and real datasets were considered. The simulation studies showed that for two considered contamination scenarios, the SSD approach outperforms the basic SSA method in terms of the RMSE and MAE criteria.

References

[1] H. Hassani, R. Mahmoudvand, H. Nabe Omer, and E. Silva, "A preliminary investigation into the effect of outlier(s) on singular spectrum analysis," *Fluctuation and Noise Letters*, vol. 13, 09 2014.

[2] J. Tolvi, "Outliers in time series : A review," 1998.

[3] A. Zaharim, R. Rajali, R. Atok, I. Mohamed, and K. Jafar, "A simulation study of additive outlier in arma (1, 1) model," *International Journal of Mathematical Models and Methods in Applied Sciences*, vol. 3, no. 2, pp. 162–169, 2009.

[4] A. M. Bianco, M. G. Ben, E. J. Martínez, and V. J. Yohai, "detection in regression models with arima errors using robust estimates," *Journal of Forecasting*, vol. 20, p. 565–579, 2001.

[5] I. Chang, G. C. Tiao, and C. Chen, "Estimation of time series parameters in the presence of outliers," *Technometrics*, vol. 30, no. 2, pp. 193–204, 1988.

[6] M. Z. H. Hassani, R. Mahmoudvand and M. Ghodsi, "On the separability between signal and noise in singular spectrum analysis," *Fluctuation and Noise Letters*, vol. 11, 2012.

[7] H. Hassani, R. Mahmoudvand, and M. Yarmohammadi, "Filtering and denoising in linear regression analysis," *Fluctuation and Noise Letters*, vol. 09, 12 2010.

[8] B. Kleiner and R. D. Martin, "Robust estimation of power spectra," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 41, no. 3, pp. 313–351, 1979.

[9] S. P. Ellis and S. Morgenthaler, "Leverage and breakdown in l1 regression," *Journal of the American Statistical Association*, vol. 87, no. 417, pp. 143–148, 1992.

[10] A. J. Fox, "Outliers in time series," *Journal of the Royal Statistical Society*, vol. 34, no. 3, pp. 350–363, 1972.

[11] V. L. Paulo Canas and R. Mahmoudvand, "A robust approach to singular spectrum analysis," *Quality And Reliability Engineering International*, vol. 34, pp. 1437–1447, 6 2018.

[12] K. Choy, "Outlier detection for stationary time series," *Journal of Statistical Planning and Inference*, vol. 99, p. 111–127, 12 2001.

[13] M. Kalantari, M. Yarmohammadi, and H. Hassani, "Singular spectrum analysis based on l_1-norm," *Fluctuation and Noise Letters*, vol. 15, no. 1, p. 1650009, 2016.

[14] N. Golyandina, V. Nekrutkin, and A. Zhigljavsky, *Analysis of Time Series Structure: SSA and Related Techniques.* Monographs on Statistics and Applied Probability, Chapman & Hall/CRC, 2001.

[15] N. Golyandina and A. Zhigljavsky, *Singular Spectrum Analysis for Time Series.* Springer Briefs in Statistics, Springer, 2013.

[16] N. Golyandina, A. Korobeynikov, and A. Zhigljavsky, *Singular Spectrum Analysis with R.* Berlin Heidelberg: Springer-Verlag, 2018.

[17] M. Movahedifar, M. Yarmohammadi, and H. Hassani, "Bicoid signal extraction: Another powerful approach," *Mathematical Biosciences*, vol. 303, pp. 52 – 61, 2018.

[18] P. Bonizzi, J. Karel, O. Meste, and R. Peeters, "Singular spectrum decomposition: a new method for time series decomposition," vol. 06, 09 2014.

[19] H. Hassani and R. Mahmoudvand, *Singular Spectrum Analysis Using R.* Palgrave Pivot, 2018.