

# Применение локальных модификаций метода анализа сингулярного спектра

Шаповал Егор Анатольевич, группа 17.Б04-мм

Санкт-Петербургский государственный университет  
Кафедра статистического моделирования  
Научный руководитель: к.ф.-м.н., доцент Голяндина Н.Э.  
Рецензент: программист Шлемов А.Ю

9 июня 2021 г.

# Постановка задачи

**Дано:** временной ряд  $X_N = (x_1, \dots, x_N)$ ,  $X_N = S_N + R_N$ , где  $S_N$  — детерминированный сигнал,  $R_N$  — случайный шум,  $\mathbb{E}r_n = 0$ .

**Задача анализа временных рядов:** построить прогноз сигнала  $S_N$ .

**Рассматриваемая стандартная модель:** сигнал управляется линейной рекуррентной формулой (ЛРФ),  $s_n = \sum_{k=1}^m a_k s_{n-k}$ .

**Прогноз:** по ЛРФ, оцененной с помощью метода анализа сингулярного спектра (SSA).

**Проблема:** сигнал приближённо управляется ЛРФ только локально, на отрезках длины  $Z < N$ .

**Предлагаемое решение:** двухуровневый прогноз сигнала.

- 1 Прогноз локальных ЛРФ, приближенно управляющих рядом в окрестности прогнозируемых точек
- 2 Прогноз ряда  $S_N$  с помощью спрогнозированных ЛРФ.

## Определение

Временной ряд  $S_\infty = (s_1, s_2, \dots)$  управляется *линейной рекуррентной формулой (ЛРФ) размерности  $m$*  если существует последовательность  $\{a_k\}_{k=1}^m$  такая, что  $a_m \neq 0$  и

$$s_{i+m} = \sum_{k=1}^m a_k s_{i+m-k} \quad \forall i = 1, 2, \dots \quad (1)$$

## Определение

ЛРФ наименьшей размерности  $m$  такой, что  $S_\infty$  управляется ЛРФ порядка  $m$  называется *минимальной ЛРФ*.

## Определение

ЛРФ размерности  $m$  с наименьшей нормой коэффициентов среди всех ЛРФ порядка  $m$ , управляющих  $S_\infty$ , называется *миннорм ЛРФ*.

## Определение

Пусть  $S_\infty$  управляется ЛРФ с коэффициентами  $\{a_k\}_{k=1}^m$ . Многочлен  $P_m(\mu) = \mu^m - \sum_{k=1}^m a_k \mu^{m-k}$  называется *характеристическим полиномом ЛРФ*.

## Определение

Корни характеристического полинома минимальной ЛРФ называются *сигнальными (главными) корнями*.

## Теорема (Hall MJ, Combinatorial theory, 1998)

Пусть  $\mu_1, \dots, \mu_p$  — корни характеристического полинома, кратностей  $k_1, \dots, k_p$ ,  $\sum_{j=1}^p k_j = d$ . Ряд  $S_\infty$  удовлетворяет ЛРФ (1) тогда и только

тогда, когда  $s_n = \sum_{j=1}^p \mu_j^n \sum_{l=0}^{k_j-1} c_{jl} n^l$ , где  $c_{jl} \in \mathbb{C}$  зависят от  $s_1, \dots, s_d$ .

# Глава 1: Оценивание подпространства сигнала с помощью алгоритма SSA

Анализ сингулярного спектра (singular spectrum analysis, SSA) [Golyandina, Nekrutkin, Zhigljavsky, 2001].

- 1 **Входные параметры** Ряд  $X_N = (x_1, \dots, x_N)$ , длина окна  $L \in 2 : (N - 1)$ , размерность подпространства  $r < \min(L, K)$ , где  $K = N - L + 1$ .

- 2 **Построение траекторной матрицы**

$$\mathbf{X} = \begin{pmatrix} x_1 & x_2 & x_3 & \dots & x_K \\ x_2 & x_3 & x_4 & \dots & x_{K+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_L & x_{L+1} & x_{L+2} & \dots & x_N \end{pmatrix}.$$

- 3 **Построение оценки базиса подпространства сигнала**  
 $\{U_i\}_{i=1}^r$  — о.н.с. из собственных векторов матрицы  $\mathbf{X}\mathbf{X}^T$ , взятых в порядке невозрастания собственных чисел.

*Ранг ряда* — ранг его траекторной матрицы.

## Нахождение оценки миннорм ЛРФ

Пусть  $\text{rank } S_N = r$  и  $U_1, \dots, U_r$  — базис подпространства сигнала, оцененный с помощью SSA. Для  $U = (u_1, \dots, u_L)^T \in \mathbb{R}^L$  обозначим  $\underline{U} := (u_1, \dots, u_{L-1})^T$ . Пусть  $\pi_i$  — последняя координата  $U_i$ ,  $i \in 1:r$ . Предположим, что  $\nu^2 := \sum_{i=1}^r \pi_i^2 < 1$ . Тогда коэффициенты прогнозирующей ЛРФ определяются как

$$\mathcal{R} := \frac{1}{1 - \nu^2} \sum_{i=1}^r \pi_i \underline{U}_i =: (a_{L-1}, \dots, a_1). \quad (2)$$

## Нахождение оценок главных корней х.п. ЛРФ

$\mathbf{U} := [U_1 : \dots : U_r]$ , рассмотрим  $\bar{\mathbf{U}}$ ,  $\underline{\mathbf{U}}$  — матрицы с удалённой первой и последней строкой соответственно. *ESPRIT-оценка главных корней*  $\{\mu_j\}_{j=1}^r$  — собственные числа матрицы  $\mathbf{D} = (\underline{\mathbf{U}}^T \underline{\mathbf{U}})^{-1} \underline{\mathbf{U}}^T \bar{\mathbf{U}}$ .

**Общий вид ряда**  $s_n = \sum_{j=1}^r c_j \mu_j^n \Leftrightarrow$  **корни х.п.**  $\mu_j = \rho_j e^{i \cdot 2\pi \omega_j} \Leftrightarrow$   
 $\Leftrightarrow$  **ЛРФ**

Рассмотрим ряд  $S_N$  с общим членом  $s_n = A \cos(2\pi\omega(n) + \varphi)$ .

## Определение

Пусть  $\omega(t)$  — непрерывно-дифференцируемая функция. Тогда мгновенной частотой в точке  $n$  называют  $\omega'(n)$ .

Смысл мгновенной частоты: в окрестности точки  $n_0$  ряд  $S_N$  хорошо приближается рядом  $s_n^{\text{approx}} = A \cos(2\pi\omega_0 n + \varphi)$ , где  $\omega_0 = \omega'(n_0)$ .

Свойства ряда с постоянной мгновенной частотой

$$s_n = A \cos(2\pi\omega n + \varphi) = \frac{Ae^{-i\varphi}}{2} \cdot e^{-i \cdot 2\pi\omega n} + \frac{Ae^{i\varphi}}{2} \cdot e^{i \cdot 2\pi\omega n} = c_1 \mu_1^n + c_2 \mu_2^n.$$

При  $0 < \omega < 0.5$  ряд имеет ранг 2, ему соответствуют корни

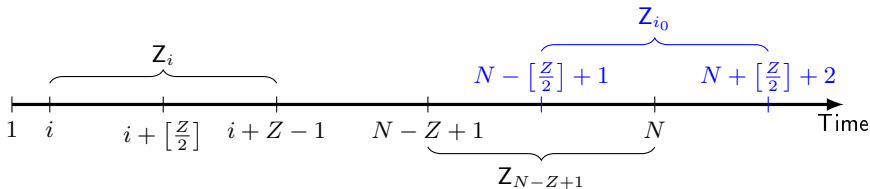
$\mu_1 = e^{-i \cdot 2\pi\omega}$ ,  $\mu_2 = e^{i \cdot 2\pi\omega}$ , он управляется ЛРФ

$$s_n = 2s_{n-1} \cos 2\pi\omega - s_{n-2}, \quad n \geq 3.$$

**Модель сигнала:**  $s_n = \sum_{j=1}^p \rho_j^n(n) \cos(2\pi\omega_j(n) + \varphi_j)$ . Предполагаем, что

- 1 На отрезках ряда  $Z_i$  длины  $Z$  каждое слагаемое хорошо приближается рядом  $s_n^{\text{approx}} = \rho_j^n(n_0) \cos(2\pi\omega_{j,0} \cdot n + \varphi)$ , где  $\omega_{j,0} = \omega'_j(n_0)$  и  $n_0$  — середина отрезка  $Z_i$ ,  $j \in 1:p$ .
- 2 Ряды  $\rho_j(n)$  и  $\omega'_j(n)$  ведут себя регулярным образом по  $n$ , в том числе, есть методы, которые могут строить их прогноз.

**Структура локальных отрезков:**



Для построения прогноза  $s_{N+1}$  нужно найти мгновенную частоту в точке  $N + 1$ .

Оценки корней, частот, модулей, коэффициентов ЛРФ будем нумеровать в соответствии с серединой локального отрезка.



## Входные параметры

- $X_N = (x_1, \dots, x_N)$  — временной ряд
- $Z$  — длина локального отрезка
- $L$  — фиксированная длина окна
- $r$  — ранг локальной аппроксимации
- $m$  — длина прогнозирующей ЛРФ
- $M$  — длина прогноза
- FOR MODs — алгоритм прогноза модулей корней
- FOR ARGs — алгоритм прогноза мгновенных частот

# Глава 2: Алгоритм двухуровневого прогноза (LocLRR SSA Forecast), часть 1

$Z_i = (x_i, \dots, x_{i+Z-1})$ ,  $i \in 1 : W$ , где  $W = N - Z + 1$ . Для каждого  $j \in 1 : r$

$$\begin{array}{ccccc}
 X_N & & \{\rho_j^{(i+[Z/2])}\}_{i=1}^W & \xrightarrow{\text{FOR MODs}} & \{\tilde{\rho}_j^{(N+i)}\}_{i=1}^M \\
 \downarrow Z & & \uparrow & & \downarrow \\
 \{Z_i\}_{i=1}^W & \xrightarrow[L, r]{\text{ESPRIT}} & \{\mu_j^{(i+[Z/2])}\}_{i=1}^W & & \{\tilde{\mu}_j^{(N+i)}\}_{i=1}^M \rightarrow \{\tilde{\mathcal{R}}_{N+i}\}_{i=1}^M \\
 & & \downarrow & & \uparrow \\
 & & \{\omega_j^{(i+[Z/2])}\}_{i=1}^W & \xrightarrow{\text{FOR ARGs}} & \{\tilde{\omega}_j^{(N+i)}\}_{i=1}^M
 \end{array}$$

$\{\tilde{\mathcal{R}}_{N+i}\}_{i=1}^M$  — последовательность прогнозирующих минимальных ЛРФ. Если  $m > r$ , удлиним каждую  $\tilde{\mathcal{R}}_{N+i}$ ,  $i \in 1 : M$ , до миннорм ЛРФ  $\tilde{\mathcal{R}}_{N+i}^{(m)}$  размерности  $m$  [Usevich, 2010].

**Результат:** последовательность прогнозирующих миннорм ЛРФ  $\{\tilde{\mathcal{R}}_{N+i}^{(m)}\}_{i=1}^M$ .

## Глава 2: Алгоритм двухуровневого прогноза (LocLRR SSA Forecast), часть 2

По  $\{\tilde{\mathcal{R}}_{N+i}^{(m)}\}_{i=1}^M$ ,  $\tilde{\mathcal{R}}_{N+i}^{(m)} = (a_m^{(N+i)}, \dots, a_1^{(N+i)})$ ,  $i \in 1 : M$ , построим прогноз сигнала  $S_N$ . Построим  $Y_{N+M}$ , у которого

$$y_n = \begin{cases} x_n, & n \in 1 : N, \\ \sum_{k=1}^m \tilde{a}_k^{(n)} y_{n-k}, & n \in (N+1) : (N+M). \end{cases}$$

**Результат:**  $\tilde{S}_{N+1, N+M} = (y_{N+1}, \dots, y_{N+M})$  — прогноз  $S_N$ .

Было проведено сравнение способов построения прогноза:

- LocLRR SSA Forecast
- Прогноз по константе  $c$  — прогноз  $\tilde{S}_{N+1,N+M} = (c, c, \dots, c)$
- Прогноз по последнему локальному промежутку — прогноз, построенный с помощью ЛРФ, посчитанной по корням последнего локального промежутка  $Z_{N-Z+1} = (x_{N-Z+1}, \dots, x_N)$ .

### Определение

Пусть  $S_{N+1,N+M} = (s_{N+1}, \dots, s_{N+M})$  — сигнал,  
 $\tilde{S}_{N+1,N+M} = (\tilde{s}_{N+1}, \dots, \tilde{s}_{N+M})$  — построенный прогноз.  
Среднеквадратичная ошибка есть

$$\text{MSE}(\tilde{S}_{N+1,N+M}, S_{N+1,N+M}) = \frac{1}{M} \sum_{i=1}^M (s_{N+i} - \tilde{s}_{N+i})^2. \quad (3)$$

Точность прогноза будем измерять как  $\text{RMSE} = \sqrt{\text{MSE}}$ .

**Модель сигнала:**  $s_n = \sum_{j=1}^p A \cos(2\pi\omega_j(n) + \varphi_j)$ , где диапазоны мгновенных частот слагаемых попарно не пересекаются и лежат в интервале  $(0, 0.5)$ .

**Примеры частотных модуляций:**

**Линейная модуляция**  $\omega(n) = (an)^2$ ,  $a \neq 0$ . Мгновенная частота  $\omega'(n) = 2a^2n$ . Прогноз  $\omega'$  строится по линейной регрессии.

**Синусоидальная модуляция**  $\omega(n) = \omega_{\text{ext}}(n + b \sin 2\pi\omega_{\text{int}}n)$ ,  $b > 0$ , частота  $\omega_{\text{int}} \in (0, 0.5)$ ,  $\omega_{\text{int}}$  гораздо меньше  $\omega_{\text{ext}}$ . Мгновенная частота  $\omega'(n) = \omega_{\text{ext}}(1 + 2\pi\omega_{\text{int}}b \cos 2\pi\omega_{\text{int}}n)$ . Прогноз  $\omega'$  строится методом SSA с длиной окна  $L$ , равной половине длины ряда и параметром  $r = 3$ .

**Сумма слагаемых с разными модуляциями** Прогноз каждой частоты строится в зависимости от вида модуляции.

Прогноз модуля корня во всех случаях строится как среднее арифметическое предыдущих значений.

# Глава 3: Сумма косинусов с линейной и синусоидальной частотой

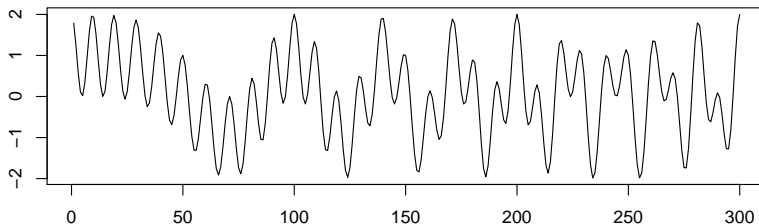


Рис.:  $s_n = \cos\left(2\pi\left(\frac{n}{100}\right)^2\right) + \cos\left(\frac{2\pi}{10}\left(n + \sin\frac{2\pi n}{100}\right)\right)$

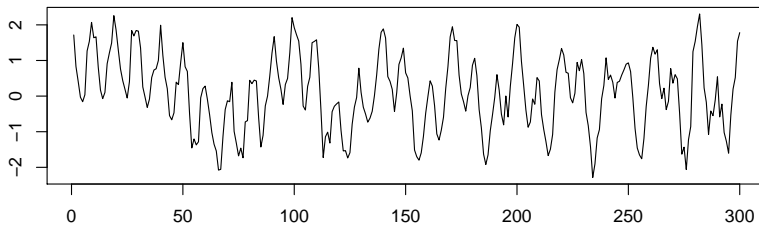


Рис.:  $x_n = s_n + \sigma \epsilon_n, \sigma = 0.25$

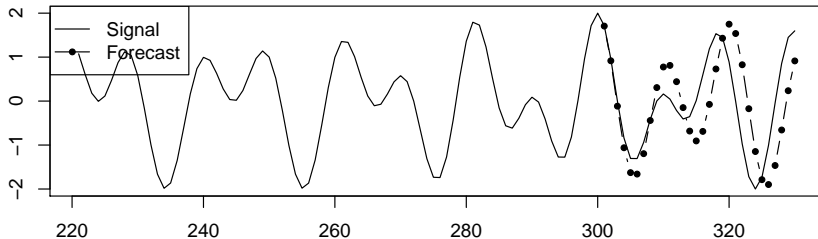


Рис.: По последнему локальному отрезку,  $m = 4$ ,  $\text{RMSE} = 0.88$

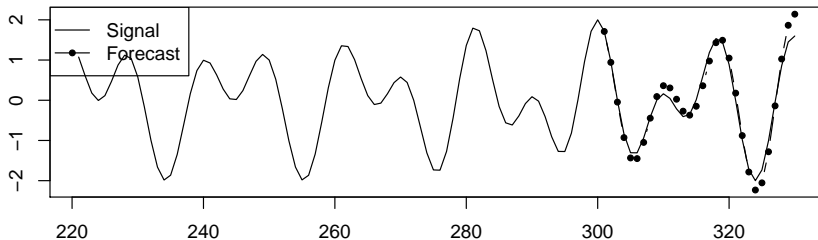


Рис.: С помощью LocLRR SSA Forecast,  $m = 4$ ,  $\text{RMSE} = 0.2$

# Глава 3: Оценки и прогноз частот $\omega^{(i)}$ и модулей $|\mu^{(i)}|$ , $\sigma = 0$

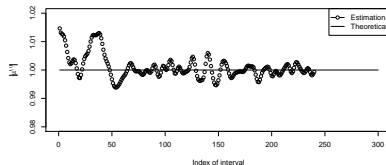
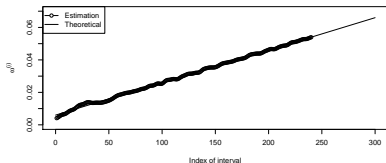


Рис.: Косинус с линейной частотой

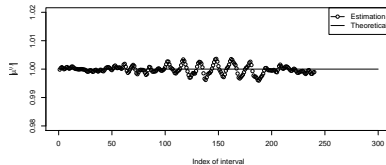
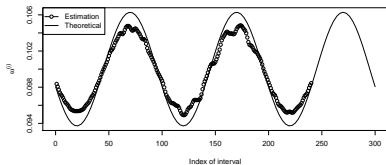


Рис.: Косинус с синусоидальной частотой



# Глава 3: Оценки и прогноз частот $\omega^{(i)}$ и модулей $|\mu^{(i)}|$ , $\sigma = 0$

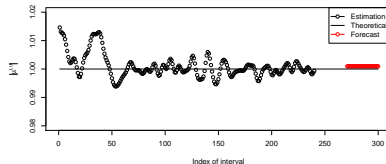
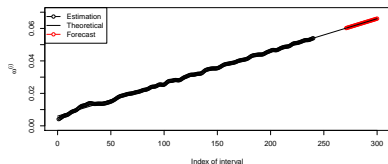


Рис.: Косинус с линейной частотой

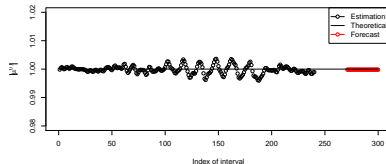
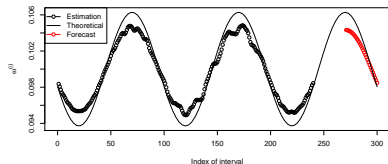


Рис.: Косинус с синусоидальной частотой

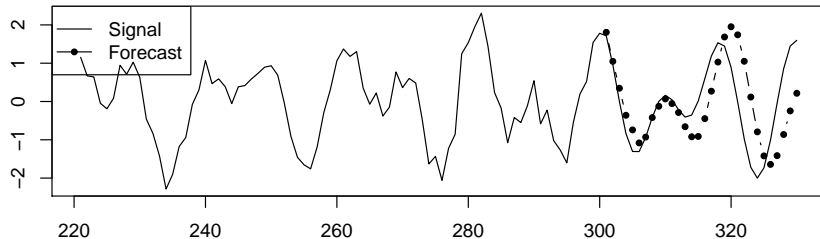


Рис.: По последнему локальному отрезку,  $m = 10$ ,  $\text{RMSE} = 0.957$

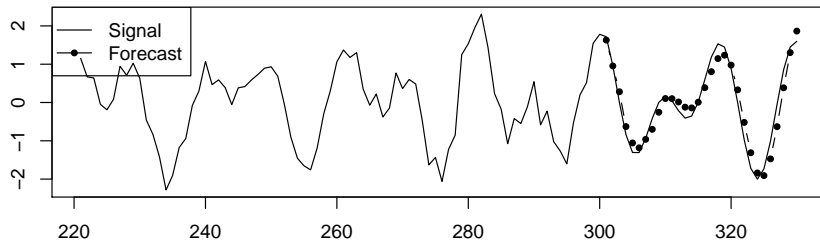


Рис.: С помощью LocLRR SSA Forecast,  $m = 12$ ,  $\text{RMSE} = 0.278$

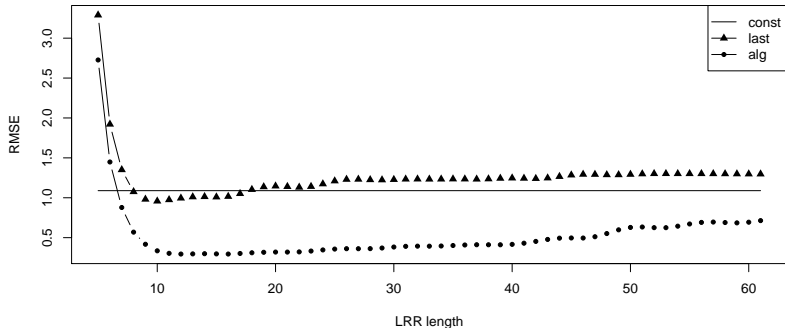


Рис.: RMSE в зависимости от длины ЛРФ,  $\sigma = 0.25$ ,  $n = 100$

Минимальная  $RMSE = 0.295$  достигается при прогнозе с помощью LocLRR SSA Forecast по ЛРФ размерности  $m = 12$ .

Модель: 
$$s_n = \sum_{j=1}^p \cos(2\pi\omega_j(n) + \varphi_j)$$

Рассмотренные примеры:

- Косинус с линейной частотой,  $s_n = \cos\left(2\pi\left(\frac{n}{100}\right)^2\right)$  (обозначим  $\cos(n^2)$ )
- Косинус с синусоидальной частотой,  
 $s_n = \cos\left(\frac{2\pi}{20}\left(n + 5\sin\frac{2\pi n}{100}\right)\right)$  (обозначим  $\cos\sin(n)$ )
- Сумма косинусов с линейной и синусоидальной частотой,  
 $s_n = \cos\left(2\pi\left(\frac{n}{100}\right)^2\right) + \cos\left(\frac{2\pi}{10}\left(n + \sin\frac{2\pi n}{100}\right)\right)$  (обозначим  $\cos(n^2) + \cos\sin(n)$ )
- Сумма двух косинусов с синусоидальной частотой,  
 $s_n = \cos\left(\frac{2\pi}{20}\left(n + \sin\frac{2\pi n}{100}\right)\right) + \cos\left(\frac{2\pi}{10}\left(n + 2\sin\frac{2\pi n}{140}\right)\right)$  (обозначим  $\cos\sin(n) + \cos\sin(n)$ )

Длина ряда  $N = 300$ , длина прогноза  $M = 30$ .

Таблица: RMSE прогноза на примерах

$S_N$	$\sigma$	$c = 0$	last		alg	
			RMSE	$m$	RMSE	$m$
$\cos(n^2)$	0	0.689	0.717	2	0.014	3
	0.25	0.733	0.754	5	0.135	11
$\cos \sin(n)$	0	0.698	0.309	2	0.097	2
	0.25	0.741	0.438	6	0.232	7
$\cos(n^2) + \cos \sin(n)$	0	1.089	0.88	6	0.2	4
	0.25	1.117	0.958	10	0.295	12
$\cos \sin(n) + \cos \sin(n)$	0	0.908	0.587	4	0.191	5
	0.25	0.942	0.656	28	0.291	15

Что сделано:

- 1 Разработан алгоритм двухуровневого прогноза (LocLRR SSA Forecast)
- 2 Разработана методика применения алгоритма LocLRR SSA Forecast к сумме частотно-модулированных косинусов с линейной и синусоидальной модуляцией
- 3 Проведено численное исследование метода на примерах. Прогноз с помощью LocLRR SSA Forecast даёт хорошую точность. Таким образом показана перспективность построения прогноза с помощью LocLRR SSA Forecast.

Что планируется сделать:

- Провести сравнение с другими методами построения прогноза
- Расширить класс рядов, к которым применим алгоритм двухуровневого прогноза
- Выработать рекомендации по выбору параметров LocLRR SSA Forecast, реализуемые на реальных данных