

# Применение метода анализа сингулярного спектра

Охотников Григорий Иванович

Санкт-Петербургский государственный университет  
Математико-механический факультет  
Статистическое моделирование

Научный руководитель — к.ф.-м.н. **Н. Э. Голяндина**  
Рецензент — к.ф.-м.н. **А. Н. Пепелышев**



Санкт-Петербург 2017

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

- координаты полюса  $x, y$  (движение полюса);
- долгота дня  $LOD$  (угловая скорость вращения);
- поправки координат небесного полюса  $dX, dY$  (движение оси вращения в пространстве — прецессия и нутация).

Прогнозы рядов ПВЗ используются для расчёта орбит спутников, в том числе спутников GPS.

Формализация задачи:

- Задан вещественный временной ряд  $\mathbb{X}_N = (x_1, \dots, x_N)$ ;
- требуется построить прогноз ряда  $\mathbb{X}_N$  на  $P$  шагов.

- Организовать сбор данных на ежедневной основе;
- применить метод анализа сингулярного спектра ([Singular Spectrum Analysis, SSA](#)) для анализа временных рядов параметров вращения Земли;
- разработать алгоритм подбора параметров метода SSA для более точного прогнозирования;
- провести сравнение с другими методами прогнозирования по точности на прошедших моментах времени;
- обеспечить регулярное обновление прогнозов рядов ПВЗ по методу SSA для текущего дня на странице в Интернете.

- Бюллетень IERS C04 содержит значения каждого из рядов ПВЗ (публикуется ежедневно, имеет запаздывание приблизительно в 30 дней);

([http://hpiers.obspm.fr/iers/eop/eopc04/eopc04\\_IAU2000.62-now](http://hpiers.obspm.fr/iers/eop/eopc04/eopc04_IAU2000.62-now))

- $x, y, LOD$  — с 1 января 1962 года;
- $dX, dY$  — с 1 января 1984 года;

- данные `finals2000A.daily` содержат оценки значений рядов  $x, y, LOD, dX, dY$  на прошедшие 90 дней, прогноз рядов  $x, y$  на следующие 90 дней и прогноз рядов  $dX, dY$  на 60 дней (публикуется ежедневно).

(<https://datacenter.iers.org/eop/-/somos/5Rgv/latest/13>)

- Бюллетень А Международной службы вращения Земли (International Earth Rotation and Reference Systems Service, IERS) содержит прогнозы рядов  $x, y, LOD$  на 365 дней (публикуется еженедельно);

(<https://datacenter.iers.org/eop/-/somos/5Rgv/latest/6>)

- прогнозы рядов  $x, y, LOD, dX, dY$  Пулковской обсерватории на 365 дней (публикуются ежедневно).

(<http://www.gao.spb.ru/english/as/persac/eopcxxx/>)

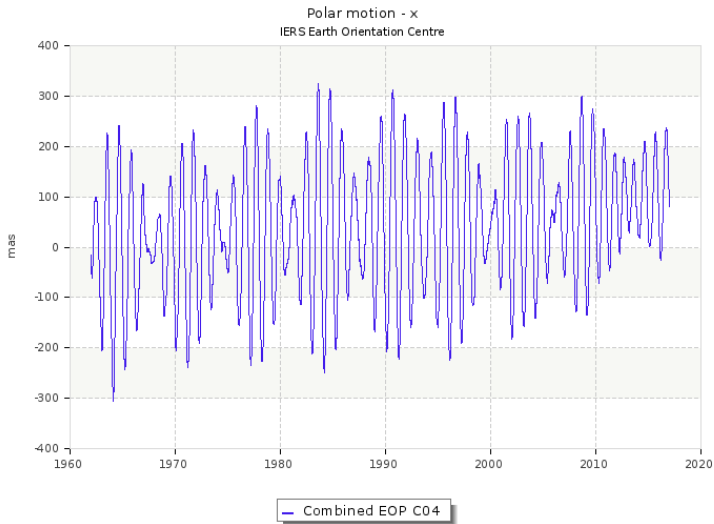


Рис. 1: График ряда  $x$  по ежедневным наблюдениям

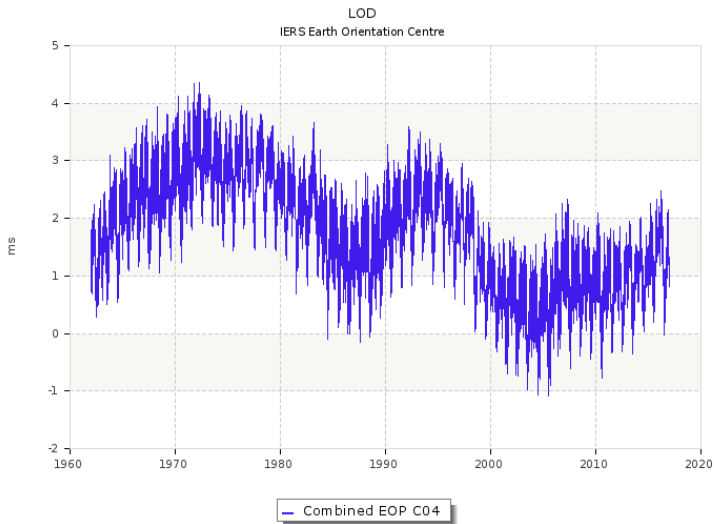


Рис. 2: График ряда  $LOD$  по ежедневным наблюдениям

# Поправки координаты небесного полюса $dX$

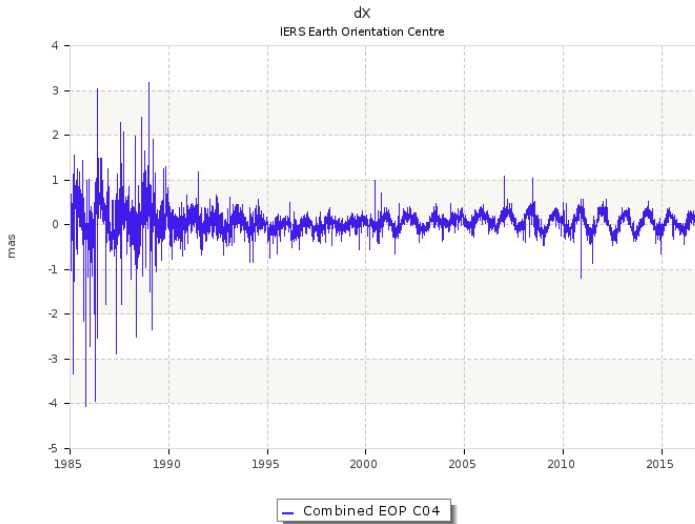


Рис. 3: График ряда  $dX$  по ежедневным наблюдениям



# Алгоритм метода SSA выделения сигнала в ряде

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

Входные данные:  $\mathbb{X}_N = (x_1, \dots, x_N)$  — временной ряд,  
параметры:  $1 < L < N$  — длина окна,  $1 < r < L$ .

- Вектора вложения  $\{X_i\}_{i=1}^K$ ,  $K = N - L + 1$ ,  
 $X_i = (x_{i-1}, \dots, x_{i+L-1})^T$ ;
- построение траекторной матрицы  $\mathbf{X} = [X_1 : \dots : X_K]$ ;
- SVD матрицы  $\mathbf{X} = \sum_{i=1}^L \sqrt{\lambda_i} U_i V_i^T$ ;
- аппроксимация сигнала  $\tilde{\mathbb{X}}$  получается в результате  
диагонального усреднения матрицы  $\hat{\mathbf{X}} = \sum_{i=1}^r \sqrt{\lambda_i} U_i V_i^T$ .

Преимущества:

- не требует задания модели ряда;
- способен выделять тренды произвольной формы;
- хорошо восстанавливает периодические составляющие.

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

Параметры:  $L$  — длина окна,  $r$  — кол-во собственных троек,  $U_i \in \mathbb{R}^L$ ,  $i = 1, \dots, r$  — собственные векторы,  $\underline{U}_i$  — первые  $L - 1$  координат и  $\pi_i$  — последняя координата.

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

Алгоритм рекуррентного прогнозирования:

❶ ряд  $\mathbb{Y}_{N+P} = (y_1, \dots, y_{N+P})$  определяется как

$$y_i = \begin{cases} \tilde{x}_i & \text{for } i = 1, \dots, N, \\ \sum_{j=1}^{L-1} a_j y_{i-j} & \text{for } i = N + 1, \dots, N + P, \end{cases} \quad (1)$$

❷ числа  $y_{N+1}, \dots, y_{N+P}$  являются результатом прогнозирования ряда на  $P$  точек вперёд.

- Метод выполняет точное прогнозирование временных рядов, управляемых линейно-рекуррентной формулой вида

$$x_n = \sum_j a_j x_{n-j};$$

т. е. подходит для аппроксимации широкого класса моделей рядов вида

$$x_n = \sum_i P_i(n) e^{\alpha_i n} \cos(2\pi\omega_i n + \phi_i);$$

- при наличии шума в данных сигнал может быть восстановлен приближённо;
- метод может использоваться для аппроксимации сигнала, содержащегося в ряде, локально управляемым линейно-рекуррентной формулой.

# Кросс-валидация для временных рядов

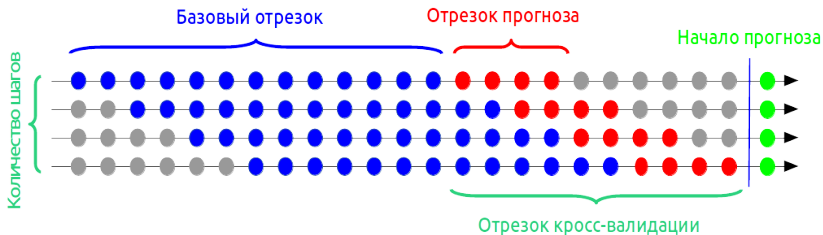


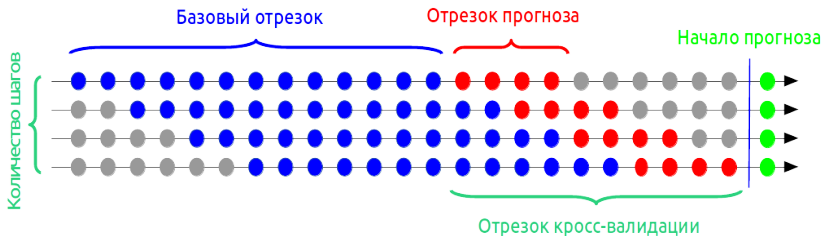
Рис. 4: Иллюстрация кросс-валидации для временных рядов (Hyndman, R. J., Athanasopoulos, G. (2013) «Forecasting: Principles & Practice»)

# Выбор параметров метода SSA и кросс-валидации

Длина прогноза рядов ПВЗ — 365 дней.

Параметры:

- длина окна  $L$ ;
- количество собственных троек  $r$ ;
- длина базового отрезка;
- длина отрезка прогноза  $P$  (фиксирована, 365 дней);
- длина отрезка кросс-валидации.



# Порядок подбора параметров

$j$  — момент начала прогноза;

$J$  — множество моментов начала прогноза, состоящее из наборов по 50 дней в каждом году из выбранных (2006, 2007, 2008, 2009, 2010) с шагом в 7 дней, начиная с 1 января;

$b$  — длина базового отрезка;

$v$  — длина отрезка кросс-валидации;

$P$  — длина прогноза (365 дней);

$T$  — множество перебираемых пар значений параметров  $L$  и  $r$ .

$$MSE_j^P(b, v) = \min_{(L, r) \in T} MSE(j, P, L, r, b, v),$$

$$MSE_{mean}^P(b, v) = \frac{\sum_{j \in J} MSE_j^P(b, v)}{|J|}.$$

- ❶  $b_0 = 5475$  (15 лет);
- ❷  $b^* = \operatorname{argmin}_b MSE_{mean}^P(b, v_0);$
- ❸  $v^* = \operatorname{argmin}_v MSE_{mean}^P(b^*, v).$

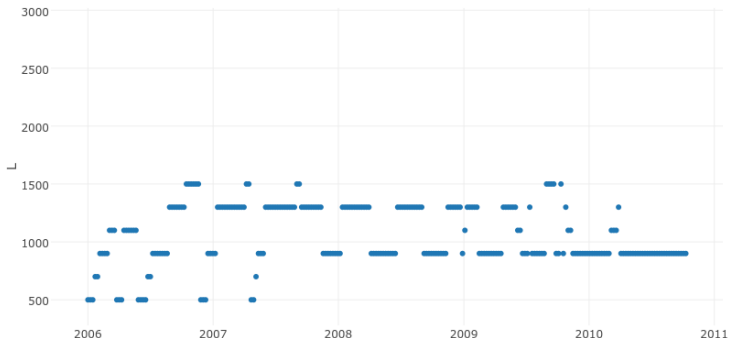


Рис. 5: Значение параметра  $L_j^{opt}$  для ряда  $y$  на 250 днях с шагом в 7 дней при  $b = 5475$  (15 лет),  $v = 1825$  (5 лет).

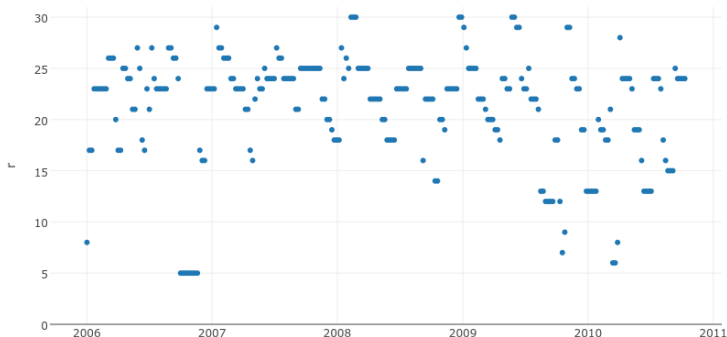


Рис. 6: Значение параметра  $r_j^{opt}$  для ряда  $y$  на 250 днях с шагом в 7 дней при  $b = 5475$  (15 лет),  $v = 1825$  (5 лет).



Таблица 1: Выбранные значения параметров для прогнозирования рядов ПВЗ на 365 дней.

Параметр	$x$	$y$	$LOD$	$dX$	$dY$
$b^*$ (лет)	15	15	15	15	15
$v^*$ (лет)	7	7	10	7	7

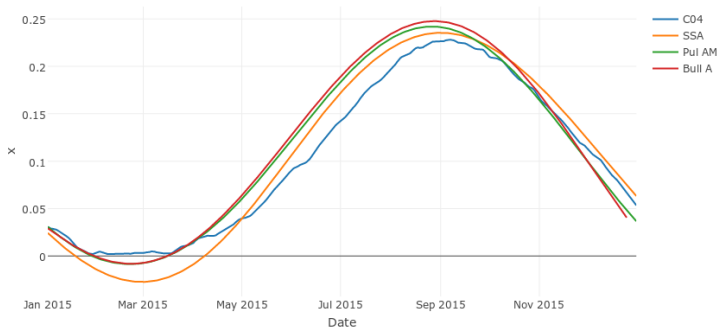


Рис. 7: Результат прогнозирования ряда  $x$  с помощью автоматического выбора параметров для 01.01.2015 ( $L_{opt} = 700, r_{opt} = 10$ )

Pul AM — прогноз Пулковской обсерватории (LS + ARIMA);

(<http://www.gao.spb.ru/english/as/persac/eopcphp/>)

Bull A — прогноз Международной службы вращения Земли.

(<https://datacenter.iers.org/eop/-/somos/5Rgv/latest/6>)

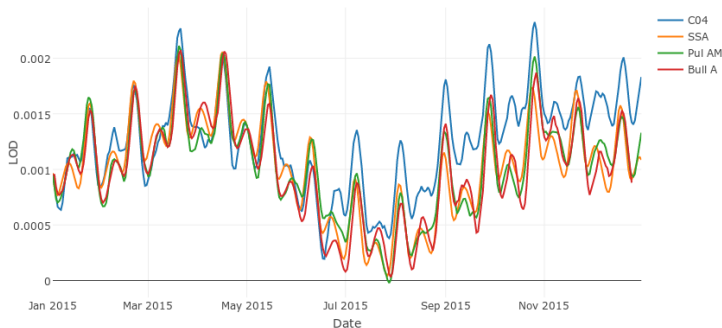


Рис. 8: Результат прогнозирования ряда  $LOD$  с помощью автоматического выбора параметров для 01.01.2015 ( $L_{opt} = 900, r_{opt} = 19$ )

Pul AM — прогноз Пулковской обсерватории (LS + AR);

(<http://www.gao.spb.ru/english/as/persac/eopcphp/>)

Bull A — прогноз Международной службы вращения Земли.

(<https://datacenter.iers.org/eop/-/somos/5Rgv/latest/6>)

# Полученные результаты: прогноз поправки полюса $dX$

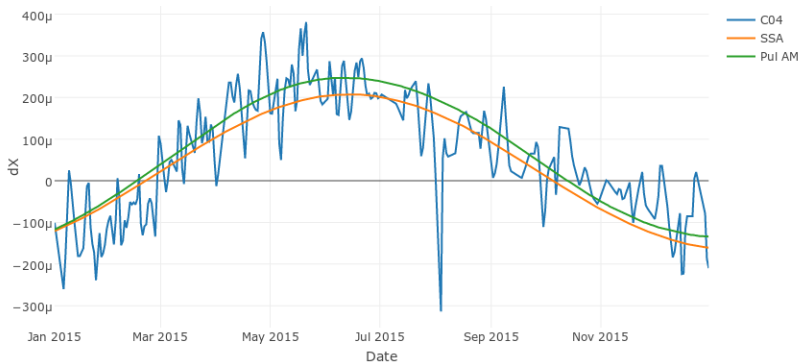


Рис. 9: Результат прогнозирования ряда  $dX$  с помощью автоматического выбора параметров для 01.01.2015,  $MJD = 57023$  ( $L_{opt} = 350, r_{opt} = 4$ )

Pul AM — прогноз Пулковской обсерватории.

(<http://www.gao.spb.ru/english/as/persac/eopcphp/>)

Таблица 2: Сравнение средних ошибок прогнозов ряда  $x$  на 365 дней с другими источниками (50 дней в каждом году с шагом в 7 дней)

	SSA	Pulkovo AM	Bulletin A
2011	$6.7 \times 10^{-4}$	$8.4 \times 10^{-4}$	$1.2 \times 10^{-3}$
2012	$8.7 \times 10^{-4}$	$7.4 \times 10^{-4}$	$4.9 \times 10^{-4}$
2013	$9.4 \times 10^{-4}$	$1.4 \times 10^{-3}$	$7.9 \times 10^{-4}$
2014	$4.4 \times 10^{-4}$	$8.8 \times 10^{-4}$	$7.2 \times 10^{-4}$
2015	$6.7 \times 10^{-4}$	$4.5 \times 10^{-4}$	$4.8 \times 10^{-4}$
Среднее	$7.2 \times 10^{-4}$	$8.6 \times 10^{-4}$	$7.5 \times 10^{-4}$
Медиана	$5.9 \times 10^{-4}$	$7.2 \times 10^{-4}$	$7.0 \times 10^{-4}$

Таблица 3: Сравнение средних ошибок прогнозов рядов ПВЗ

	SSA	Pulkovo AM	Bulletin A
$x$	$7.2 \times 10^{-4}$	$8.6 \times 10^{-4}$	$7.5 \times 10^{-4}$
$y$	$6.1 \times 10^{-4}$	$7.6 \times 10^{-4}$	$8.5 \times 10^{-4}$
$LOD$	$9.1 \times 10^{-8}$	$1.0 \times 10^{-7}$	
$dX$	$1.3 \times 10^{-8}$	$1.1 \times 10^{-8}$	
$dY$	$1.6 \times 10^{-8}$	$2.2 \times 10^{-8}$	

Таблица 4: Сравнение медиан ошибок прогнозов рядов ПВЗ

	SSA	Pulkovo AM	Bulletin A
$x$	$5.9 \times 10^{-4}$	$7.2 \times 10^{-4}$	$7.0 \times 10^{-4}$
$y$	$4.9 \times 10^{-4}$	$6.3 \times 10^{-4}$	$7.7 \times 10^{-4}$
$LOD$	$7.8 \times 10^{-8}$	$8.2 \times 10^{-8}$	
$dX$	$1.1 \times 10^{-8}$	$1.0 \times 10^{-8}$	
$dY$	$1.5 \times 10^{-8}$	$1.9 \times 10^{-8}$	

Для ежедневной публикации результатов прогнозирования рядов ПВЗ было создано веб-приложение с использованием пакета RShiny.

<http://eoppredict.ru>

Возможности:

- скачивание прогнозов ПВЗ на 365 дней для текущего или произвольного дня в прошлом, начиная с 1 января 2010 года;
- генерация прогнозов при заданных параметрах;
- сравнение прогнозов с другими источниками.

Date Input is limited between 27.08.2010 and 1.03.2016

Choose starting date

01.01.2015

☐ MJD labels

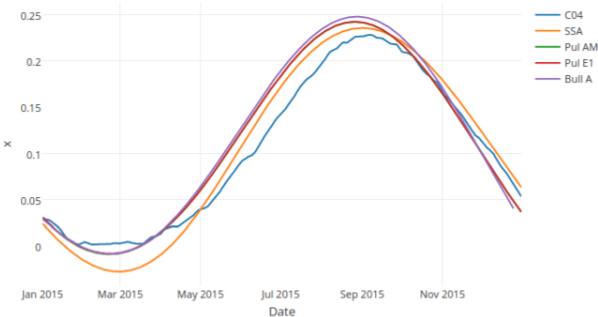
Download

MSE

	x	y	LOD	dX	dY
SSA	0.000331596675	0.000295618073	0.000000122859	0.000000006213	0.000000011058
Pul AM	0.000439326995	0.000809755335	0.000000104876	0.000000006143	0.000000018532
Pul E1	0.000441758041	0.000835092190	0.000000104657	0.000000006143	0.000000018532
Bull A	0.000601392920	0.000953560285	0.000000125508		

Showing 1 to 4 of 4 entries

Pole x





- Разработана методика подбора оптимальных параметров алгоритма рекуррентного прогнозирования метода SSA для прогнозирования рядов ПВЗ;
- реализован на языке R алгоритм автоматического подбора параметров для ежедневного прогнозирования пяти рядов ПВЗ с использованием пакета Rssa;
- выполнено сравнение прогнозов, полученных с использованием автоматически подобранных параметров, с прогнозами, предоставляемыми еженедельно МСВЗ и ежедневно Пулковской обсерваторией;
- создано веб-приложение, с помощью которого результаты работы программы прогнозирования публикуются ежедневно, а также выполняется сравнение прогнозов из различных источников с точными значениями на указанный день в прошлом, начиная с 1 января 2010 года.