Robust singular spectrum analysis: Methodology and application

Paulo Canas Rodrigues[1,2]

(joint work with J. Pimentel, P. Messala and M. Kazemi, V. Lourenço and R. Mahmoudvand)

[1]Department of Statistics, Federal University of Bahia, Brazil

[2]Statistical Learning Laboratory (SaLLy)

UFBA Universidade Federal da Bahia

Departamento de Estatística IME - UFBA

Statistical Learning Laboratory SaLLy

Outline

1. Introduction and background
2. Singular Spectrum Analysis
   - SSA – the four steps
   - SSA – parameter selection
   - SSA forecasting
3. Robust SSA
   - SSA vs. Robust SSA
4. Concluding Remarks

Singular Spectrum Analysis

- Singular Spectrum Analysis (SSA), introduced in the seminal work of Broomhead and King (1986), is an increasingly popular extension of Principal Component Analysis (Jolliffe, 2002; Golyandina and Zhigljavsky, 2013), suited for data sets on which the dependence constraint is not fulfilled

- The core idea of singular spectrum analysis lies in the decomposition of the series of interest into several building blocks that can be classified as trends, oscillatory, or noise components

- SSA is a non-parametric approach for analyzing time series data that incorporates elements of (Golyandina et al., 2001)
  - classical time series analysis,
  - multivariate statistics, and
  - matrix algebra.

- The main assumption behind Basic SSA is that the time series can be represented as a sum of different components such as trend (which we define as any slowly varying series), modulated periodicities, and noise.

SSA – The four steps

- Basic SSA (Golyandina et al., 2001) performs four steps:
  - Stage 1: decomposition
    1. Embedding
    2. Singular Value Decomposition (SVD)
  - Stage 2: reconstruction
    3. Grouping
    4. Diagonal Averaging

SSA – The four steps

The first step – Embedding

- Let $\boldsymbol{y} = [y_1 \;\dots\; y_n]'$
- The *window length*, L, is such that $1 < L < n$
- $K = n - L + 1$ lagged vectors of length L can be obtained:

$$\boldsymbol{Y}_i = [y_i \;\; y_{i+1} \;\; \dots \;\; y_{i+L-1}]' \quad \text{for} \quad i = 1, \dots, K$$

- The *trajectory matrix* is defined as follows

$$\boldsymbol{Y} = [\boldsymbol{Y}_1 \;\dots \boldsymbol{Y}_K]' = \begin{bmatrix} y_1 & y_2 & \cdots & y_k \\ y_2 & y_3 & & y_{k+1} \\ \vdots & & \ddots & \vdots \\ y_L & y_{L+1} & \cdots & y_{L+(K-1)} \end{bmatrix}$$

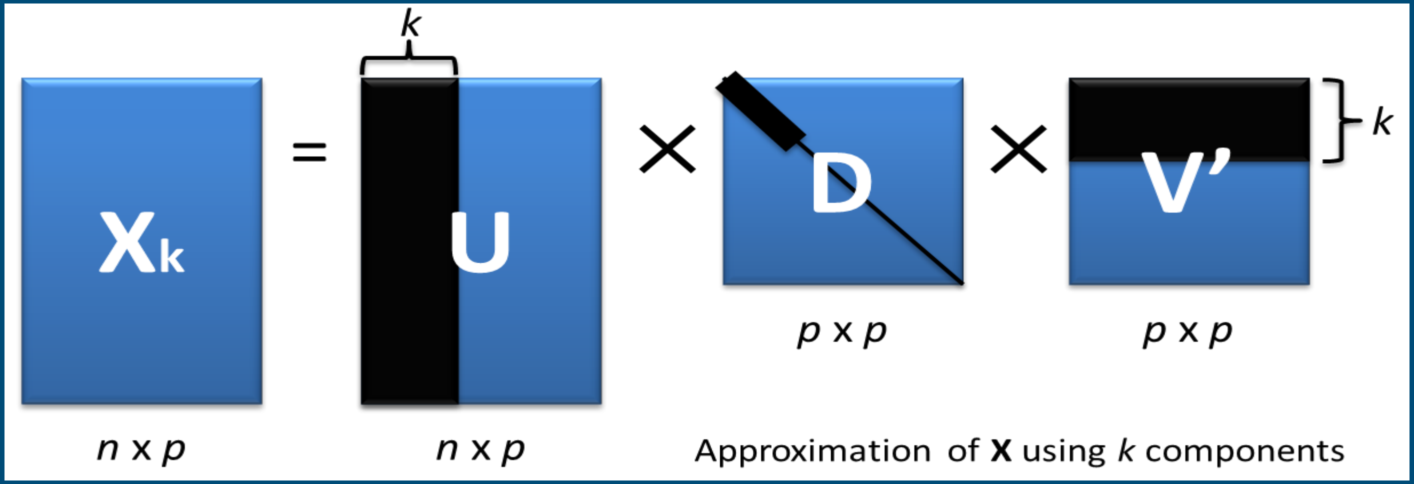
Singular Value Decomposition

We decompose the matrix $\boldsymbol{X}, rank(X) = r$ as

$$\boldsymbol{X} = \boldsymbol{UDV}^T$$

where $\boldsymbol{U}$ and $\boldsymbol{V}$ have orthonormal columns and $\boldsymbol{D}$ is diagonal; $d_1 \geq d_2 \geq \cdots \geq d_p \geq 0$. Columns of $\boldsymbol{U}$ and $\boldsymbol{V}$ are the left and right singular vectors. The diagonal of $\boldsymbol{D}$ contains the p singular values.

A low-rank approximation of $\boldsymbol{X}$ using only k factors can be written as:



The approximation minimizes the error $||\boldsymbol{X} - \boldsymbol{X}_k||_F$ (Frobenius norm).

SSA – The four steps

The second step – Singular Value Decomposition (SVD)

- At the second step we perform the SVD of the trajectory matrix $\boldsymbol{Y}$. Considering $d = rank(YY')$, we can rewrite the trajectory matrix into a sum of rank-one bi-dimensional matrices:

$$\boldsymbol{Y} = \sum_{i=1}^{d} \boldsymbol{Y}_i = \sum_{i=1}^{d} \sqrt{\lambda_i}\, \boldsymbol{U}_i\, \boldsymbol{V}_i'$$

where λ_i, $\boldsymbol{U}_i$ and $\boldsymbol{V}_i$, are the eigenvalues, the left and right singular vectors, respectively.

- The collection $(\lambda_i, \boldsymbol{U}_i, \boldsymbol{V}_i)$ is called the *i*-th *eigentriple* of the SVD of matrix $\boldsymbol{S} = \boldsymbol{YY}'$; and the similarities between this equation and the Karhunen-Loève decomposition are obvious.

SSA – The four steps

The third step – Grouping

- In the grouping step, the selection of the m principal components takes place. Let $I = 1, \dots, m$, and $I^c = m+1, \dots, d$, $d = rank(\boldsymbol{YY'})$

- The point here is to choose the first m leading eigentriples associated to the signal and exclude the remaining $d - m$ associated to the noise. I.e. we search for a 'suitable' selection of the set I, which allows us to disentangle the series $\boldsymbol{Y}$ into

$$\boldsymbol{Y} = \sum_{i \in I} \sqrt{\lambda_i}\, \boldsymbol{U}_i\, \boldsymbol{V}_i' + \varepsilon$$

where ε denotes an error term, and the remainder represents the signal.

SSA – The four steps

The fourth step – Diagonal Averaging

- Formally, consider the linear space $M_{L,K}$, formed by the collection of all the $L \times K$ matrices, and let $\{h_l\}_{l=1}^{n}$ denote the canonical basis of IR^n; Let $\boldsymbol{X} = [x_{i,j}] \in M_{L,K}$;
- The diagonal averaging procedure is hence carried on by the mapping $\overline{D}\colon M_{L,K} \to IR^n$, defiled as

$$\overline{D}(\boldsymbol{X}) = \sum_{w=2}^{K+1} \boldsymbol{h}_{w-1} \sum_{(i,j)\in A_w} \frac{x_{i,j}}{|A_w|}$$

where $A_w = \{(\mathrm{i},\mathrm{j})\colon \mathrm{i}+\mathrm{j}=\mathrm{w}\}; \mathrm{i}=1,\dots,\mathrm{L}, \mathrm{j}=1,\dots,\mathrm{K}$, and $|.|$ is the cardinal operator.

- We are now able to write the signal component of the series through the diagonal averaging procedure described above:

$$\tilde{\boldsymbol{y}} = \overline{D}\left(\sum_{i\in I} \sqrt{\lambda_i}\, \boldsymbol{U}_i\, \boldsymbol{V}_i'\right).$$

SSA – The four steps

The fourth step – Diagonal Averaging

- The central idea in this step is the reconstruction of the deterministic component of the series--the signal;
- A natural way to do this is to transfigure the matrix $\boldsymbol{Y} - \boldsymbol{\varepsilon}$ obtained in the previous step into a Hankel matrix;
- The point here is to reverse the process done so far, returning to a reconstructed variant of the trajectory matrix, and thus the signal component of the series. An optimal way to do this is to average over all the elements of the several antidiagonals.

$$\begin{bmatrix} y_{1,1} & y_{1,2} & \cdots & y_{1,k} \\ y_{2,1} & y_{2,2} & & y_{2,k} \\ \vdots & & \ddots & \vdots \\ y_{L,1} & y_{L,2} & \cdots & y_{L,k} \end{bmatrix} \rightarrow \begin{bmatrix} y_{1,1} \\ (y_{2,1} + y_{2,1})/2 \\ \vdots \\ y_{L,k} \end{bmatrix}$$

SSA – Parameter selection

- Two parameters have to be decided by the analyst:
  - The window lenght, L;
  - the number of singular values, r, to be selected for filtering/reconstruct the time series.

- Choosing improper values for the parameters L and/or r yields incomplete reconstruction and misleading results when doing forecasting.

SSA – Parameter selection

Window length L

- Considering theoretical results for the structure of the trajectory matrix and separability, it seems mostly suitable to proposed L close as half of the time series length and proportional to the number of observations per period (e.g. proportional to 12 for monthly time series), which does not guaranteed the best predictions;

- This will yield a more detailed decomposition of the time series, however it is always better to repeat the SSA analysis several times using different values of L (Golyandina and Zhigljavsky, 2013).

SSA – Parameter selection

Number of eigentriples r

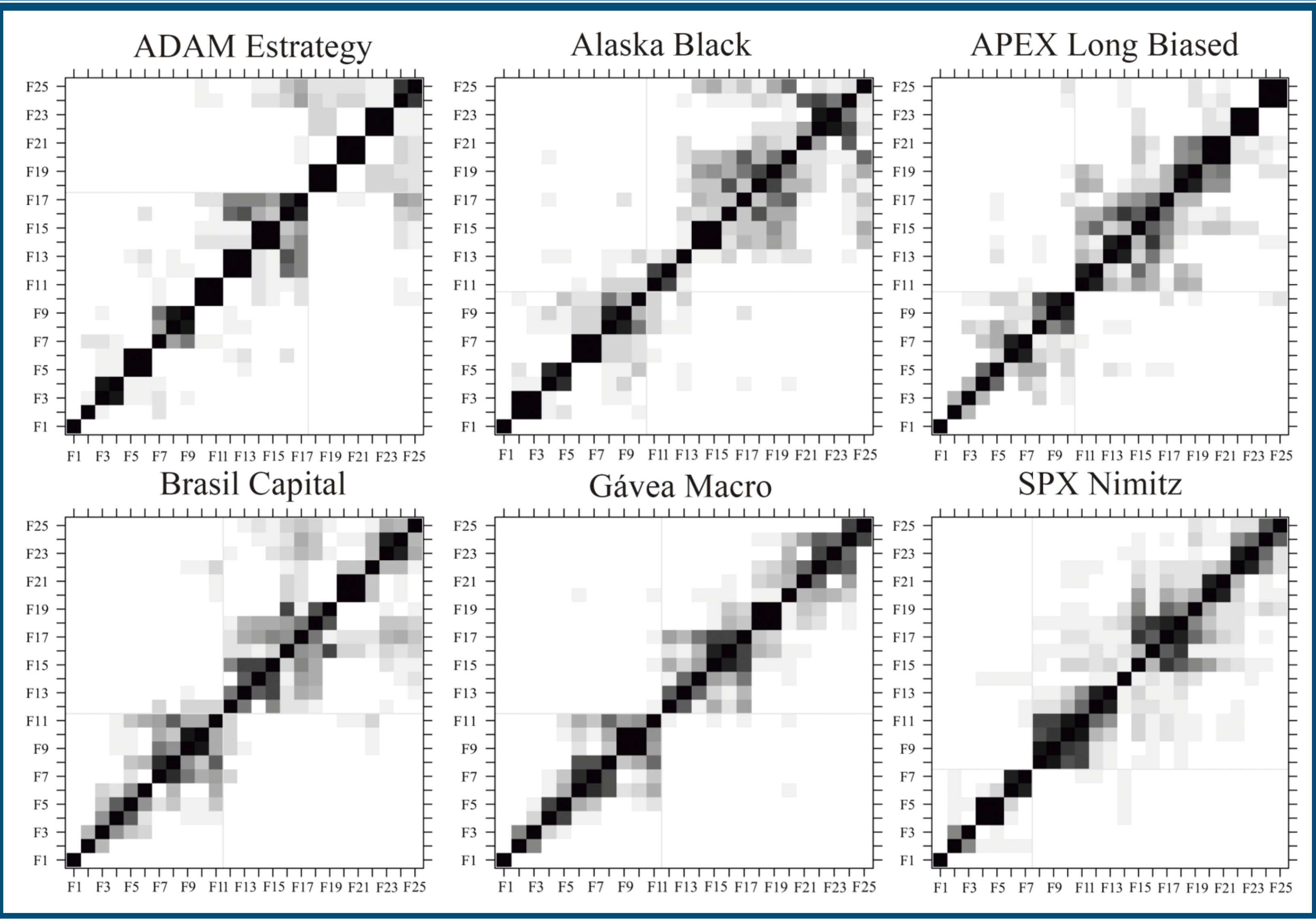
- SSA decomposition of the series Y_T can only be successful if the resulting additive components of the series are approximately separable from each other. The w-correlation is a natural measure of dependence between two series:

$$\rho_{12}^{(w)} = \frac{\left(Y_T^{(1)}, Y_T^{(2)}\right)_w}{\left\|Y_T^{(1)}\right\|_w \times \left\|Y_T^{(2)}\right\|_w},$$

where $\left\|Y_T^{(i)}\right\|_w = \sqrt{\left(Y_T^{(i)}, Y_T^{(i)}\right)_w}$; $\left(Y_T^{(i)}, Y_T^{(j)}\right)_w = \sum_{k=1}^{T} w_k\, y_k^i\, y_k^j$, $i, j = 1,2$, and $w_k = \min\{k, L, T-k\}$.

- If two reconstructed components have (near) zero w-correlation it means that these two components are separable. Large values of w-correlations between reconstructed components indicate that the components should possibly be gathered into one group and correspond to the same (signal or noise) component in SSA decomposition.

SSA – Parameter selection



SSA – Forecasting

- The basic requirement to make SSA forecasting is that the time series satisfies a linear recurrent formula (LRF). Recall that a time series $Y_T = (y_1, \dots, y_T)$ satisfies LRF of order d if

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + \cdots + a_d y_{t-d}, \qquad t = d+1, \dots, T.$$

- The time series governed by LRFs admits a natural recurrent continuation because each term of such a series is equal to a linear combination of several (d) preceding terms. The coefficients of this linear combination can be used for out-of-sample predictions.

- Although there are several versions of univariate SSA forecasting algorithms the two of the most widely used are: the Recurrent SSA (RSSA, Danilov, 1997a, b) and the Vector SSA (VSSA, Nekrutkin, 1999).

SSA – Forecasting

- Essentially, the method relies on the presumption that we are able to write the *i*-th observation y_i as a linear combination of the preceding $(L-1)$ observations.

- We are then faced with the question: What coefficients should we use in this linear recurrent formula?

- Assuming that U_j^{∇} denotes the vector of the first $L-1$ components of the eigenvector U_j, π_j is the last component of U_j, $j = 1, \dots, r$, and r the number of eigenvalues used for reconstruction, we can define the coefficient vector $\boldsymbol{a}$ as

$$\boldsymbol{a}^T = (a_{L-1}, \dots, a_1) = \frac{1}{1 - \upsilon^2} \sum_{j=1}^{r} \pi_j \, U_j^{\nabla} \, ,$$

where $\upsilon^2 = \sum_{j=1}^{r} \pi_j^2$.

SSA – Forecasting

Recurrent SSA forecast algorithm

- The 1-step-ahead out-of-sample forecast proposed by the method is then given by the following linear combination of the last $(L-1)$ reconstructed values of the series

$$\vec{y}_{n+1} = \sum_{i=1}^{L-1} a_i \tilde{y}_{n-i}$$

- In general we have that for further steps-ahead, the out-of-sample forecasts are given by

$$\begin{cases} \vec{y}_{n+2} = a_1 \vec{y}_{n+1} + \sum_{i=2}^{L-1} a_i \tilde{y}_{n-i} \\ \vdots \\ \vec{y}_{n+(L-1)} = \sum_{i=2}^{L-1} a_i \vec{y}_{n-i} \end{cases}$$

Limitations of SSA

- The Era of Big Data has brought very long and contaminated time series. Although SSA have provided advantages over traditional methods, the computational time needed for the analysis of long time series and the lack of robustness against outliers might make it unappropriated.

- If the length of the time series is very large then the conventional software performing SVD (most time-consuming step of SSA) may be computationally costly.

Alternative: Randomized SSA (*Rodrigues et al., 2018a*)

- If the data is contaminated with outlying observations, the standard SSA is likely to led to erroneous and inadequate results for model fit and forecasting.

Alternative: Robust SSA (*Rodrigues et al. 2018b; Rodrigues et al. 2020, Kezemi and Rodrigues 2023*)

Outline

Robust SSA

- The proposed robust SSA is to be an alternative to SSA for contaminated time series without losing the quality of the analysis.

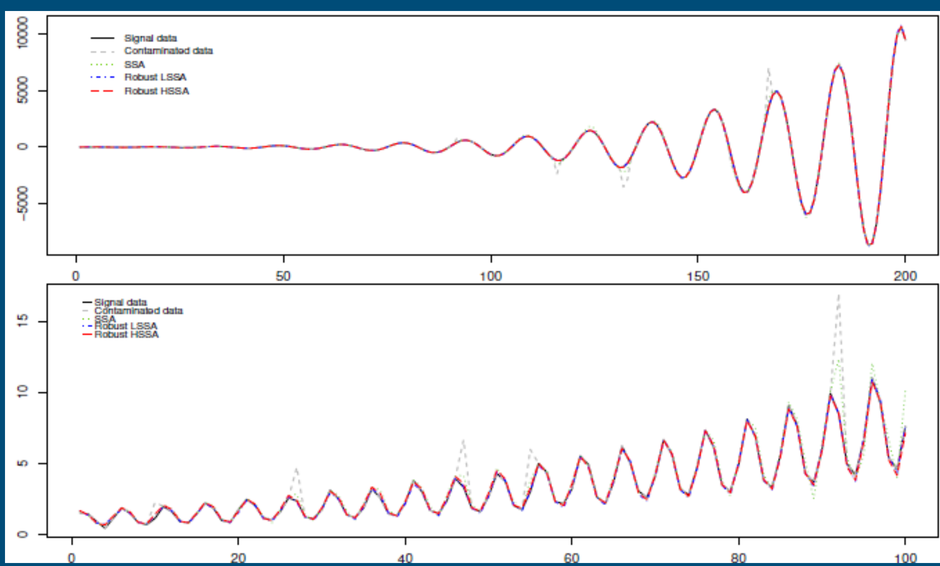
- In this new algorithm, the SVD in step two is replaced by six alternatives for robust SVD/PCA:
  - Stage 1: decomposition
    - Embedding
      - Robust SVD (Hawkins, et al., 2001; L1 norm);
      - Robust regularized SVD (2 algorithms; Zhang et al., 2013);
      - Robust PCA algorithm (Hubert et al., 2005);
      - Robust PCA based on the grid algorithm and projection pursuit (Croux and Ruiz-Gazen, 2005);
      - Robust PCA based on a robust covariance matrix (Todorov et al., 1994)

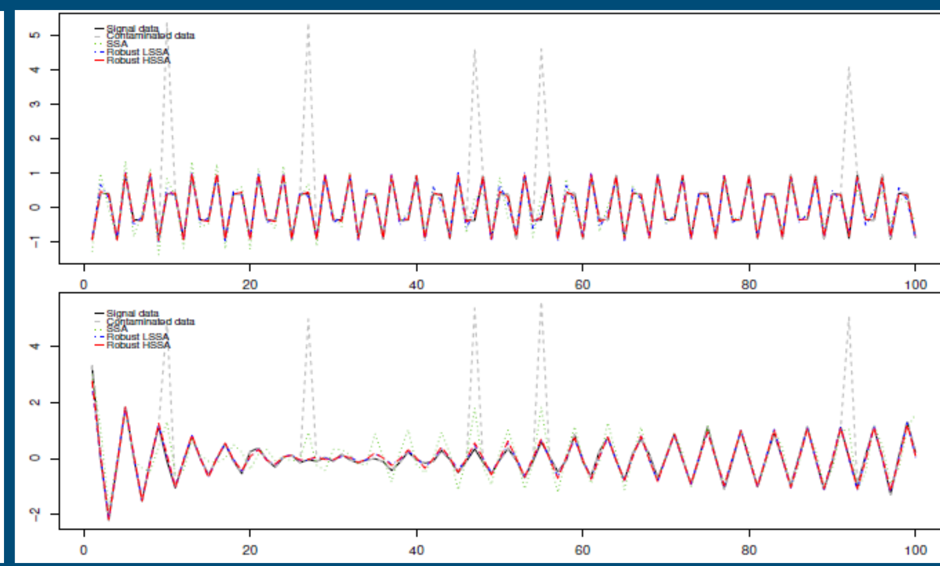
  - Stage 2: reconstruction
    - Grouping
    - Diagonal Averaging

SSA vs. Robust SSA – Simulation study

- Synthetic data

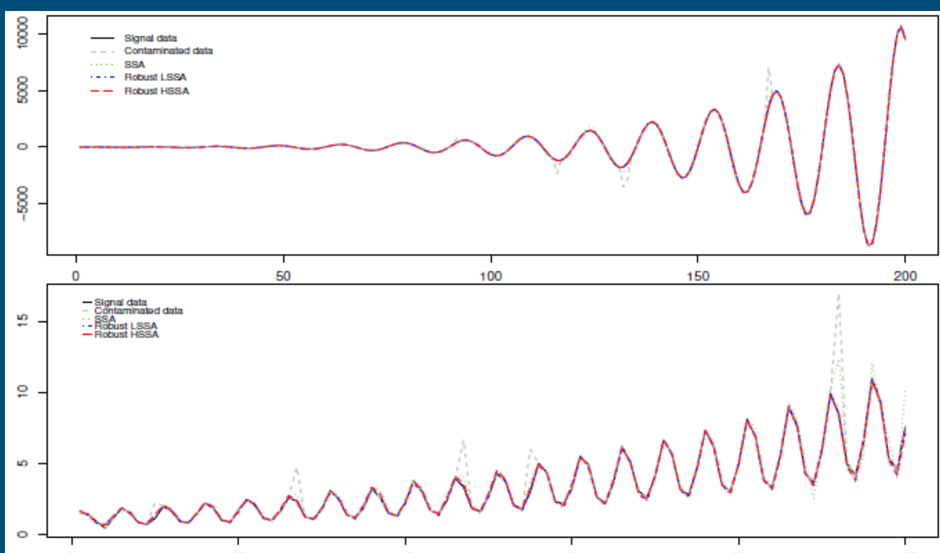
1. $f(t) = t \times \exp\left(\frac{4t}{200}\right) \times \sin\left(\frac{2\pi t}{15}\right) + \varepsilon;\ t = 1, \dots, 200;\ \varepsilon \sim N(0, 1)$
2. $f(t) = \exp\left(0.02t + 0.5 \times \sin\left(\frac{2\pi t}{5}\right)\right) + \varepsilon;\ t = 1, \dots, 100;\ \varepsilon \sim N(0, 0.1)$
3. $f(t) = \cos(2\pi \times w \times t + \varphi) + \varepsilon;\ t = 1, \dots, 100;\ \varepsilon \sim N(0, 0.01); w = \frac{3}{8}; \varphi = \frac{\pi}{8}$
4. $f(t) = \log(\alpha \times t)\cos(2\pi \times w \times t + \varphi) + \varepsilon;\ t = 1, \dots, 100;\ \varepsilon \sim N(0, 0.1); w = \frac{1}{4}; \alpha = \frac{1}{30}; \varphi = \frac{\pi}{8}$

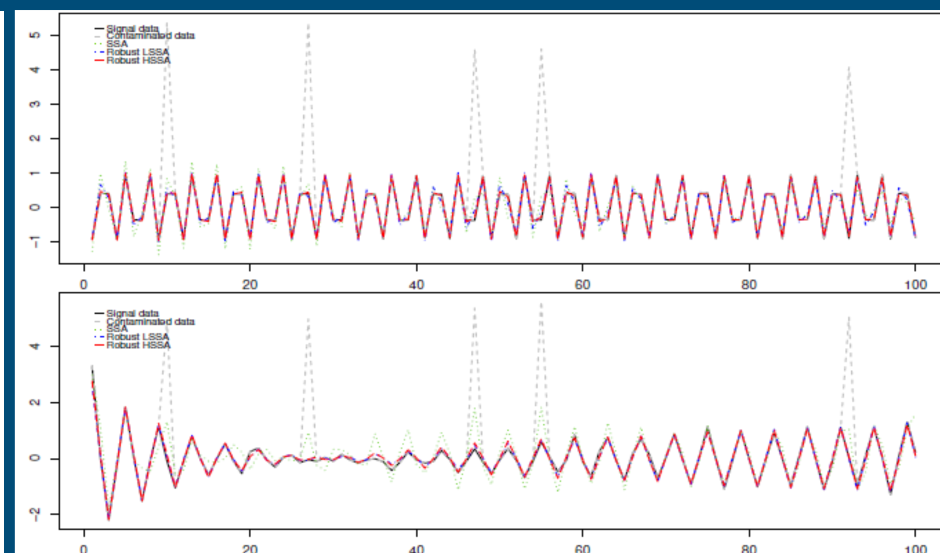




SSA vs. Robust SSA – Simulation study

- Contamination for the synthetic data
  - Magnitude increase: 2% and 5% of the time points y_i are randomly chosen to be replaced by $1.5 \times y_i$, by $2.0 \times y_i$, and by $3.0 \times y_i$, i.e., the time point magnitude of y_i is increased by a factor of 1.5, 2.0 and 3.0, respectively, resulting in three different contamination schemes. This is applied to the Simulation 1 and to the Simulation 2
  - Additive outliers: 2%, 5% and 10% of the time points y_i are randomly chosen to be replaced by $2 + y_i$, by $5 + y_i$ and by $10 + y_i$, i.e., the values of y_i are increased by a constant value of 2, 5 and 10, respectively, resulting in three different contamination schemes. This is applied to the Simulation 3 and to the Simulation 4





SSA vs. Robust SSA – Simulation study – Model fit

Table 2 Mean of the root mean square errors for model fit, computed for classic and robust SSA reconstructions for simulation 2, based on 500 runs, using $L = 24$ and $r = k = 5$

% of data contamination	Shift	Method						
		SSA	RLSSA	RHSSA	RRSSA	Hubert	Grid	Cov
0%	–	0.083	0.107	0.126	0.228	0.082	0.166	0.119
2%	$1.5 \times y_i$	0.160	0.124	0.132	0.230	0.145	0.219	0.147
5%	$1.5 \times y_i$	0.261	0.136	0.146	0.231	0.230	0.265	0.223
2%	$2 \times y_i$	0.299	0.128	0.147	0.229	0.250	0.282	0.241
5%	$2 \times y_i$	0.530	0.151	0.221	0.231	0.440	0.404	0.404
2%	$3 \times y_i$	0.584	0.130	0.191	0.232	0.464	0.449	0.438
5%	$3 \times y_i$	1.061	0.159	0.401	0.243	0.847	0.726	0.766

SSA vs. Robust SSA – Simulation study – Model fit

Table 3 Mean of the root mean square errors for model fit, computed for classic and robust SSA reconstructions for simulation 3, based on 500 runs, using $L = 24$ and $r = k = 2$

% of data contamination	Shift	Method						
		SSA	RLSSA	RHSSA	RRSSA	Hubert	Grid	Cov
0%	–	0.009	0.009	0.010	0.706	0.009	0.041	0.009
2%	$2 + y_i$	0.072	0.049	0.018	0.708	0.082	0.125	0.071
5%	$2 + y_i$	0.113	0.090	0.031	0.719	0.151	0.173	0.136
10%	$2 + y_i$	0.157	0.212	0.053	0.736	0.254	0.266	0.236
2%	$5 + y_i$	0.183	0.049	0.021	0.710	0.208	0.223	0.177
5%	$5 + y_i$	0.297	0.090	0.040	0.720	0.387	0.373	0.342
10%	$5 + y_i$	0.601	0.211	0.097	0.760	0.683	0.632	0.620
2%	$10 + y_i$	0.399	0.049	0.025	0.710	0.439	0.366	0.353
5%	$10 + y_i$	0.836	0.090	0.076	0.726	0.920	0.632	0.759
10%	$10 + y_i$	1.412	0.227	0.386	0.840	1.494	1.126	1.267

SSA vs. Robust SSA – Simulation study – Model forecasting

Table 10 Trimmed means, with 10% trimmed observations, of the root mean square errors for model forecasting ($M = 1, 3, 6$ steps-ahead), computed for classic and robust SSA reconstructions for simulation 2, based on 100 runs, using $L = 24$ and $r = 5$ for SSA and $r = k = 10$ for the robust methods

M	Cont. % of	Cont Shift of	Method					
			SSA	RLSSA	RHSSA	Hubert	Grid	Cov
	0%	–	0.390	0.203	0.235	0.189	0.192	0.193
	2%	$1.5 \times y_i$	0.386	0.240	0.661	0.379	0.396	0.380
M=1	5%	$1.5 \times y_i$	0.746	0.329	1.851	0.589	0.583	0.597
	2%	$3 \times y_i$	2.062	0.298	5.236	1.076	1.086	1.244
	5%	$3 \times y_i$	3.366	0.699	23.056	1.944	1.710	1.993
	0%	–	0.315	0.233	0.304	0.204	0.203	0.200
	2%	$1.5 \times y_i$	0.374	0.250	1.056	0.344	0.377	0.357
M=3	5%	$1.5 \times y_i$	0.790	0.458	3.130	0.572	0.554	0.561
	2%	$3 \times y_i$	4.292	0.362	25.897	1.017	1.016	1.205
	5%	$3 \times y_i$	3.442	3.966	106.550	1.888	1.743	1.894
	0%	–	0.404	0.289	0.325	0.204	0.218	0.218
	2%	$1.5 \times y_i$	0.372	0.338	2.507	0.357	0.420	0.371
M=6	5%	$1.5 \times y_i$	0.905	0.786	13.615	0.605	0.634	0.609
	2%	$3 \times y_i$	31.686	0.634	589.616	1.053	1.109	1.233
	5%	$3 \times y_i$	4.654	4.288	640.446	2.145	1.908	2.218

UFBA Universidade Federal da Bahia

Departamento de Estatística IME - UFBA

Statistical Learning Laboratory SaLLy

Outline

SSA vs. Robust SSA – USAccDeaths

(iii) the time series is contaminated by randomly assigning five isolated additive outliers along the time series; the chosen observations y_i are shifted by k units, with $k = -\bar{Y}_N, +\bar{Y}_N, +\frac{\bar{Y}_N}{2}, +\frac{\bar{Y}_N}{4}$ (dashed grey line in the top plot of Figs. 2 and 3).

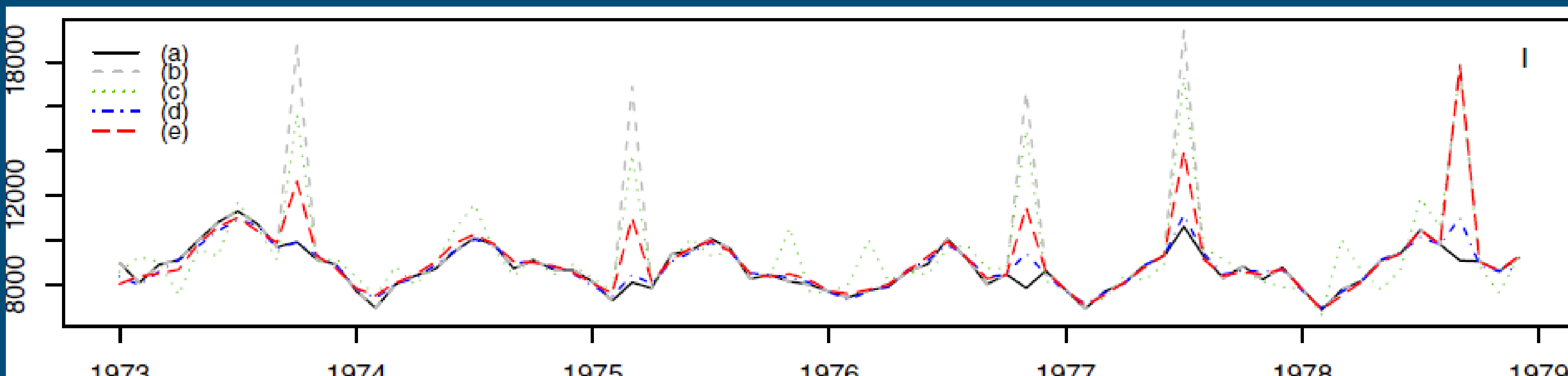


Fig. 3 a Signal data. b Contaminated data. c SSA fitted/reconstructed values. d Robust SSA fitted/reconstructed values by RLSSA and e Robust fitted/reconstructed values by RHSSA, for the simulation scenarios: (**I**) 1(**iii**) with $k = \bar{Y}_N$; (**II**) 2 with $l = 3$ and $k = \bar{Y}_N/2$; (**III**) 3(**i**); and (**IV**) 4(**i**), for the USAccDeaths data

UFBA Universidade Federal da Bahia

Departamento de Estatística IME - UFBA

Statistical Learning Laboratory SaLLy

SSA vs. Robust SSA – USAccDeaths

2. *Additive outlier patches:* The time series is contaminated by assigning $l = 3, 6, 12$, consecutive additive outliers to the l first observations of the year of 1953 for the AirPassengers dataset ($t = 49$), and of the year of 1976 for the USAccDeaths data-set ($t = 37$); the l observations y_i are shifted by k units, with $k = +\frac{\bar{Y}_N}{2}, +\frac{\bar{Y}_N}{4}, +\frac{\bar{Y}_N}{8}$ (dashed grey line in the second plot of Figs. 2 and 3).

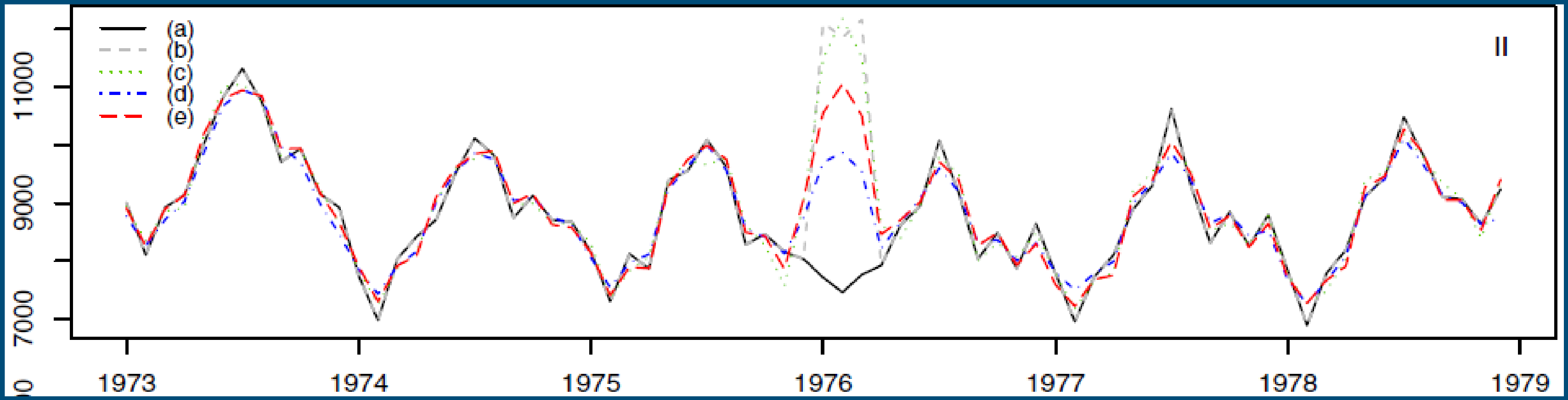


Fig. 3 a Signal data. b Contaminated data. c SSA fitted/reconstructed values. d Robust SSA fitted/reconstructed values by RLSSA and e Robust fitted/reconstructed values by RHSSA, for the simulation scenarios: (**I**) 1(**iii**) with $k = \bar{Y}_N$; (**II**) 2 with $l = 3$ and $k = \bar{Y}_N/2$; (**III**) 3(**i**); and (**IV**) 4(**i**), for the USAccDeaths data

UFBA Universidade Federal da Bahia
Departamento de Estatística IME - UFBA
Statistical Learning Laboratory SaLLy

SSA vs. Robust SSA – Model fit

Table 8 Mean of the root mean square errors for model fit, computed for classic and robust SSA reconstructions for the USAccDeaths data, considering the contamination scenario 1, based on 500 runs, using $L = 24, r = 13$

Scenario	Shift	Method						
		SSA	RLSSA	RHSSA	RRSSA	Hubert	Grid	Cov
No cont.	–	108.37	162.56	112.36	262.09	117.78(182.27)	184.90(256.72)	119.27(154.82)
1(i)	$-\bar{Y}$	1036.34	228.23	514.59	326.73	672.75(609.4)	701.00(589.3)	680.29(562.1)
	$+\bar{Y}$	1029.52	195.42	535.17	268.54	650.87(573.3)	645.35(559.0)	664.49(535.6)
	$+\frac{1}{2}\bar{Y}$	499.37	187.65	204.11	269.20	375.14(354.4)	357.74(346.5)	345.33(301.5)
	$+\frac{1}{4}\bar{Y}$	238.90	187.23	141.28	274.09	230.82(249.4)	239.51(278.0)	201.48(211.2)
1(ii)	$-\bar{Y}$	863.97	222.57	244.09	274.86	703.75(640.55)	673.23(608.25)	658.39(588.80)
	$+\bar{Y}$	862.97	208.50	223.29	269.18	697.47(627.74)	656.21(574.15)	644.58(580.56)
	$+\frac{1}{2}\bar{Y}$	438.28	201.12	155.73	261.41	370.71(354.54)	357.73(356.61)	351.35(340.30)
	$+\frac{1}{4}\bar{Y}$	226.84	197.58	139.92	264.22	220.18(236.96)	230.90(269.44)	209.52(226.40)
1(iii)	$-\bar{Y}$	2027.49	680.28	1379.83	468.70	1912.07(1781.4)	1654.69(1438.5)	1779.77(1653.4)
	$+\bar{Y}$	2022.91	616.37	1389.46	447.91	1890.97(1745.9)	1621.03(1391.7)	1761.64(1620.1)
	$+\frac{1}{2}\bar{Y}$	1009.71	384.47	656.02	382.51	930.68(857.33)	820.48(725.97)	891.42(823.60)
	$+\frac{1}{4}\bar{Y}$	505.39	295.81	355.61	346.02	471.71(444.62)	430.09(410.95)	458.99(433.82)

SSA vs. Robust SSA – Model forecast

M	Cont.		Method					
			SSA	RLSSA	RHSSA	Hubert	Grid	Cov
	No cont.	–	572.67	672.75	604.03	558.47	553.23	452.61
	Con. 1(i)	$+\bar{Y}$	9070.91	2801.96	8045.17	2951.07	3184.54	2968.44
		$+\frac{1}{2}\bar{Y}$	2608.62	1545.16	2953.89	1529.85	1543.24	1479.82
M=1	Con. 1(ii)	$+\bar{Y}$	2076.08	603.53	848.88	1002.05	891.21	1080.18
		$+\frac{1}{2}\bar{Y}$	1005.35	596.46	637.11	712.76	680.63	671.97
	Con. 1(iii)	$+\bar{Y}$	3889.34	2730.14	7603.73	2876.41	2563.99	2813.67
		$+\frac{1}{2}\bar{Y}$	2086.66	1517.48	2327.57	1556.94	1447.67	1570.45
	No cont.	–	640.14	779.42	592.28	502.33	617.37	430.22
	Con. 1(i)	$+\bar{Y}$	39734.71	2867.18	12363.35	2828.56	3189.99	2799.70
		$+\frac{1}{2}\bar{Y}$	3043.70	1582.43	9319.68	1464.19	1524.42	1438.18
M=3	Con. 1(ii)	$+\bar{Y}$	2284.97	779.14	2933.24	1061.55	1045.60	1209.82
		$+\frac{1}{2}\bar{Y}$	1166.66	778.01	899.44	790.80	763.76	854.35
	Con. 1(iii)	$+\bar{Y}$	4203.61	5081.90	1.55e+4	3026.98	2761.09	3047.03
		$+\frac{1}{2}\bar{Y}$	2229.57	1893.21	3464.55	1702.30	1561.83	1705.65

UFBA Universidade Federal da Bahia

Departamento de Estatística IME - UFBA

Statistical Learning Laboratory SaLLy

Outline

1. Introduction and background

2. Singular Spectrum Analysis
   - SSA – the four steps
   - SSA – parameter selection
   - SSA forecasting

3. Robust SSA
   - SSA vs. Robust SSA – Mutual investment funds

4. Concluding Remarks

Concluding Remarks

- SSA is a non-parametric approach for analyzing time series data which incorporates elements of: (i) classical time series analysis; (ii) multivariate statistics; and (iii) matrix algebra
- Although SSA have provided advantages over traditional methods, the results might be inadequate due to data contamination with outlying observations;
- The robust SSA proposed here performs well when the data is contaminated with outlying observations.

Thank you for your attention!

Questions/Remarks/Suggestions?

E-mail: paulocanas@gmail.com

Web: www.paulocanas.org

Statistical learning laboratory: www.SaLLy.ufba.br

/company/88682389

/sally.laboratory

References

➢ Kazemi, M. and Rodrigues, P.C. (2023). Robust singular spectrum analysis: Comparison between classic and robust approaches for model fit and forecasting. Computational Statistics. DOI: 10.1007/s00180-022-01322-4

➢ Rodrigues, P.C., Lourenço, V.M. and Mahmoudvand, R. (2018). A robust approach to singular spectrum analysis. Quality and Reliability Engineering International. 34:1437–1447.

➢ Rodrigues, P.C., Tuy, P.G.S.E. and Mahmoudvand, R. (2018). Randomized singular spectrum analysis for long time series. Journal of Statistical Computation and Simulation. 88:1921–1935.

➢ Rodrigues, P.C., Pimentel, J., Messala, P., and Kazemi, M. (2020). Decomposition and forecasting of mutual investment funds using singular spectral analysis. Entropy 22, 83.

UFBA Universidade Federal da Bahia

Departamento de Estatística IME - UFBA

Statistical Learning Laboratory SaLLy