Title: Singular Spectrum Analysis

Name: U. Radojičić[1], K. Nordhausen[2], S. Taskinen[2]

Affil./Addr.: [1]Institute of Statistics & Mathematical Methods in Economics

TU Wien, Austria

una.radojicic@tuwien.ac.at

[2]Department of Mathematics and Statistics

University of Jyväskylä,

Jyväskylä, Finland

sara.l.taskinen@jyu.fi, klaus.k.nordhausen@jyu.fi

Singular Spectrum Analysis

Definition

Singular Spectrum Analysis (SSA) is a family of methods for time series analysis and forecasting, which seeks to decompose the original series into a sum of a small number of interpretable components such as trend, oscillatory components, and noise.

Introduction

Singular Spectrum Analysis (SSA) aims at decomposing the observed time series into the sum of a small number of independent and interpretable components such as a slowly varying trend, oscillatory components, and noise (Elsner and Tsonis, 1996; Golyandina et al, 2001). SSA can be used, for example, for finding trends and seasonal components (both short and large period cycles) in time series, smoothing, and forecasting. When SSA is used as an exploratory tool, one does not need to know the underlying model of the time series.

The origins of SSA are usually associated with nonlinear dynamics studies (Broomhead and King, 1986a,b), and over time, the method has gained a lot of attention. As stated in (Golyandina et al, 2001), SSA has proven to be very successful in time series analysis and is now widely applied in the analysis of climatic, meteorological, and geophysical time series. Many variants of SSA exist. However, in the following we focus on a specific SSA method, often referred to as *basic SSA*.

Singular spectrum analysis

We define $\boldsymbol{x} = \{x_t : t = 1, \ldots, N\}$ for an observable time series. The SSA algorithm consists of the following four steps (Golyandina et al, 2018).

1. **Embedding**: First, the so-called $l \times k$-dimensional *trajectory matrix*

$$\boldsymbol{T}(\boldsymbol{x}) = \begin{pmatrix} x_1 & x_2 & x_3 & \cdots & x_k \\ x_2 & x_3 & x_4 & \cdots & x_{k+1} \\ x_3 & x_4 & x_5 & \cdots & x_{k+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_l & x_{l+1} & x_{l+1} & \cdots & x_N \end{pmatrix} \tag{1}$$

is constructed. Here N is the time series length, l is so-called *window length* and $k = N - l + 1$. The trajectory matrix $\boldsymbol{T}$ is a linear map mapping $\boldsymbol{x} \in \mathbb{R}^N$ into a $l \times k$-dimensional Hankel matrix, *i.e.*, to $l \times k$-dimensional matrix $\boldsymbol{X} = \boldsymbol{T}(\boldsymbol{x})$ with equal elements on the off-diagonals. Columns $\boldsymbol{x}_i = (x_i, \ldots, x_{i+l-1})'$, $i = 1, \ldots, k$, of $\boldsymbol{X}$ are called as *lagged vectors* of dimension l. Golyandina et al (2001) suggests that l should be smaller than $N/2$, but sufficiently large so that l-lagged vector $\boldsymbol{x}_i$, $i = 1, \ldots, k$, incorporates the essential part of the behaviour in the initial time series $\boldsymbol{x}$.

2. **Decomposition**: Once the time series $\boldsymbol{x}$ is embedded into a trajectory matrix $\boldsymbol{X}$, it is decomposed into a sum of rank-1 matrices. Denote now $\boldsymbol{S} = \boldsymbol{X}\boldsymbol{X}'$ and write $\boldsymbol{S} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}'$ for an eigendecomposition of $\boldsymbol{S}$. Here $\boldsymbol{\Lambda}$ is a $l \times l$ diagonal matrix with non-negative eigenvalues, $\lambda_1 \geq \cdots \geq \lambda_l$, as diagonal values and $\boldsymbol{U} = (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_l)'$, $\boldsymbol{u}_i \in \mathbb{R}^l$, is an $l \times l$ orthogonal matrix that includes the corresponding eigenvectors as columns. If we further write $\boldsymbol{v}_i = \boldsymbol{X}'\boldsymbol{u}_i/\sqrt{\lambda_i}$, $i = 1, \ldots d$, where $d = \text{rank}(\boldsymbol{X})$, then $\boldsymbol{X}$ can be decomposed into a sum of rank-1 matrices as follows

$$\boldsymbol{X} = \sum_{i=1}^{d} \boldsymbol{X}_i = \sum_{i=1}^{d} \sqrt{\lambda_i}\, \boldsymbol{u}_i \boldsymbol{v}_i'. \tag{2}$$

Notice that $\boldsymbol{u}_i \in \mathbb{R}^l$ and $\boldsymbol{v}_i \in \mathbb{R}^k$ are left and right singular vectors of $\boldsymbol{X}$, respectively, and $\sqrt{\lambda_i} > 0$ are the corresponding singular values. In SSA literature these ordered singular values are often referred to as the *singular spectrum* thus giving the name to the method (Elsner and Tsonis, 1996). It is important to mention that the above method, where $\boldsymbol{X}$ is being decomposed into rank-1 components using singular value decomposition, is called basic SSA. In practice, also other decompositions can be used. For more details, see (Golyandina et al, 2001), for example.

3. **Grouping**: In this step, the rank-1 components in decomposition (2) are grouped into predefined groups, where the group membership is described by a set of indices $I = \{i_1, \ldots, i_p\} \subset \{1, \ldots, d\}$, $p \leq d$. Then, the matrix corresponding to the group I is defined as $\boldsymbol{X}_I = \boldsymbol{X}_{i_1} + \cdots + \boldsymbol{X}_{i_p}$. If the set of indices $\{1, \ldots, d\}$ is partitioned into m disjoint subsets $I_1, \ldots, I_m$, $m \leq d$, then one obtains so-called *grouped decomposition* of matrix $\boldsymbol{X}$,

$$\boldsymbol{X} = \boldsymbol{X}_{I_1} + \cdots + \boldsymbol{X}_{I_m}. \tag{3}$$

Especially, if $I_k = k$, for $k = 1, \ldots, d$, the grouping is called elementary and the corresponding components in (3) are called elementary matrices. Furthermore, if only one group $I \subset \{1, \ldots, d\}$ is specified, one proceeds by assuming that the given partition is $\{I, I^c\}$, where $I^c = \{1, \ldots, d\} \setminus I$. In that case, $\boldsymbol{X}_I$ usually corresponds to the pattern of interest, while $\boldsymbol{X}_{I^c} = \boldsymbol{X} - \boldsymbol{X}_I$ is treated as the residual.

4. **Reconstruction**: As final step, matrices $\boldsymbol{X}_{I_j}$, $j = 1, \ldots, m$, from decomposition (3) are transformed into new time series of length N. Reconstructed $\tilde{\boldsymbol{x}}_{I_k}$ are obtained by sequentially averaging the elements of matrix $\boldsymbol{X}_{I_k}$, $k = 1, \ldots, m$, that lay on the off-diagonals, *i.e.*,

$$[\tilde{\boldsymbol{x}}_{I_k}]_t = \frac{1}{|S_t|} \sum_{i+j=t+1} [\boldsymbol{X}_{I_k}]_{i,j}, \tag{4}$$

where $|S_t|$ denotes the cardinality of the finite set $S_t = \{(i, j) : i + j = t + 1\}$. In the literature, this process is known as *diagonal averaging*. If one applies the given reconstruction to all components in group decomposition (3) and, for simplicity of the notation, denotes $\tilde{\boldsymbol{x}}_k := \tilde{\boldsymbol{x}}_{I_k}$, $k = 1, \ldots, m$, the resulting decomposition of initial time series $\boldsymbol{x}$ is given by

$$\boldsymbol{x} = \tilde{\boldsymbol{x}}_1 + \cdots + \tilde{\boldsymbol{x}}_m. \tag{5}$$

In the case of basic SSA for univariate time series, the tuning parameters one needs to specify *a priori* are window length l and the partition of the set of indices. In more general SSA, the trajectory matrix, as well as its rank-1 decomposition, can be chosen more freely.

Separability

A central concept in SSA is separability, which to some extent ensures the validity of the method. Assume that one can decompose the time series into two univariate series,

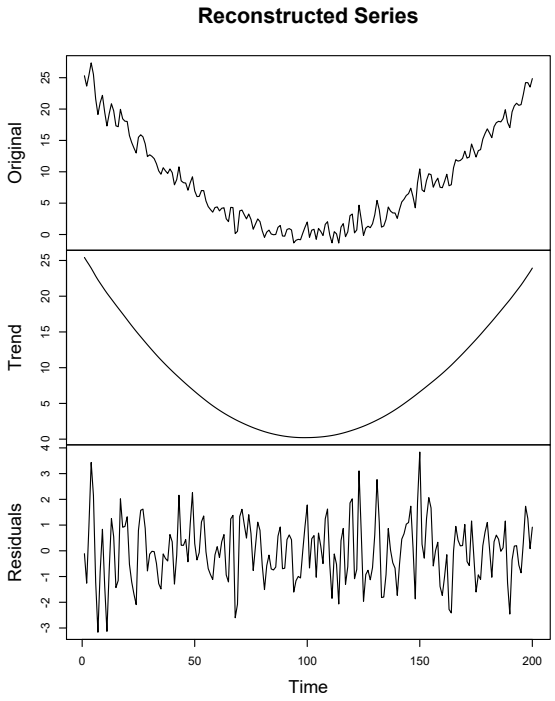
that is, $\boldsymbol{x} = \boldsymbol{x}_1 + \boldsymbol{x}_2$. This representation is usually associated with a *signal plus noise* model, *trend plus the remainder* model and other structured models. (Approximate) separability of the components $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ implies that there exist a grouping (see Step 4) such that reconstructed $\tilde{\boldsymbol{x}}_1$ and $\tilde{\boldsymbol{x}}_2$ are (approximately) equal to $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, respectively, *i.e.*, $\tilde{\boldsymbol{x}}_i \approx \boldsymbol{x}_i$, $i = 1, 2$. In basic SSA, the (approximate) separability corresponds to (approximate) orthogonality of the trajectory matrices.

Consider as an example a time series of length N. If N is large enough, the trend, which is a slowly varying smooth component, and periodic components are (approximately) separable and both are (approximately) separable from the noise. For illustration see Figure 1, where a trend and a periodic component are separated from a noise component. In Figure 1 (a), the trend is extracted using the first three significant elementary components, with window length size $l = 100$. In Figure 1 (b), the seasonal component is extracted using the first six significant elementary components, with window length size $l = 150$. The extracted components are approximately equal to the theoretical trend and seasonal component.

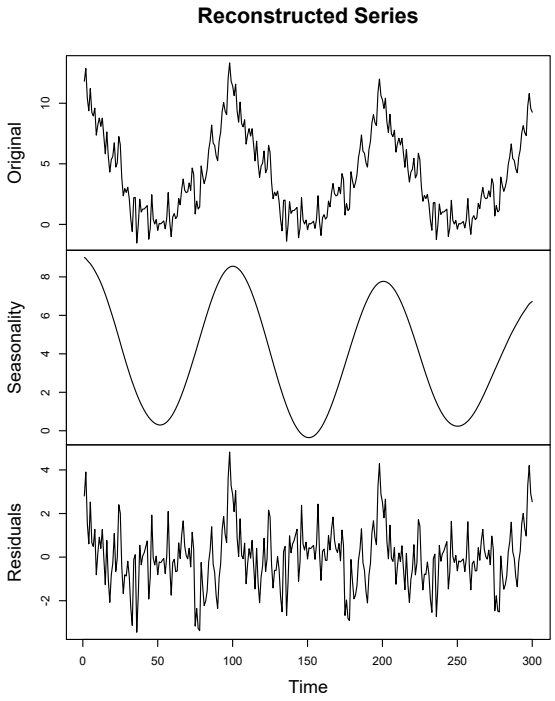
In order to check the separability of the reconstructed components $\tilde{\boldsymbol{x}}_1$ and $\tilde{\boldsymbol{x}}_2$, with corresponding trajectory matrices $\tilde{\boldsymbol{X}}_1$ and $\tilde{\boldsymbol{X}}_2$, respectively, the normalized orthogonality measure given in (Golyandina et al, 2018) is

$$\rho(\tilde{\boldsymbol{X}}_1, \tilde{\boldsymbol{X}}_2) = \frac{\langle \tilde{\boldsymbol{X}}_1, \tilde{\boldsymbol{X}}_2 \rangle_F}{\|\tilde{\boldsymbol{X}}_1\|_F \|\tilde{\boldsymbol{X}}_2\|_F},$$

where $\langle \cdot \rangle_F$ and $\| \cdot \|_F$ are Frobenius matrix inner product and norm, respectively. This orthogonality measure further induces dependence measure between two time series called *w-correlation* (Golyandina and Korobeynikov, 2014). A large value of w-correlation between a pair of elementary components suggests that, in the grouping step, these should perhaps be in the same group. Furthermore, additive sub-series can be identified using the principle that the form of an eigenvector resembles the form



(a) Extraction of the trend



(b) Extraction of the oscillatory component

Fig. 1: Left: extraction of the polynomial trend in time series $x_t = (0.05(t-1)-5)^2 + y_t$, $t = 1, \ldots, 200$, where y_t is $MA(0.9)$ process. Right: extraction of the oscillatory component in time series $x_t = a_t y_t$, where $a_t = 1$, for $t = 1, \ldots, 100$, $a_t = 0.9$ for $t = 101, \ldots, 200$ and $a_t = 0.9^2$ for $t = 201, \ldots, 300$. $y_t = (0.1(t-1)-5)^2 + z_t$, where z_t is $MA(0.9)$ process.

of the sub-series produced by the eigenvector. For example, the eigenvectors produced by a slowly-varying component (trend) are slowly-varying and the eigenvectors produced by a sine wave (periodic component) are again sine waves with the same period (Golyandina and Korobeynikov, 2014). Therefore, plots of eigenvectors are often used in the process of identification. For more details, see (Golyandina and Korobeynikov, 2014; Golyandina et al, 2018), for example.

Prediction using SSA

A class of time series especially suited for SSA are so-called *finite-rank* time series, where we say that the time series $\boldsymbol{x}$ is of finite-rank if its trajectory matrix is of rank $d < \min(k, l)$ and does not depend on window length l, for l large enough. In that case, under mild conditions, there exists one-to-one correspondence between the trajectory matrix of the time series $\boldsymbol{x}$ and a linear recurrent relation (LRR)

$$x_{t+d} = \sum_{i=1}^{d-1} a_i x_{t+i}, \quad t = 1, \ldots, N-d, \tag{6}$$

that governs the time series $\boldsymbol{x}$. Furthermore, the solution to LRR (6) can then be expressed as sums and products of polynomial, exponential and sinusoidal components, whose identification leads to the reconstruction of trend and various periodic components, among others. If one applies obtained LRR (6) to the last terms of the initial time series $\boldsymbol{x}$, one obtains the continuation of $\boldsymbol{x}$ which serves as a prediction of the future.

The real-data time series are not in general finite-rank time series. However, if time series $\boldsymbol{x}$ is a sum of a finite-rank signal $\boldsymbol{x}_1$ and additive noise, then SSA may approximately separate the signal component and one can further use the methods designed for analysis and forecasting of the finite-rank series, thus obtaining the continuation (forecast) of a signal component $\boldsymbol{x}_1$ of $\boldsymbol{x}$ (Golyandina and Korobeynikov, 2014; Golyandina et al, 2001). Such a problem is known as forecasting the signal (trend, seasonality...) in the presence of additive noise. The confidence intervals for the forecast can be obtained using bootstrap techniques. For more details, see (Golyandina et al, 2001; Golyandina and Korobeynikov, 2014).

Example

To illustrate the SSA method, we use the average monthly temperatures (in °C) measured at the Jyväskylä airport, in Finland, from 1960 to 2000. To perform basic SSA as described above, we use the `R` (R Core Team, 2020) package `Rssa` (Golyandina et al, 2015). As suggested in (Golyandina et al, 2001), for extracting a periodic component in short time series, it is preferable to take the window length l proportional to the period. Thus, we take $l = 48$. After investigating the singular values and paired scatter plots of the left singular vectors, we identify the first component as depicting a slowly varying trend, and the grouped fifth and sixth components to depict the yearly seasonality of the average monthly temperature. The original time series is shown together with its decomposition into seasonal component, trend component and residuals in Figure 2. As seen in Figure 2, the seasonal component is clearly dominating. However, there exists also a trend that looks almost linear. Eliminating the noise and using only the reconstructed trend and yearly seasonality, we predict the average monthly temperature for years 2000 and 2001 based on the components shown in Figure 2. The predictions together with the true observed values are shown in Figure 3. Figure 3 shows that SSA can be used to predict the temperatures in Jyväskylä quite nicely. For more details and guidelines for grouping and identification of trend and oscillatory components, we refer to Golyandina et al (2001).

Extensions and relations to other approaches

SSA has been extended to multivariate time series in which case it is known as multi-dimensional or multi-channel SSA (M-SSA) and for the analysis of images where it is known as 2D singular spectrum analysis (2D-SSA).

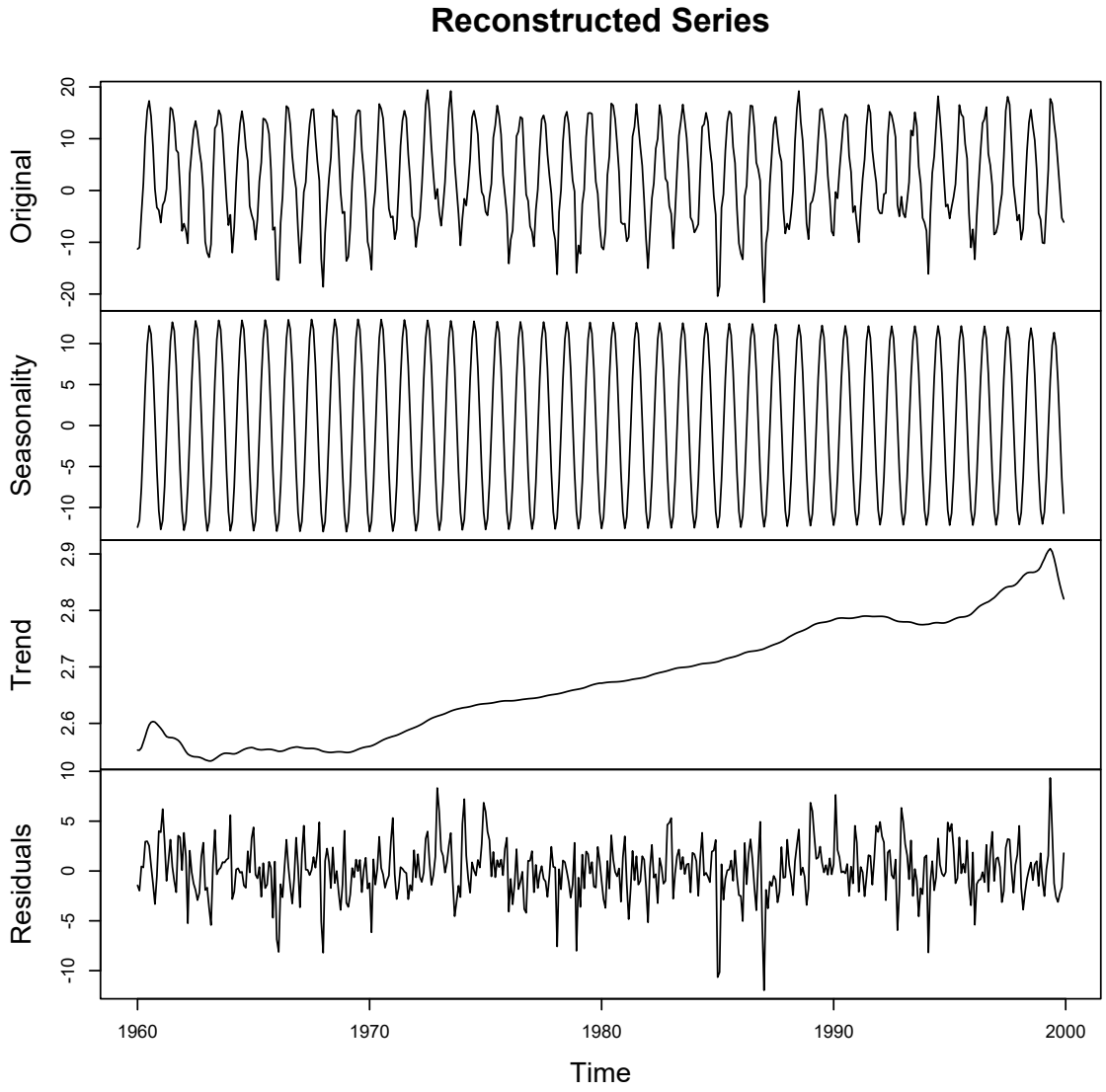


Fig. 2: Top to bottom: Average monthly temperature (°C) at the Jyväskylä airport from 1960 to 2000, the reconstructed seasonal component with 12 month period, reconstructed trend component, and residuals.

While usually time series analysis is either performed in the time or the frequency domain one can see SSA as a compromise, as a time-frequency method. For example, SSA can be seen as a method that chooses an adaptive basis generated by the time series itself while, for example, Fourier Analysis uses a fixed basis of sine and cosine functions. For more detailed discussions and interpretations of SSA, we refer to Elsner and Tsonis (1996); Golyandina et al (2001); Ghil et al (2002).

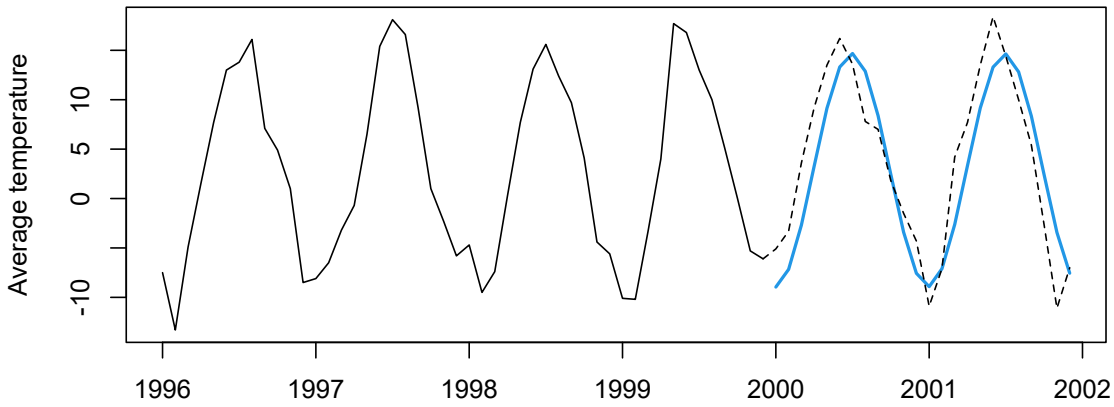


Fig. 3: The predicted average monthly temperature (°C) at the Jyväskylä airport in °C for 2000 and 2001 (blue line). The dashed black line shows the true observed average temperatures.

Summary

Singular Spectrum Analysis (SSA) is a model-free family of methods for time series analysis, that can be used, for example, for eliminating noise in the data, identifying interpretable components such as trend and oscillatory components, forecasting and imputing missing values. The method consists of the four main steps: embedding, decomposition, grouping, and reconstruction. The method is currently widely applied in the analysis of climatic, meteorological, and geophysical time series.

Cross References

Chaos and Singularity Analysis in Geosciences, Singular Value Decomposition, Time Series Analysis, Time Series Analysis in the Geosciences

References

Broomhead D, King GP (1986a) Extracting qualitative dynamics from experimental data. Physica D: Nonlinear Phenomena 20(2):217–236

Broomhead D, King GP (1986b) On the qualitative analysis of experimental dynamical systems. Nonlinear Phenomena and Chaos pp 113–144

Elsner JB, Tsonis AA (1996) Singular Spectrum Analysis: A New Tool in Time Series Analysis. Plenum Press, New York

Ghil M, Allen MR, Dettinger MD, Ide K, Kondrashov D, Mann ME, Robertson AW, Saunders A, Tian Y, Varadi F, Yiou P (2002) Advanced spectral methods for climatic time series. Reviews of Geophysics 40(1):3–1–3–41

Golyandina N, Korobeynikov A (2014) Basic singular spectrum analysis and forecasting with R. Computational Statistics and Data Analysis 71:934–954

Golyandina N, Nekrutkin V, Zhigljavsky A (2001) Analysis of time series structure: SSA and related techniques. Chapman & Hall/CRC, Boca Raton

Golyandina N, Korobeynikov A, Shlemov A, Usevich K (2015) Multivariate and 2D extensions of singular spectrum analysis with the Rssa package. Journal of Statistical Software 67(2):1–78

Golyandina N, Korobeynikov A, Zhigljavsky A (2018) Singular spectrum analysis with R. Springer, Berlin

R Core Team (2020) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, URL https://www.R-project.org/