

Применение метода SSA для анализа и прогноза параметров вращения Земли

Оленев Роман Олегович, 23.M03-мм

Санкт-Петербургский Государственный Университет
Математико-механический факультет
Кафедра статистического моделирования

Научный руководитель — д.ф.-м.н. **Н.Э. Голяндина**
Рецензент — главный научный сотрудник, д.ф.-м.н. **З.М. Малкин**, Главная
(Пулковская) астрономическая обсерватория Российской академии наук

Санкт-Петербург
2025г.

Дано: Временные ряды параметров вращения Земли (ПВЗ).

Цель: Сравнить точность различных методов прогнозирования и заполнения пропусков для временных рядов параметров вращения Земли.

Мотивация: Прогнозы временных рядов ПВЗ используются для расчёта орбит спутников, в том числе спутников GPS. Навигация космических миссий, точное наведение астрономических приборов и позиционирование спутников на Земле в значительной степени зависят от точных прогнозов ПВЗ.

Параметры вращения Земли (ПВЗ) — одномерные временные ряды, описывающие положение Земли в пространстве:

- координаты полюса X, Y (движение полюса);
- долгота суток LOD (угловая скорость вращения);
- отклонение небесного полюса dX, dY .

Наблюдением ПВЗ занимается Международная служба вращения Земли (IERS, International Earth Rotation and Reference Systems Service).

Параметры движения полюса

Параметры X , Y являются координатами полюса в земной системе координат. Полюс находится в точке поверхности Земли, через которую проходит мгновенная ось вращения планеты.

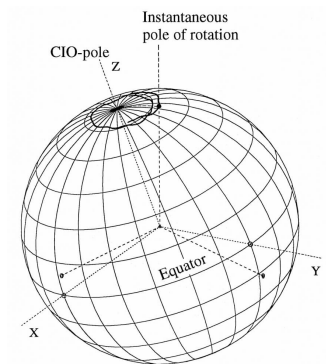


Рис.: Движение полюса.

Временные ряды координат полюса X , Y состоят из тренда и двух основных периодик — годового и чандеровского колебаний.

Данные были взяты из бюллетеня EOP 14 C04 (IAU2000A).

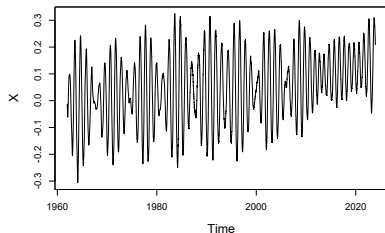


Рис.: Временной ряд X .

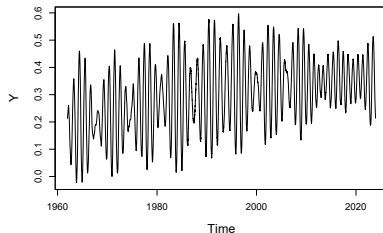


Рис.: Временной ряд Y .

Структура сигнала рядов координат полюса:

$$S = P_a + P_c + T,$$

где P_a , P_c — годовая и чандлеровская периодики, T — тренд.

Периоды (в днях): $P_a \approx 365$, $P_c \approx 433$.

Методы для временных рядов с несколькими периодическими компонентами:

- МНК: $s_n = b_0 + b_1 n + b_2 \cos\left(\frac{2\pi n}{P_a} + \phi_a\right) + b_3 \cos\left(\frac{2\pi n}{P_c} + \phi_c\right)$;
- MSTL (Bandura et al., 2021);
- MSARIMA (Svetunkov, 2023);
- Singular spectrum analysis (SSA)
(Golyandina, Nekrutkin, Zhigljavsky, Analysis of Time Series Structure: SSA and Related Techniques, 2001).

Данная работа посвящена применению метода SSA для прогноза и заполнения пропусков временных рядов координат полюса X , Y .

SSA для прогноза X , Y : Kong et. al, 2023; Jin et. al, 2021; Modiri et. al, 2018; Okhotnikov, Golyandina, 2019.

В данной работе рассмотрены две новые задачи:

- Прогнозирование методом CSSA (Complex SSA).
- Заполнение исторических пропусков во временном ряде Y .

Выделение сигнала методом SSA

Дан временной ряд $X = (x_1, \dots, x_N)$.

Алгоритм SSA (1D-SSA) для выделения сигнала
(параметры $1 < L < N$ — длина окна, $1 < r < L$).

1 Построение траекторной матрицы

$$\mathcal{T}(X) = \mathbf{X} = [X_1 : \dots : X_K] = \begin{pmatrix} x_1 & x_2 & \dots & x_K \\ x_2 & \ddots & \ddots & x_{K+1} \\ \vdots & \ddots & \vdots & \vdots \\ x_L & x_{L+1} & \dots & x_N \end{pmatrix}$$

2 Сингулярное разложение траекторной матрицы

$$\begin{aligned} \mathbf{X} &= \mathbf{X}_1 + \dots + \mathbf{X}_d, \\ \mathbf{X}_i &= \sqrt{\lambda_i} U_i V_i^T, \quad i = 1, \dots, d. \end{aligned}$$

3 Получение оценки сигнала

$$\tilde{X} = \mathcal{T}^{-1}(\mathcal{H}(\tilde{\mathbf{X}})), \quad \tilde{\mathbf{X}} = \sum_{i=1}^r \mathbf{X}_i.$$

Вариации метода SSA: 1D-SSA, MSSA, CSSA (Complex SSA).

MSSA: многомерные временные ряды (X, Y) .

CSSA: комплексные временные ряды $X + iY$, оператор транспонирования T заменяется на оператор эрмитова сопряжения H .

$$p_n^{(1)} = A_1 \cdot \cos(2\pi\omega n + \varphi_1), p_n^{(2)} = A_2 \cdot \sin(2\pi\omega n + \varphi_2), 0 < \omega < 0.5.$$

1D-SSA: ранг P_1, P_2 равен 2.

MSSA: ранг (P_1, P_2) равен 2.

CSSA: если $A_1 = A_2, \varphi_1 - \varphi_2 \equiv 0 \pmod{\pi}$, то ранг $P_1 + iP_2$ равен 1.

$$\text{rank}_L(X) \stackrel{\text{def}}{=} \text{rank } X.$$

Дан временной ряд $X = (x_1, \dots, x_N)$. Требуется получить прогноз ряда на h точек $(\tilde{x}_{N+1}, \dots, \tilde{x}_{N+h}) = \tilde{X}_{N+1, N+h}$.

Виды прогнозов:

- Краткосрочные (short-term).
- Долгосрочные (long-term). Устойчивость прогноза играет бóльшую роль при долгосрочном прогнозировании.

Алгоритм рекуррентного прогнозирования

Параметры: L, r .

$U_i \in \mathbb{R}^L, i = 1, \dots, r$ — левые сингулярные векторы траекторной матрицы \mathbf{X} , \underline{U}_i — первые $L - 1$ координат и π_i — последняя координата.

Определим $(a_{L-1}, \dots, a_1)^T = \sum_{i=1}^r \pi_i \underline{U}_i / (1 - \sum_i \pi_i^2)$.

Алгоритм рекуррентного прогнозирования:

1 Ряд $Y_{1, N+h} = (y_1, \dots, y_{N+h})$ определяется как

$$y_i = \begin{cases} \tilde{x}_i, & i = 1, \dots, N, \\ \sum_{j=1}^{L-1} a_j y_{i-j}, & i = N + 1, \dots, N + h, \end{cases}$$

2 числа y_{N+1}, \dots, y_{N+h} являются результатом прогнозирования ряда X на h точек.

Задача: Сравнить метод 1D-SSA с CSSA при прогнозировании рядов X , Y , а также исследовать применимость одношаговой валидации для подбора числа компонент.

Метод оценивания точности: MAE (mean absolute error).

$$\text{MAE}(X_{N+1, N+h}, \tilde{X}_{N+1, N+h}) = \frac{1}{h} \sum_{i=1}^h |x_{N+i} - \tilde{x}_{N+i}|.$$

Способ выбора r :

- r при котором ошибка $\text{MAE}(X_{N-h+1, N}, \tilde{X}_{N-h+1, N})$ минимальна.
- r при котором средняя ошибка по нескольким прогнозам на всем временном ряде минимальна.

Прогноз на 500 точек: 94 прогноза, с 01.01.1995 до 01.01.2021 со сдвигом в 100 дней.

Прогноз на 30 точек: 316 прогнозов, с 01.01.1995 до 01.01.2021 со сдвигом в 30 дней.

Выбор N : 6 лет, 12 лет, 19 лет, 31 год.

Выбор L : от 3 до 15 лет с шагом в 3 года.

Выводы по численным исследованиям:

- Прогноз с параметром L , равным 3 годам, и параметром N , равным 19 годам, оказался более точным при прогнозировании на 500 точек.
- Прогноз с параметром L , равным 3 годам, и параметром N , равным 31 году, оказался более точным при прогнозировании на 30 точек.
- Подход с выбором r , дающего минимальную среднюю ошибку, оказался предпочтительным.

Таблица: Средние MAE по прогнозам для 1D-SSA и CSSA с фиксированными r при прогнозировании на 500 и 30 точек.

MAE, Mean	X	Y
1D-SSA, $h = 500$	0.0201	0.0214
CSSA, $h = 500$	0.0200	0.0192
1D-SSA, $h = 30$	0.0050	0.0036
CSSA, $h = 30$	0.0048	0.0029

Вывод: метод CSSA более точный, особенно для Y .

Задача: Сравнить методы 1D-SSA, CSSA, MSTL, MSARIMA, и SSA + ARMA (Kong et al., 2023) при прогнозировании рядов X, Y .

Метод оценивания точности: MAE.

Были рассмотрены:

- Методы 1D-SSA и CSSA с параметром N , равным 19 годам, параметром L , равным 3 годам, $r = 100$.
- Метод MSTL с параметром $n_s = 5$, и метод MSARIMA(1, 0, 0)(0, 1, 0)₃₆₅(0, 1, 0)₄₃₃. Параметр $N = 31$ год.
- SSA + ARMA с параметрами из (Kong et al., 2023).

Было построено 105 прогнозов, начиная с 04.01.2019, со сдвигом в 7 дней. Проведено сравнение прогнозов при различных h .

Таблица: Средние MAE по прогнозам для ряда Y в зависимости от h .

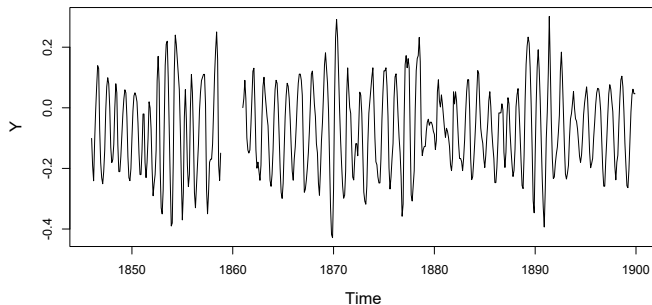
h	1D-SSA, $r = 100$	CSSA, $r = 100$	MSTL	MS ARIMA	SSA + ARMA
1	0.0012	0.0011	0.0092	0.0008	0.0003
2	0.0012	0.0012	0.0093	0.0012	
3	0.0013	0.0012	0.0093	0.0015	
4	0.0014	0.0013	0.0093	0.0018	
7	0.0017	0.0016	0.0094	0.0027	
15	0.0023	0.0022	0.0097	0.0044	
30	0.0034	0.0031	0.0102	0.0071	0.0043
60	0.0052	0.0042	0.0114	0.0117	0.0078
120	0.0086	0.0064	0.0130	0.0186	0.0132
240	0.0108	0.0087	0.0139	0.0234	0.0122
365	0.0117	0.0102	0.0151	0.0273	0.0142
420	0.0120	0.0109	0.0156	0.0289	0.0149
500	0.0124	0.0116	0.0166	0.0305	0.0156

Выводы:

- Методы MSARIMA и SSA+ARMA оказались более точными при краткосрочном прогнозировании.
- Начиная с $h = 3$, метод CSSA оказался более точным.

Заполнение исторических пропусков в ряде Y

Задача: Заполнить пропуски во временном ряде Y (1846–1899г.) из бюллетеня EOP C01.



Данные представлены с частотой, равной 1/10 года.

$P_a \approx 10$, $P_c \approx 11.9$.

Заполнение исторических пропусков в ряде Y

Были рассмотрены 2 метода заполнения пропусков:

- igarfill: пропуски итеративно заполняются значениями, полученными после выделения сигнала методом $SSA(L, r)$.
- Метод наименьших квадратов, модель сигнала Y_p :

$$\begin{aligned} y_n^p = & b_0 + b_{c_0}^a \cos\left(\frac{2\pi n}{P_a} + \phi_{c_0}^a\right) + b_{c_1}^a n \cos\left(\frac{2\pi n}{P_a} + \phi_{c_1}^a\right) \\ & + b_{c_0}^{c_1} \cos\left(\frac{2\pi n}{P_{c_1}} + \phi_{c_0}^{c_1}\right) + b_{c_1}^{c_1} n \cos\left(\frac{2\pi n}{P_{c_1}} + \phi_{c_1}^{c_1}\right) \\ & + b_{c_0}^{c_2} \cos\left(\frac{2\pi n}{P_{c_2}} + \phi_{c_0}^{c_2}\right) + b_{c_1}^{c_2} n \cos\left(\frac{2\pi n}{P_{c_2}} + \phi_{c_1}^{c_2}\right), \end{aligned}$$

$P_{c_1} = 1.181$, $P_{c_2} = 1.231$, $P_a = 1$ (в годах), применяется к G точкам вокруг пропусков.

Заполнение исторических пропусков в ряде Y

Для оценки качества заполнения был применен **алгоритм вычисления ошибки на искусственных пропусках**.

Вход : ряд Y с пропусками, длина интервала с искусственными пропусками M , индексы начала i_1, \dots, i_Q интервалов с искусственными пропусками, мера ошибки Err , r , L .

Выход: средняя ошибка заполнения \overline{Err} .

for $q \leftarrow 1$ **to** Q **do**

$Y_{i_q, i_q+M-1} \leftarrow \text{NA};$

$\tilde{Y} \leftarrow \text{igapfill}(Y, L, r);$

$Err_q \leftarrow Err(Y_{i_q, i_q+M-1}, \tilde{Y}_{i_q, i_q+M-1});$

return $\overline{Err} = \frac{1}{Q} \sum_{q=1}^Q Err_q;$

igapfill: Подбирается L , r .

МНК: Подбирается G (результат $G = N$, средняя MAE = 0.88).

Заполнение исторических пропусков в ряде Y

Таблица: Средние MAE по заполнениям для моделей порядка r на искусственных пропусках.

r	1D-SSA, L=119	CSSA, L=119	1D-SSA, L=238	CSSA, L=238	1D-SSA, L=60	CSSA, L=60
1	0.121	0.104	0.121	0.106	0.125	0.104
2	0.104	0.084	0.107	0.087	0.112	0.084
3	0.089	0.076	0.091	0.080	0.093	0.070
4	0.082	0.072	0.085	0.076	0.081	0.066
5	0.085	0.070	0.086	0.072	0.080	0.064
6	0.083	0.065	0.083	0.071	0.078	0.067
7	0.085	0.064	0.084	0.066	0.081	0.067
8	0.086	0.064	0.083	0.073	0.078	0.071
9	0.090	0.065	0.084	0.075	0.083	0.073
10	0.088	0.068	0.085	0.076	0.086	0.077
11	0.088	0.071	0.085	0.074	0.094	0.079
12	0.080	0.072	0.083	0.074	0.092	0.083

Заполнение исторических пропусков в ряде Y

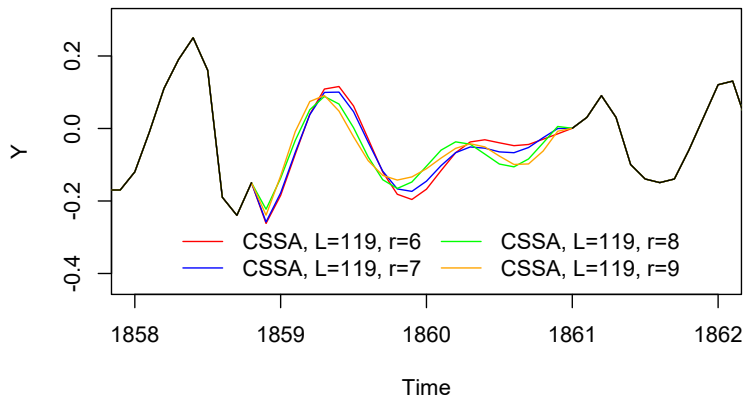


Рис.: Заполнения методом CSSA с $L = 119$ для $r \in \{6, 7, 8, 9\}$.

Выводы:

- CSSA с $L = 119$, $r \in \{6, 7, 8, 9\}$ — лучший среди вариантов SSA.
- CSSA точнее заполнения по параметрической модели Y_p (0.64 против 0.88).
- В качестве финального заполнения можно рекомендовать среднее из заполнений по лучшим моделям CSSA.

Прогноз X и Y :

- Прогноз с фиксированным r более точный, чем прогноз с одношаговой валидацией для подбора r .
- Для прогноза следует выбирать большое r .
- При краткосрочном прогнозировании методы MSARIMA и SSA+ARMA более точные.
- Метод CSSA более точный при среднесрочном и долгосрочном прогнозировании.

Заполнение исторических пропусков в Y :

- Метод CSSA более точный.
- Результаты приняты к публикации в научном журнале (Малкин, Голяндина, Оленев, 2025).

R-скрипты выложены в открытый доступ

<https://doi.org/10.5281/zenodo.15498356>.

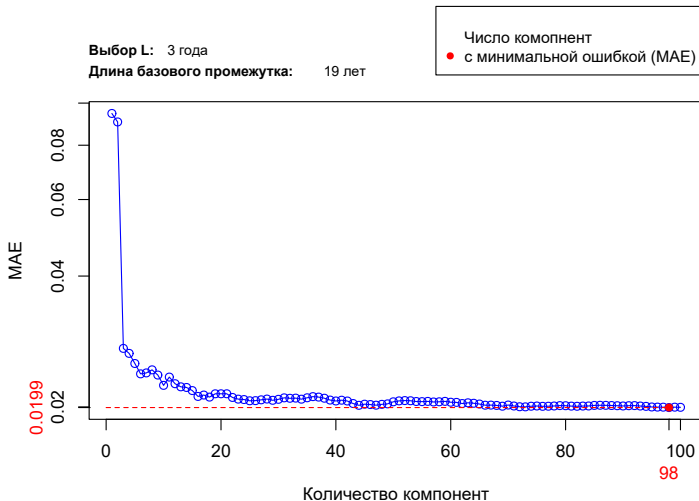


Рис.: CSSA, зависимость средней ошибки по промежуткам валидации от параметра r при прогнозе X на 500 точек.

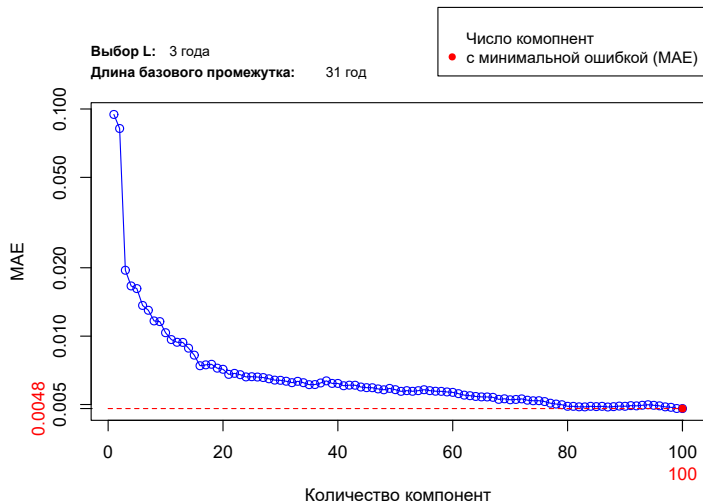


Рис.: CSSA, зависимость средней ошибки по промежуткам валидации от параметра r при прогнозе X на 30 точек.

Таблица: r , дающие минимальную среднюю ошибку по прогнозам.

r	X	Y
1D-SSA, $h = 500$	82	88
CSSA, $h = 500$	98	67
1D-SSA, $h = 30$	98	82
CSSA, $h = 30$	100	86