

Monte Carlo SSA for Extracting Weak Signals

E. POTESHKIN¹, N. GOLYANDINA²

^{1 2}*St. Petersburg State University
St. Petersburg, Russia*

e-mail: ¹egor.poteshkin@yandex.ru, ²n.golyandina@spbu.ru

The paper addresses the issue of extracting signals from noise using singular spectrum analysis (SSA). We propose an algorithm that automatically selects significant modulated harmonics without specifying their periods. This algorithm relies on the Monte Carlo SSA criterion to identify the significant frequencies, which are then extracted.

Keywords: time series, signal extraction, signal detection, singular spectrum analysis

1 Introduction

Consider the following model: $X = S + R$, where $X = (x_1, \dots, x_N)$ is the observed time series, S is the signal and R is the noise, i.e., the realisation of some stationary process. This paper considers two problems: signal detection and signal extraction in cases where the signal is present.

To solve the first problem, we use the Monte Carlo SSA (MC-SSA) [2] method, which tests the hypothesis $H_0 : S = 0$. The second problem is solved by applying singular spectrum analysis (SSA) [4, 7], which decomposes the series into elementary components. After grouping these components, we obtain a decomposition of the series into trend, periodic components, and noise. As one of the SSA steps involves visual analysis to identify the signal components, there is a need to automate this step. This problem has been addressed in works such as [1, 3, 6, 8], which are mainly devoted to trend extraction or smoothing. This paper aims to define an approach to the automatic extraction of weak oscillating signals detected by the MC-SSA criterion.

2 The autoMCSSA algorithm

Let us introduce some notation and assumptions.

The proposed algorithm uses a modified version of MC-SSA that corrects for multiple comparisons [5]. Sine waves with equidistant frequencies $\omega_k = k/(2L)$, $k = 1, \dots, L$, were chosen as the projection vectors needed to construct the criterion statistics. With this choice of projection vectors, we can identify significant frequencies present in the signal.

For a series X of length N and $0 \leq \omega_1 \leq \omega_2 \leq 0.5$, we will use the same frequency measure used in [1] for trend extraction. Let the measure $T(X; \omega_1, \omega_2)$ reflect the contribution of frequencies from the interval $[\omega_1, \omega_2]$ calculated from the periodogram of the series.

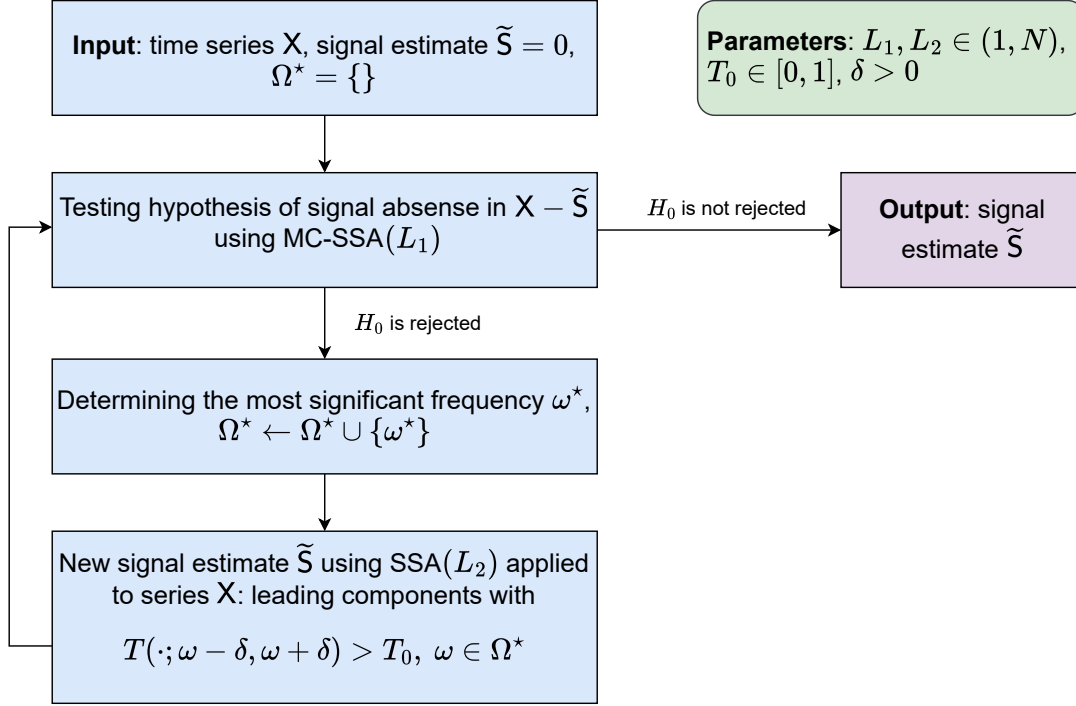


Figure 1: Flowchart of autoMCSSA

Figure 1 shows the flowchart of the autoMCSSA algorithm. It consists of applying the MC-SSA criterion sequentially until the hypothesis is no longer rejected. If the hypothesis is rejected at the next iteration of the algorithm, the most significant frequency ω^* is determined and a new signal estimate is calculated using SSA applied to the original series: for each frequency ω from the set of significant frequencies Ω^* , the leading components with a measure T exceeding the threshold T_0 are selected. Once the null hypothesis is no longer rejected, the algorithm terminates, with the final signal estimate being given by \tilde{S} .

We will estimate the frequency ω^* using a weighted average of the nearest significant frequencies, where the weights are determined by the frequencies' significance. This estimation method enables us to obtain a more accurate estimate of ω in cases when it does not fall on the $k/(2L)$ frequency grid.

Let us consider an example of the work of the proposed algorithm. Let $\mathbf{X} = \mathbf{S} + \boldsymbol{\xi}$, where $\boldsymbol{\xi}$ is the AR(1) model with parameters $\phi = 0.7$ and $\sigma^2 = 1$, $N = 200$, $\mathbf{S} = (s_1, \dots, s_N)$,

$$s_n = 0.075 e^{0.02n} \cos(2\pi n/8) + 2 \cos(2\pi n/4) + 0.2 \cdot (-1)^n.$$

Figure 2 shows the first 15 elementary components reconstructed using SSA. The signal corresponds to components with indices 1, 2, 5, 6, and 13. Without knowing the formula by which this signal is generated, it is difficult to say with certainty which components are non-random, since the components (3, 4) and (11, 12) look like pairs of harmonics. We applied the autoMCSSA algorithm to this series and found that

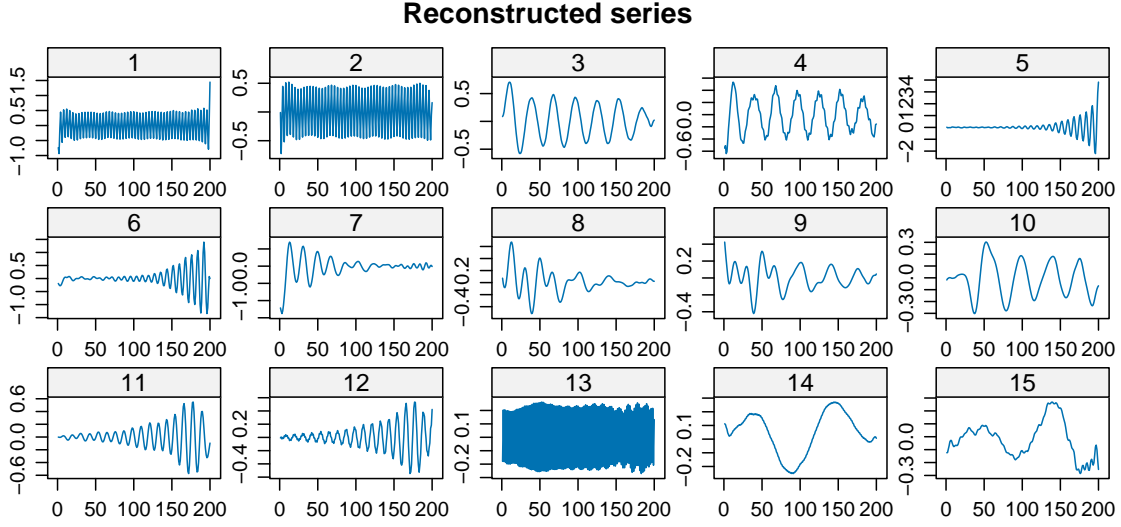


Figure 2: Elementary series components

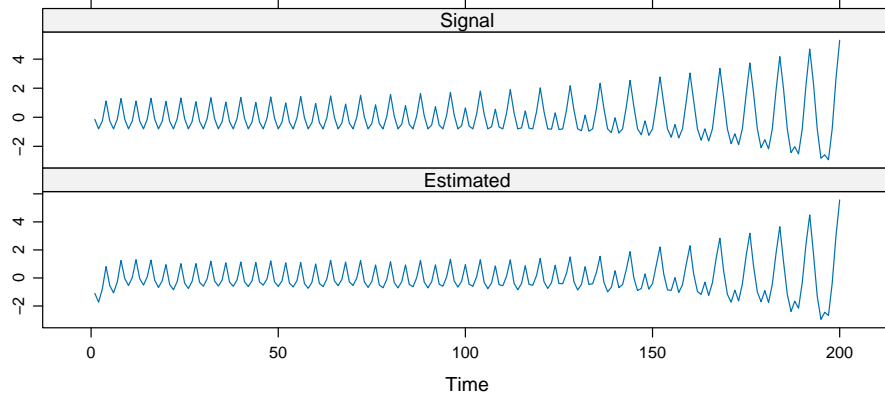


Figure 3: Estimated signal using autoMCSSA ($L_1 = 50$, $L_2 = 100$, $\delta = 1/80$, $T_0 = 0.5$)

the developed method correctly identified the components corresponding to the signal. Figure 3 shows the signal S and its estimate by the autoMCSSA method.

3 Conclusion

The paper proposes an algorithm capable of extracting only the significant time series components without specifying the period of a periodic component. Multiple modulated periodic components may be present, each with different amplitude modulation.

References

- [1] Alexandrov Th. (2009). A method of trend extraction using singular spectrum analysis. *RevStat*. Vol. **7**, Num. **1**, pp. 1-22.
- [2] Allen M., Smith L. (1996). Monte Carlo SSA: detecting irregular oscillations in the presence of coloured noise. *Journal of Climate*. Vol. **9**, pp. 3373-3404.
- [3] Bogalo J., Poncela P., Senra E. (2021). Circulant singular spectrum analysis: A new automated procedure for signal extraction. *Signal Processing*. Vol. **179**, 107824.
- [4] Broomhead D., King G. (1986). Extracting qualitative dynamics from experimental data. *Physica D: Nonlinear Phenomena*. Vol. **20**, Num. **2-3**, pp. 217-236.
- [5] Golyandina N. (2023). Detection of signals by Monte Carlo singular spectrum analysis: multiple testing. *Statistics and Its Interface*. Vol. **16**, Num. **1**, pp. 147-157.
- [6] Golyandina N., Dudnik P., Shlemov A. (2023). Intelligent identification of trend components in singular spectrum analysis. *Algorithms*. Vol. **16**, Num. **7**, 353.
- [7] Golyandina N., Nekrutkin V., Zhigljavsky A. (2001). *Analysis of Time Series Structure: SSA and Related Techniques*. Chapman and Hall/CRC: New York.
- [8] Kalantari M., Hassani. H. (2019). Automatic grouping in singular spectrum analysis. *Forecasting*. Vol. **1**, Num. **1**, pp. 189-204.