

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

В работе рассматривается задача выделения сигнала из шума методом анализа сингулярного спектра (SSA). Предлагается алгоритм автоматического выделения значимых модулированных гармоник без задания их периодов, который на основе критерия Monte Carlo SSA определяет значимые частоты и затем выделяет их.

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

1 Введение

Рассмотрим следующую модель: $X = S + R$, где $X = (x_1, \dots, x_N)$ — наблюдаемый временной ряд, S — сигнал, R — шум, т.е. реализация некоторого стационарного процесса. В работе рассматриваются две проблемы: проблема обнаружения сигнала S и проблема выделения сигнала при его наличии.

Для решения первой проблемы используется метод Monte Carlo SSA (MC-SSA) [1], проверяющий гипотезу $H_0 : S = 0$, а для решения второй — метод анализа сингулярного спектра (singular spectrum analysis, SSA) [2, 3], раскладывающий ряд на элементарные компоненты, после группировки которых получается разложение ряда на тренд, периодические компоненты и шум. Один из шагов SSA подразумевает визуальный анализ для определения компонент сигнала, поэтому возникает потребность в автоматизации этого шага, этой проблеме посвящены, например, работы [4–7], посвященные, главным образом, выделению тренда или сглаживанию. Целью работы является определение подхода к автоматическому выделению слабых осциллирующих сигналов, обнаруживаемых критерием MC-SSA.

2 Метод autoMCSSA

Для описания разработанного метода необходимо ввести некоторые обозначения и предположения.

В предлагаемом алгоритме используется модификация MC-SSA с поправкой на множественные сравнения [8]. В качестве векторов для проекции, необходимых для построения статистики критерия, были выбраны косинусы с равноотстоящими частотами $\omega_k = k/(2L)$, $k = 1, \dots, L$. В этом случае можно говорить о значимых частотах, присутствующих в сигнале.

Для ряда X длины N и $0 \leq \omega_1 \leq \omega_2 \leq 0.5$ будем использовать ту же меру, что в [4] для выделения тренда. Пусть мера $T(X; \omega_1, \omega_2)$ отражает долю вклада частот из интервала $[\omega_1, \omega_2]$, сосчитанную на основе значения периодограммы ряда.

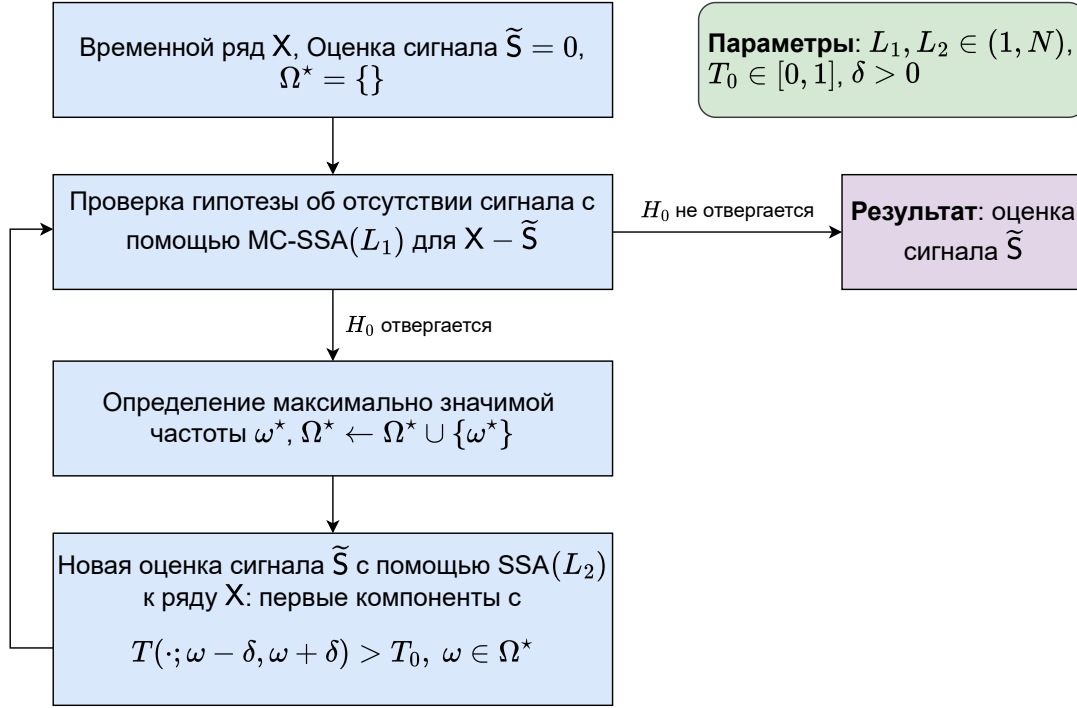


Рис. 1: Алгоритм autoMCSSA

На рис. 1 изображена блок-схема алгоритма autoMCSSA. Заключается он в последовательном применении критерия MC-SSA до тех пор, пока гипотеза не перестанет отвергаться. Если на очередной итерации алгоритма гипотеза отвергается, определяется максимально значимая частота ω^* и вычисляется новая оценка сигнала с помощью SSA, примененного к исходному ряду: для каждой частоты ω из множества значимых частот Ω^* выбираются первые компоненты с мерой T превышающей порог T_0 . Как только гипотеза перестает отвергаться, алгоритм завершает свою работу и тогда \tilde{S} — итоговая оценка сигнала.

Оценивать частоту ω^* будем с помощью взвешенного среднего ближайших значимых частот с весами, определяемыми их значимостью. Такой способ оценки позволяет получить более точную оценку ω в случае, когда она не попадает в решетку $k/(2L)$.

Рассмотрим пример работы рассмотренного алгоритма. Пусть $X = S + \xi$, где ξ — модель AR(1) с параметрами $\phi = 0.7$ и $\sigma^2 = 1$, $N = 200$, $S = (s_1, \dots, s_N)$,

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

На рис. 2 представлены первые 15 элементарных восстановленных с помощью SSA компонент. Сигналу соответствуют компоненты с индексами 1, 2, 5, 6 и 13. Не видя формулы, по которой этот сигнал задается, сказать наверняка, какие компоненты неслучайные, проблематично, поскольку компоненты 3, 4 и 11, 12 похожи на пары гармоник. Мы применили алгоритм autoMCSSA к этому ряду и получили, что разработанный метод правильно идентифицировал компоненты,

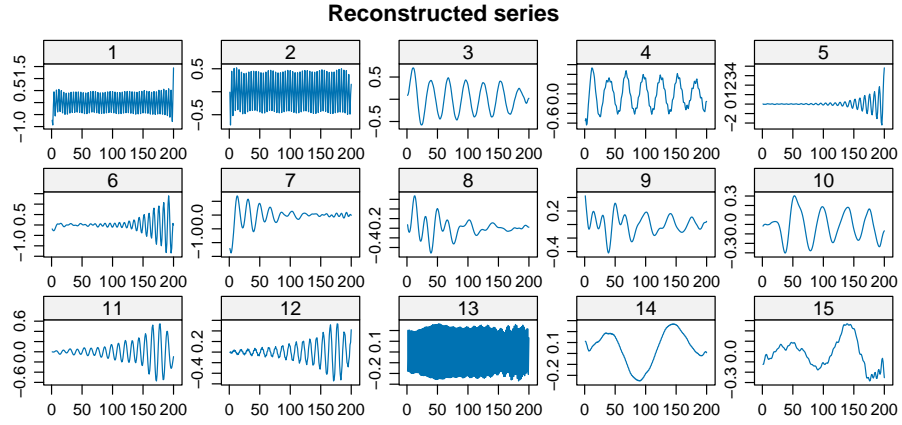


Рис. 2: Элементарные восстановленные компоненты

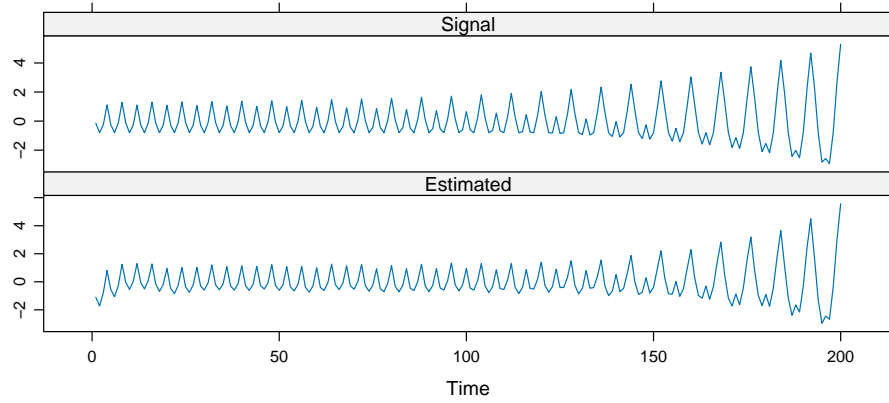


Рис. 3: Результат autoMCSSA ($L_1 = 50$, $L_2 = 100$, $\delta = 1/80$, $T_0 = 0.5$)

соответствующие сигналу, на рис. 3 представлены истинная форма сигнала S и его оценка методом autoMCSSA.

3 Заключение

Таким образом, в работе предлагается алгоритм, способный выделить только значимые компоненты, при этом не требуется задавать период для периодической компоненты, периодик может быть несколько с неизмеримыми периодами и они могут быть модулированными с разной амплитудной модуляцией.

Список литературы

- [1] Allen M., Smith L. Monte Carlo SSA: detecting irregular oscillations in the presence of coloured noise // Journal of Climate. — 1996. — Vol. 9. — P. 3373–3404.
- [2] Broomhead D., King G. Extracting qualitative dynamics from experimental data // Physica D: Nonlinear Phenomena. — 1986. — Vol. 20, no. 2–3. — P. 217–236.
- [3] Golyandina N., Nekrutkin V., Zhigljavsky A. Analysis of Time Series Structure: SSA and Related Techniques. — Chapman and Hall/CRC, 2001.
- [4] Alexandrov Th. A method of trend extraction using singular spectrum analysis // RevStat. — 2009 — Vol. 7 no 1. — P. 1–22. . — 2004. — Vol. 7-80. — P. 54–61.
- [5] Kalantari M., Hassani. H. Automatic grouping in singular spectrum analysis // Forecasting. — 2019. — Vol. 1, no 1. — P. 189–204.
- [6] Bogalo J., Poncela P., Senra E. Circulant singular spectrum analysis: A new automated procedure for signal extraction // Signal Processing. — 2021. — Vol. 179.
- [7] Golyandina N., Dudnik P., Shlemov A. Intelligent Identification of Trend Components in Singular Spectrum Analysis // Algorithms. — 2023. — Vol. 16. — ID 353.
- [8] Golyandina N. Detection of signals by Monte Carlo singular spectrum analysis: multiple testing // Statistics and Its Interface. — 2023. — Vol. 16. no 1. — P. 147–157.