

Идентификация компонент временных рядов в методе SSA

Дудник Павел Дмитриевич, 23М.03-мм

Санкт-Петербургский государственный университет
Математико-механический факультет
Кафедра статистического моделирования

Научный руководитель — д.ф.-м.н. **Н.Э. Голяндина**
Рецензент — старший программист **А.Ю. Шлемов**, Майкрософт
Швейцария

Санкт-Петербург
2025г.

Введение: Постановка задачи

$X_N = (x_1, \dots, x_N)$ — временной ряд длины N , $x_i \in \mathbb{R}$ — наблюдение в момент времени i .

$X_N = X_{\text{Trend}} + X_{\text{Periodics}} + X_{\text{Noise}}$, — временной ряд, где

- $X_{\text{Trend}} + X_{\text{Periodics}}$ — сигнал
- X_{Trend} — тренд, медленно меняющаяся компонента
- $X_{\text{Periodics}}$ — сумма периодических компонент
- X_{Noise} — шум, случайная составляющая

Метод: Анализ сингулярного спектра (Singular spectrum analysis, SSA) — непараметрический метод, позволяющий раскладывать временной ряд в сумму тренд+периодичность+шум.

(Golyandina, N., Nekrutkin, V., Zhigljavsky, A., 2001, Analysis of Time Series Structure: SSA and Related Techniques)

Задача: Придумать и реализовать алгоритм, который автоматически идентифицирует компоненты сигнала и строит полное разложение в рамках метода SSA.

$X_N = (x_1, \dots, x_N)$ — временной ряд, N — его длина. .

Алгоритм SSA для идентификации тренда и периодик (параметр $1 < L < N$ — длина окна, наборы индексов $\{i_1, \dots, i_p\}, \{j_1, \dots, j_k\}$):

- 1 **Построение траекторной матрицы:** $K = N - L + 1$, ряд переводится в траекторную матрицу $\mathbf{X} = \mathcal{T}(X_N) = [X_1 : \dots : X_K]$, где $X_i = (x_i, \dots, x_{i+L-1})^T$, $1 \leq i \leq K$.
- 2 **Сингулярное разложение (SVD)** траекторной матрицы:
$$\mathbf{X} = \sum_{i=1}^d \sqrt{\lambda_i} U_i V_i^T = \sum_{i=1}^d \mathbf{X}_i, \quad d = \text{rank } \mathbf{X}.$$
 Компоненты SVD \mathbf{X}_i — элементарные матрицы ранга 1.
- 3 **Идентификация трендовой и периодической компоненты** с номерами $I = \{i_1, \dots, i_p\}$ и $J = \{j_1, \dots, j_k\}$.
- 4 **Получение оценки тренда и периодики** $\tilde{X}_{Trend} = \sum_{i \in I} \tilde{X}_i$, $\tilde{X}_{Periodics} = \sum_{j \in J} \tilde{X}_j$, где $\tilde{X}_i = \mathcal{T}^{-1}(\mathcal{H}(\mathbf{X}_i))$ — элементарные временные ряды.

Этап идентификации компонент:

- В базовом варианте идентификация визуальная.
- Есть алгоритмы автоматической идентификации.

Проблема: Идентификация компонент сигнала возможна, только если в разложении трендовые компоненты не смешались с периодическими (если есть *разделимость*).

Решение: совместить методы улучшения разделимости и автоматическую идентификацию компонент.

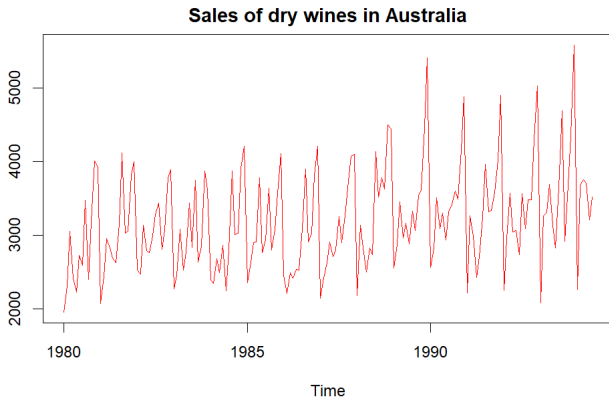


Рис.: График продаж сухих вин в Австралии по месяцам

Введение: Пример применения метода SSA

Выделение тренда — идентификация медленно-меняющихся компонент и их суммирование.

Выделение периодик — идентификация быстро меняющихся компонент, имеющих период.

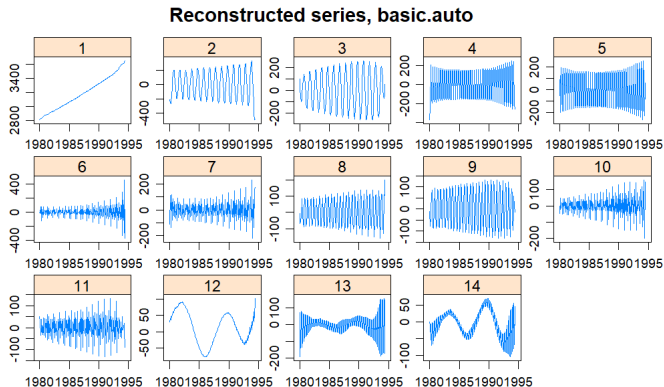


Рис.: Разложение в сумму элементарных временных рядов.

Введение: Пример применения метода SSA

Если компоненты смешались, то их нельзя корректно идентифицировать.

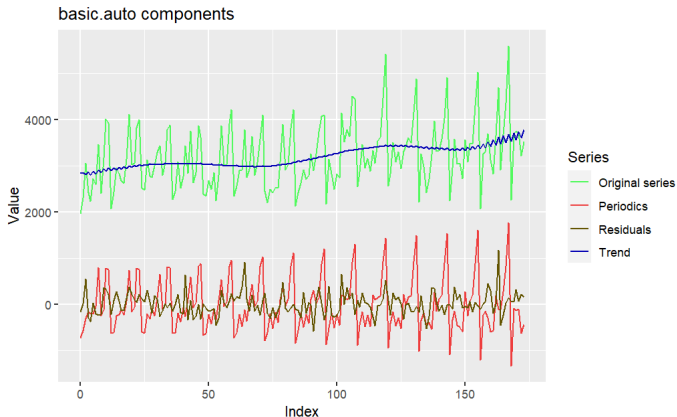


Рис.: Идентифицированные компоненты

Схема алгоритма автоматической идентификации компонент

Схема предлагаемого алгоритма:

- ❶ Улучшение разделимости
 - ❷ Автоматическая идентификация тренда
 - ❸ Автоматическая идентификация периодик после выделения тренда
 - ❹ «Отсеивание» ложных периодик
- } Бак. работа
-
- Для **улучшения разделимости** и автоматической идентификации **тренда** использовался метод **autoEOSSA**.
 - Для автоматической идентификации **периодик** используем **модифицированный метод регулярности углов**.
 - Для **«отсеивания»** ложных периодик используется **модифицированный критерий MC-SSA**.

Улучшение разделимости и автоматическая идентификация тренда

Для автоматической идентификации тренда используется алгоритм $\text{autoEOSSA}(t, \omega_0)$, представленный в [1].

Подход: улучшение разделимости (EOSSA) и автоматическая группировка трендовых компонент.

Параметры:

- t — количество элементарных компонент, отнесенных к сигналу.
- ω_0 — порог трендовых частот.
- T_0 — порог меры трендовых частот.

Результаты [1] показали, что алгоритм устойчив к превышению ранга сигнала (когда оценка t больше истинного ранга r).

[1] autoEOSSA — Golyandina, Dudnik, Shlemov 2023

Метод регулярности углов (Жорникова 2016):

- **Что делает метод?** Метод идентифицирует э-м гармоники с частотами $0 < \omega < \frac{1}{2}$.
- **Вход и параметры.** Левые сингулярные векторы траекторной матрицы. Порог τ_0 .
- **Идея метода.** Метод использует факт, что углы между последовательными отрезками на двумерной диаграмме сингулярных векторов равны.
- **Что было дополнено.** Метод расширен на случай $\omega = \frac{1}{2}$. К параметрам добавился порог p_0 . Используется регулярность смены знака компоненты.

Критерий **Multiple MC-SSA** (Golyandina 2020):

- **Что делает метод?** Проверяется гипотеза об отсутствии сигнала в белом или красном шуме.
- **Вход и параметры.** Временной ряд, длина окна, базис, уровень значимости α .
- **Идея метода.** Для вычисления статистики критерия траекторная матрица временного ряда домножается на базис. Критерий контролирует групповую ошибку.
- **Что было дополнено.** Критерий модифицирован на случай проверки гипотезы для ряда с сигналом, из которого была извлечена компонента с некоторым диапазоном частот. Для оценки параметров используется модификация правдоподобия Whittle.

Полученные результаты. Алгоритм SSA-Aid

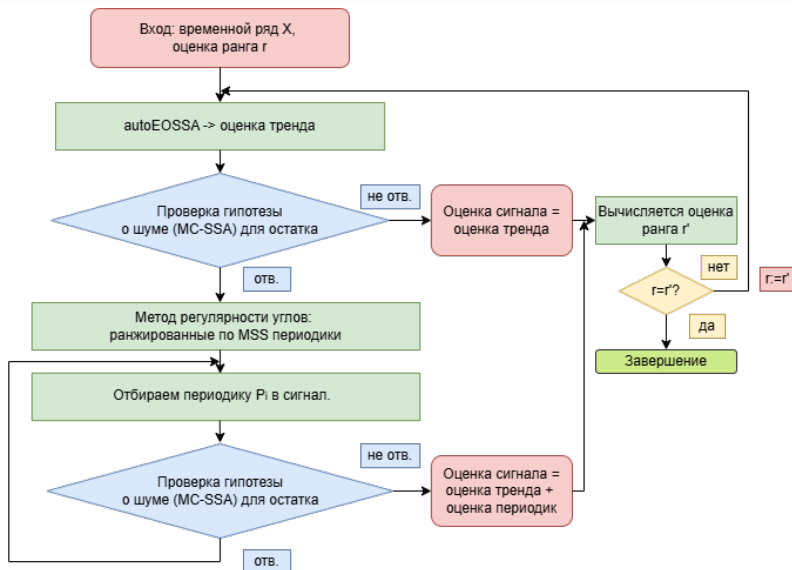


Рис.: Схема алгоритма SSA-Aid

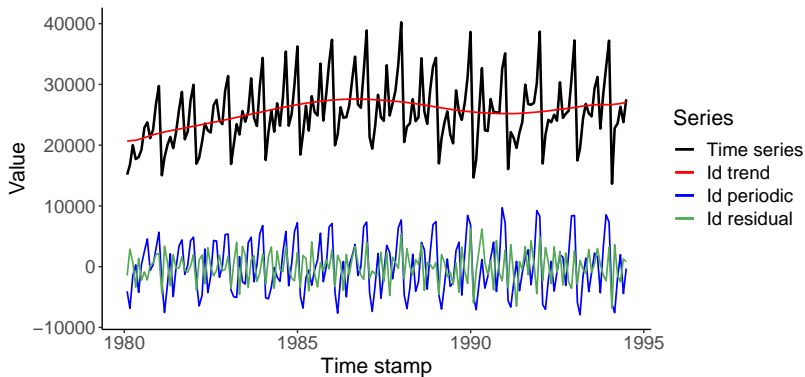
Снова рассмотрим временной ряд продажи австралийских вин, применим алгоритм к ряду. Длина ряда $N = 174$. Параметры:

- Длина окна $L = 84$.
- Порог низких частот $\omega_0 = \frac{1}{24}$.

Параметры по умолчанию:

- Число эл. компонент, отнесенных к сигналу $t = 30$.
- Порог меры низких частот $T_0 = 0.5$.
- Порог для меры регулярности углов: $\tau_0 = 0.05$.
- Порог для меры чередования знаков $p_0 = 0.03$.
- Уровень значимости $\alpha = 0.001$.
- Модель красного шума.

Полученные результаты. SSA-Aid



Индексы идентифицированных компонент: Тренд: 1 – 3.

Исходно идентифицированные
периодические компоненты:
4 – 16, 19 – 30.

Периодики после отсеивания:
4 – 14

Периоды: 2, 2.4, 3, 4, 6, 12.

Численное сравнение SSA-Aid с методами оценки ранга сигнала

- Информационные критерии для оценки ранга сигнала **TRMAT**, **MDL** (Golyandina N., Zvonarev N., 2024).
- В большинстве случаев результаты SSA-Aid сравнимы по точности и оценке ранга с методами.

Пример: рассматривается временной ряд длины $N = 200$:

- Тренд $t_n = 0.2e^{0.05n} + 2 \cos\left(\frac{2\pi n}{60}\right)$
- Периодика $s_n = 4.12 \cos\left(\frac{2\pi n}{4}\right) + 4.12 \cos\left(\frac{2\pi n}{3}\right) + 6 \cos\left(\frac{2\pi n}{2}\right)$
- Белый стандартный гауссовский шум

Таблица: MSE выделения сигнала

	Mean MSE	Med MSE	Процент неверной оценки ранга	
			Переоценка	Недооценка
Rank TRMAT	0.243	0.209	5.5	0
Rank MDL	0.218	0.207	2.4	0
SSA-Aid	0.217	0.207	3.4	0

Численное сравнение SSA-Aid с методами декомпозиции и прогнозирования

Сравнение с методами **STL** (Cleveland et al., 1990), **ETS** (Hyndman et al., 2008), **TBATS** (De Livera et al., 2011), **Prophet** (Taylor et al., 2018).

- STL: декомпозиция через LOESS
- ETS: экспоненциальное сглаживание
- TBATS: расширение ETS для сложных сезонностей и авторегрессионных шумов
- Prophet: кусочно-линейное моделирование тренда, сезонность через Фурье-ряды

Пример 1:

- Тренд $t_n = 0.2e^{0.005n}$
- Периодика
 $s_n = 4.12 \cos\left(\frac{2\pi n}{4}\right)$

Пример 2:

- Тренд
 $t_n = 0.2e^{0.005n} + 2 \cos\left(\frac{2\pi n}{60}\right)$
- Периодика
 $s_n = 4e^{-0.002n} \cos\left(\frac{2\pi n}{4}\right) + 4 \cos\left(\frac{2\pi n}{3}\right) + 6e^{-0.002n} \cos\left(\frac{2\pi n}{2}\right)$

Используется красный шум $\phi = 0.3, \sigma = 1$, длина ряда $N = 1000$.

Численное сравнение SSA-Aid с методами декомпозиции, выделение тренда

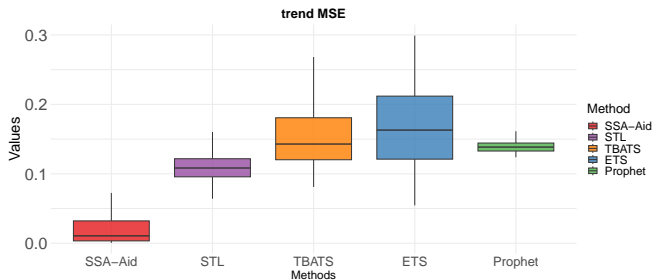


Рис.: MSE выделения тренда. Пример 1.

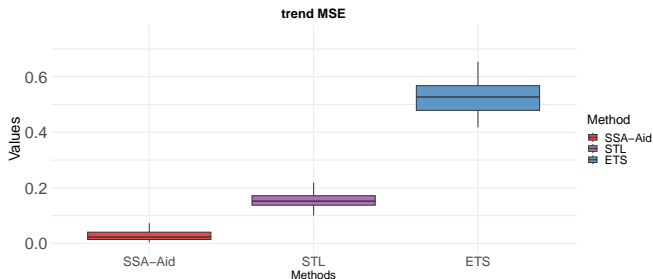


Рис.: MSE выделения тренда. Пример 2.

Численное сравнение SSA-Aid с методами декомпозиции, выделение сезонности

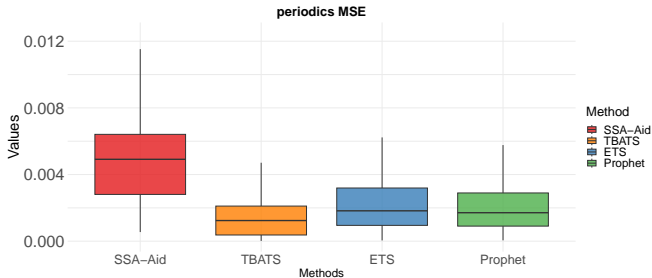


Рис.: MSE выделения сезонности. Пример 1.

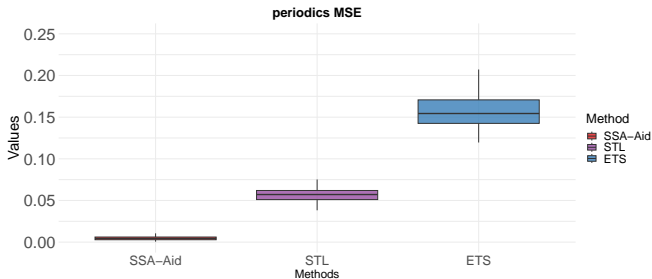


Рис.: MSE выделения сезонности. Пример 2.

Численное сравнение SSA-Aid с методами прогнозирования

Сравнение с методами **ETS**, **TBATS**, **Prophet**, **SARIMA** (Box et. al. 2015). Рассмотрим те же примеры.

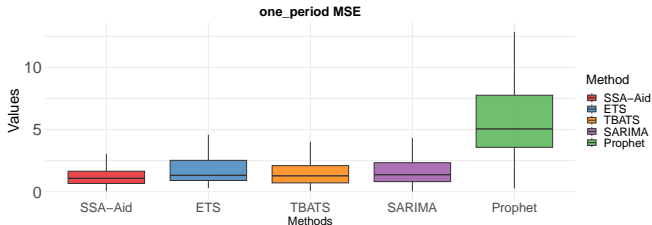


Рис.: MSE прогнозирования на один период.
Пример 1.

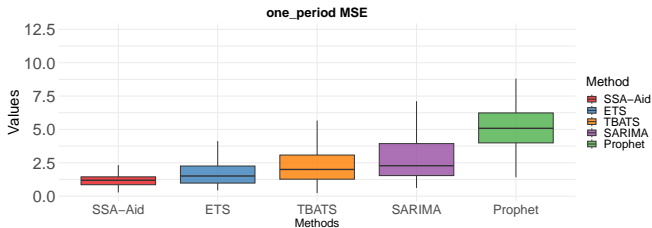


Рис.: MSE прогнозирования на один период.
Пример 2.

- Разработан алгоритм SSA-Aid автоматической идентификации компонент и построения разложения временного ряда на тренд, периодичность и шум для случаев белого и красного шума.
- Все шаги алгоритма исследованы, подобраны базовые параметры.
- Проведено численное сравнение точности выделения сигнала с методами оценки ранга: SSA-Aid сравним по точности с методами.
- Проведено численное сравнение SSA-Aid с методами декомпозиции и прогнозирования для разных типов сигналов.

Предложенный алгоритм и модификации реализованы в виде набора программ на языке R и выложены в публичный доступ.