

# О способах выделения сигнала на основе критерия Monte Carlo SSA

Потешкин Е.П., СПбГУ, Санкт-Петербург egor.poteshkin@yandex.ru,  
Голяндина Н.Э., СПбГУ, Санкт-Петербург n.golyandina@spbu.ru

## Аннотация

Рассматриваются методы анализа сингулярного спектра (SSA) и Monte Carlo SSA (MC-SSA) для решения задач обнаружения и выделения сигналов во временных рядах. Предложены три подхода к восстановлению сигнала: адаптивный, полуадаптивный и метод с фиксированной проекцией. Для оценки частоты сигнала используется метод MC-SSA. Проведен численный эксперимент, сравнивающий точность восстановления при различных уровнях шума, типах сигнала и значениях параметра  $\delta$ , определяющего длину частотного интервала при отборе компонент. Результаты показывают, что полуадаптивный вариант является универсальным выбором, наиболее устойчивым к наличию умеренной амплитудной модуляции.

## Введение

Рассмотрим следующую модель:  $X = S + R$ , где  $X$  — 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]. Один из шагов SSA подразумевает визуальный анализ для определения компонент сигнала, поэтому возникает потребность в автоматизации этого шага, этой проблеме посвящены, например, работы [4, 5, 6, 7]. Целью работы является определение подходов к автоматическому выделению слабых сигналов, обнаруживаемых критерием MC-SSA, и их сравнение по точности их выделения.

## Метод SSA

### *Базовый алгоритм*

Пусть  $X = (x_1, \dots, x_N)$ ,  $x_i \in \mathbb{R}$ , — временной ряд длины  $N$ . Зафиксируем параметр  $L$ ,  $1 < L < N$ , называемый длиной окна и построим так называемую траекторную матрицу  $X = [X_1 : \dots : X_K]$ , состоящую из  $K = N - L + 1$  векторов вложения  $X_i = (x_i, \dots, x_{i+L-1})^T \in \mathbb{R}^L$ .

Следующий шаг — разложение в сумму матриц единичного ранга  $X = \sum_{i=1}^d X_i$ . В базовом SSA используется сингулярное разложение матрицы  $X$ , где столбцы  $X_i$  состоят из проекций столбцов матрицы  $X$  на порождаемые ею самой левые сингулярные векторы.

Далее компоненты полученного матричного разложения группируются на основе свойств левых сингулярных векторов, и каждая сгруппированная матрица преобразуется во временной ряд. Таким образом, результатом SSA является разложение временного ряда.

### *SSA с проекцией*

Метод SSA использует адаптивный базис, но существует возможность зафиксировать некоторые компоненты разложения. Пусть  $D \in \mathbb{R}^{L \times m}$  — матрица, проекцию на столбцы которой мы хотим зафиксировать в разложении  $X$ . Тогда SSA с проекцией отличается от базового алгоритма только шагом разложения:

1. В случае, если столбцы матрицы  $D$  не ортонормированны,  $D$  приводится к нужному виду путем ортогонализации Грамма-Шмидта.
2. Вычисляется матрица  $C = DD^T X$ .
3. Вычисляется матрица  $X^* = X - C$ .
4. Матрица  $X^*$  раскладывается в сумму матриц ранга 1.

## Метод Monte Carlo SSA

Рассмотрим задачу поиска сигнала во временном ряде. Модель временного ряда имеет вид

$$X = S + \xi,$$

где  $S$  — сигнал,  $\xi$  — стационарный процесс с нулевым средним. Тогда нулевая гипотеза  $H_0 : S = 0$  и альтернатива  $H_1 : S \neq 0$ .

Зафиксируем длину окна  $L$  и обозначим траекторную матрицу ряда  $\xi$  как  $\Xi$ . Рассмотрим вектор  $W \in \mathbb{R}^L$  единичной длины, называемый проекционным вектором. Введем величину

$$p = \|\Xi^T W\|^2.$$

Статистикой критерия является величина

$$\hat{p} = \|\mathbf{X}^T W\|^2.$$

Распределение статистики критерия оценивается с помощью моделирования согласно нулевой гипотезе, отсюда и название метода.

Если вектор  $W$  — синусоида с частотой  $\omega$ , то  $\hat{p}$  отражает вклад частоты  $\omega$  в исходный ряд. Так как частота ожидаемого сигнала неизвестна, то необходимо рассматривать несколько векторов  $W_k$ ,  $k = 1, \dots, H$ . Решение возникающей при этом проблемы множественного тестирования рассматривается в [8]. Гипотеза об отсутствии сигнала отвергается, если хотя бы для одного вектора  $W = W_k$  значение  $\hat{p}$  оказывается значимым.

Важной частью метода MC-SSA является способ выбора векторов  $W_k$ . В данной работе в качестве векторов для проекции берутся косинусы с равноотстоящими частотами  $\omega_k = k/(2L)$ ,  $k = 1, \dots, L$ . В этом случае можно говорить о значимых частотах, присутствующих в сигнале.

## Подходы к выделению сигнала

Для ряда  $X$  длины  $N$  и  $0 \leq \omega_1 \leq \omega_2 \leq 0.5$  определим меру, следуя [4]

$$T(X; \omega_1, \omega_2) = \frac{1}{\|X\|^2} \sum_{k: \omega_1 \leq k/N \leq \omega_2} I_N(k/N),$$

где  $I_N$  — периодограмма ряда  $X$ . Величину  $T(X, \omega_1, \omega_2)$  можно рассматривать как долю вклада частот, содержащегося в интервале  $[\omega_1, \omega_2]$ .

В данной работе будем считать, что сигнал представляет из себя экспоненциально-модулированную гармонику:

$$S = \{Ae^{\alpha n} \cos(2\pi\omega n)\}_{n=1}^N,$$

где  $\omega \in (0, 0.5)$ . Пусть  $\hat{\omega}$  — оценка  $\omega$ . Обозначим

$$D_1 = \begin{pmatrix} \cos(2\pi\hat{\omega}1) \\ \dots \\ \cos(2\pi\hat{\omega}L) \end{pmatrix}, D_2 = \begin{pmatrix} \sin(2\pi\hat{\omega}1) \\ \dots \\ \sin(2\pi\hat{\omega}L) \end{pmatrix} \in \mathbb{R}^L.$$

Рассмотрим следующие варианты выделения сигнала  $S$  по частоте  $\hat{\omega}$ :

1. «adaptive»: применить SSA и выбрать первые две компоненты разложения, у которых мера  $T$  на интервале  $[\hat{\omega} - \delta, \hat{\omega} + \delta]$ ,  $\delta > 0$ , больше некоторого порога  $T_0 \in [0, 1]$ ;
2. «semi-adaptive»: применить SSA с проекцией с  $\mathbf{D} = D_1 \in \mathbb{R}^{L \times 1}$  и выбрать, помимо компоненты, соответствующей вектору  $D_1$ , первую компоненту разложения, у которой мера  $T$  на интервале  $[\hat{\omega} - \delta, \hat{\omega} + \delta]$ ,  $\delta > 0$ , больше некоторого порога  $T_0 \in [0, 1]$ ;
3. «fixed»: применить SSA с проекцией с  $\mathbf{D} = [D_1 : D_2] \in \mathbb{R}^{L \times 2}$  и выбрать компоненты разложения, соответствующие векторам  $D_1, D_2$ .

Оценивать частоту  $\omega$  будем с помощью MC-SSA:

1. Найти индекс наиболее значимой частоты, т.е.  $k = \operatorname{argmax}_i (\hat{p}_i - c_i)$ , где  $c_i$  — верхняя граница доверительного интервала для  $\hat{p}_i$ ;
2. Вычислить значение  $\hat{\omega}$  как взвешенное среднее частот  $\omega_{k-1}, \omega_k, \omega_{k+1}$  с весами  $w_i = \max(0, \hat{p}_i - c_i)$ ;

Такой способ оценки позволяет получить более точную оценку  $\omega$  в случае, когда она не попадает в решетку  $k/(2L)$ .

### ***Численное сравнение подходов***

Проведем численный эксперимент с целью понять, какой из предложенных способов восстановления сигнала наиболее точен. Пусть  $N = 99$ , процесс  $\xi$  — модель AR(1) с параметрами  $\phi = 0.7$ ,  $\sigma^2 \in \{0.2, 0.4, 0.6, 0.8, 1\}$ . Для SSA  $L = 50$ , для MC-SSA  $L = \tilde{L} = 40$  (выводы устойчивы к выбору длины окна). В вариантах «adaptive» и «semi-adaptive»  $\delta = 0.025$  и  $T_0 = 0.5$ . Рассмотрим два типа сигнала  $S$ , один из которых является частным случаем другого:

1.  $\alpha = 0$ ,  $A = 1$  — гармоника с постоянной амплитудой.

2.  $\alpha = 0.05$ ,  $A = 0.025$  — экспоненциально-модулированная гармоника.

Возьмем  $\omega = 0.115$ . Заметим, что при таком выборе частоты сигнала  $\tilde{L}\omega$  не целое, а значит  $\omega$  не попадает в решетку  $k/(2\tilde{L})$ .

На рис. 1 изображена зависимость MSE восстановления сигнала от дисперсии белого шума  $\sigma^2$ . По графикам видно, что в случае постоянной амплитуды ( $\alpha = 0$ ) выигрывает вариант «fixed», однако в случае непостоянной амплитуды фиксированный базис оказывается наихудшим. Полуадаптивный базис, являясь неким компромиссом между адаптивным и фиксированным базисами, оказывается вторым по точности в случае  $\alpha = 0$  и сравнимым с адаптивным в рассмотренном случае  $\alpha \neq 0$ . При увеличении  $|\alpha|$ , начиная с какого-то момента, фиксированная половина базиса ухудшает восстановление сигнала.

Теперь посмотрим, как будут изменяться ошибки при уменьшении/увеличении  $\delta$  для фиксированного  $\tilde{L}$ . На рис. 2  $\delta$  уменьшена, а на рис. 3 увеличена в два раза ( $\delta = 0.0125$  и  $0.05$  соответственно). Из этих графиков видно, что слишком маленькое  $\delta$  приводит к ухудшению точности адаптивного и полуадаптивного вариантов в случае  $\alpha \neq 0$ . Связано это с тем, что частота экспоненциально-модулированной гармоники, в отличие от гармоники с постоянной амплитудой, всегда растекается по спектру и чем больше абсолютное значение показателя экспоненты  $\alpha$ , тем сильнее это растекание. Увеличение  $\delta$  в два раза не привело к значительному изменению точности методов, однако, если и дальше увеличивать  $\delta$ , ошибки, как и в случае слишком маленького  $\delta$ , опять возрастут.

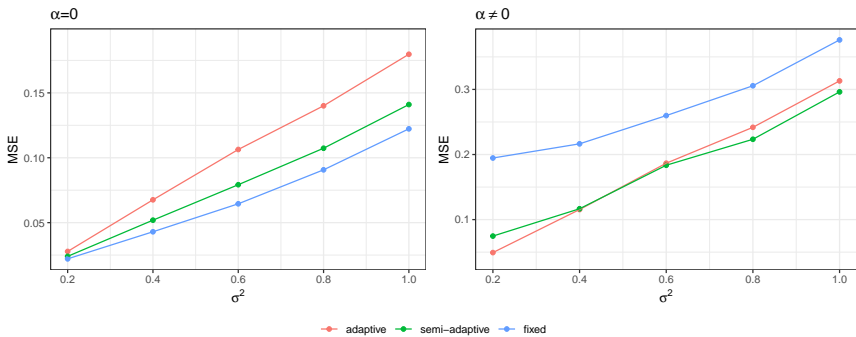
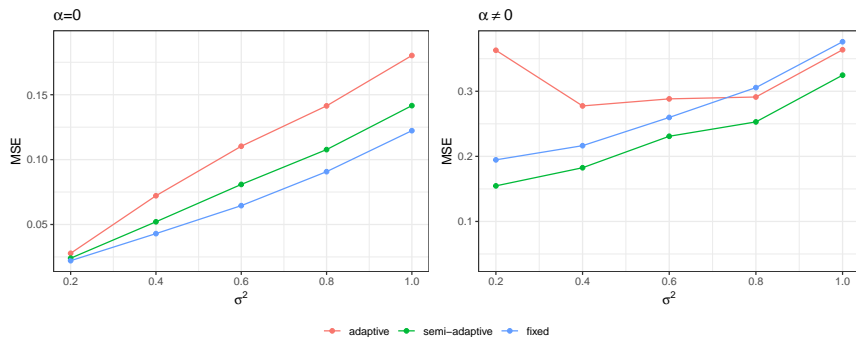
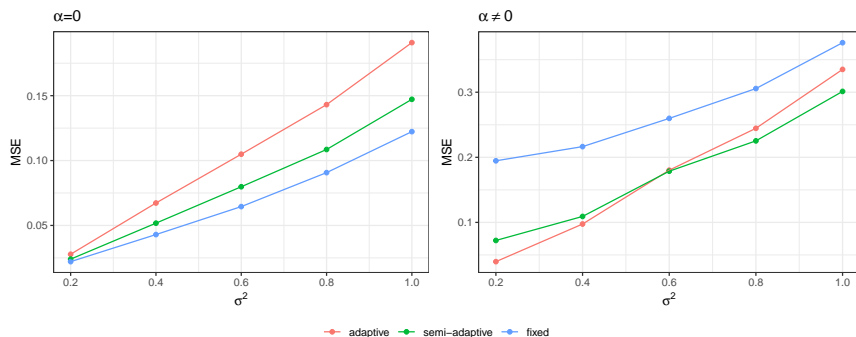


Рис. 1: MSE восстановления сигнала ( $\delta = 0.025$ )

Рис. 2: MSE восстановления сигнала ( $\delta = 0.0125$ )Рис. 3: MSE восстановления сигнала ( $\delta = 0.05$ )

## Заключение

В статье рассмотрены три подхода к автоматическому выделению сигнала во временных рядах с использованием критерия MC-SSA и метода анализа сингулярного спектра. Исследование показало, что полуадаптивный вариант может быть использован в качестве базового в ситуации, когда неизвестно наличие или отсутствие амплитудной модуляции, при этом модуляция не очень сильная. Для выбора параметра  $\delta$  нужно делать предположения о силе модуляции. Заметим, что при выборе длины окна для MC-SSA нужно учитывать сочетание мощности критерия и точности оценивания частоты.

### 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] Александров Ф., Голяндина Н. Автоматизация выделения трендовых и периодических составляющих временного ряда в рамках метода «Гусеница»-SSA // *Exponenta Pro (Математика в приложениях)*. — 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.