

Общие компоненты временных рядов

Чёрный Артём Константинович, 422 гр.

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

Научный руководитель: к.ф.-м.н., доц. Голяндина Н. Э.

Рецензент: ассистент Шлемов А. Ю.



Санкт-Петербург, 2015

- Временной ряд $X = (x_1, \dots, x_N)$, $x_i \in \mathbb{R}$.
- Система временных рядов (X_1, \dots, X_m) , длина каждого ряда равна N , $X_k = F_k + R_k$, где F_k — интересующая нас компонента ряда, R_k — остаток.

Задача

Выделить из системы рядов (X_1, \dots, X_m) компоненту (F_1, \dots, F_m) ;

Известные методы:

- Singular Spectrum Analysis (**SSA**);
- Multivariate Singular Spectrum Analysis (**MSSA**).

Задачи данной работы:

- Построение новых модификации метода MSSA **JointSSA** и **Simultaneous Components Analysis (SCA-SSA)**;
- Исследование свойств новых классов модификаций;
- Численное сравнение результатов.

Определение

Траекторной матрицей временного ряда $X = (x_1, \dots, x_N)$ называется матрица

$$\mathbf{X} = (x_{ij})_{i,j=1}^{L,K} = \begin{pmatrix} x_1 & x_2 & \cdots & x_K \\ x_2 & x_3 & \cdots & x_{K+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_L & x_{L+1} & \cdots & x_N \end{pmatrix},$$

где $L, 1 < L < N$ — **длина окна**, $K = N - L + 1$.

- Вложение:
временные ряды \mapsto матрица

$$\mathcal{T}: \mathbb{R}^{N \times m} \rightarrow \mathbb{R}^{L \times mK}$$

$$X_1, \dots, X_m \mapsto [X_1 : X_2 : \dots : X_m].$$

- Диагональное усреднение:
матрица \mapsto временные ряды

$$\mathcal{T}^{-1}\mathcal{H}: \mathbb{R}^{L \times mK} \rightarrow \mathbb{R}^{N \times m}$$

$$[X_1 : X_2 : \dots : X_m] \mapsto X_1, \dots, X_m.$$

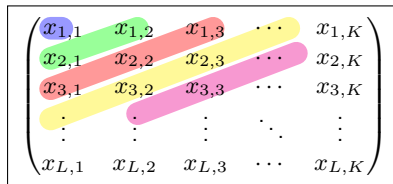


Рис. Диагональное усреднение

Определение

L — длина окна, $\mathbf{X} = \mathcal{T}(X)$.

Ряд X имеет L -ранг d , если $\text{rk } \mathbf{X} = d$.

Ряды, для которых L -ранг равен d для любого допустимого L , называются **рядами конечного ранга**.

Обозначение для ранга ряда: $\text{rk } X$.

Пример: $x_n = \sum_i p_i(n) e^{\alpha_i n} \cos(2\pi\omega_i n + \varphi_i)$, где $p_i(n) = \sum_k a_{ik} n^k$.

Определение

$\mathbf{X}_k = \mathcal{T}(X_k)$, $\mathbf{X} = [\mathbf{X}_1 : \dots : \mathbf{X}_m]$.

Система временных рядов X_1, \dots, X_m имеет L -ранг d , если $\text{rk } \mathbf{X} = d$.

Обозначение для ранга системы рядов: $\text{rk}[X_1, \dots, X_m]$.

Фиксируем $r \leq L$. Обозначим:

$$\mathcal{D}_r \stackrel{\text{def}}{=} \{\mathbf{D} \in \mathbb{R}^{r \times r} : \mathbf{D} = \text{diag}(d_1, \dots, d_r), d_i \geq 0, i = 1, \dots, r\},$$

$$\mathcal{O}_{L,r} \stackrel{\text{def}}{=} \{\mathbf{U} \in \mathbb{R}^{L \times r} : \mathbf{U}^T \mathbf{U} = \mathbf{I}_r\},$$

$$\mathcal{P}_{L,r} \stackrel{\text{def}}{=} \{\mathbf{P} \in \mathbb{R}^{L \times L} : \mathbf{P}^2 = \mathbf{P}, \text{Im } \mathbf{P} \perp \text{Ker } \mathbf{P}, \text{rk } \mathbf{P} = r\}.$$

Вход: Система временных рядов (X_1, \dots, X_m) , $X_k = F_k + R_k$, длина каждого ряда равна $N > 2$. Параметры: длина окна L , степень аппроксимации r .

Выход: оценка F_k .

1 Вложение (\mathcal{T})

$$X_k \mapsto \mathbf{X}_k \in \mathbb{R}^{L \times K}, \quad K = N - L + 1, \quad k = 1, \dots, m;$$

$$(X_1, X_2, \dots, X_m) \mapsto \mathbf{X} = [\mathbf{X}_1 : \mathbf{X}_2 : \dots : \mathbf{X}_m] \in \mathbb{R}^{L \times mK}.$$

2 Поиск ортонормированной системы векторов

Поиск $\mathbf{U}^{(r)} = [U_1^{(r)} : \dots : U_r^{(r)}] \in \mathcal{O}_{L,r}$, зависящий от $\mathbf{X}_k, k = 1, \dots, m$.

3 Группировка и проектирование

Выбор группы индексов $I = \{i_1, \dots, i_n\} \subset \{1, \dots, r\}$

$$\mathbf{U}_I^{(r)} = [U_{i_1}^{(r)} : \dots : U_{i_n}^{(r)}] \mapsto \mathbf{P}_I^{(r)} = \mathbf{U}_I^{(r)} \left(\mathbf{U}_I^{(r)} \right)^T \in \mathcal{P}_{L,n},$$

$$\mathbf{Y}_k = \mathbf{P}_I^{(r)} \mathbf{X}_k, \quad k = 1, \dots, m.$$

4 Диагональное усреднение и переход к рядам ($\mathcal{T}^{-1}\mathcal{H}$)

$$\hat{F}_k = \mathcal{T}^{-1}\mathcal{H}(\mathbf{Y}_k) \text{ — оценка } F_k, k = 1, \dots, m.$$

Определение (Сингулярное разложение)

Сингулярным разложением матрицы $\mathbf{X} \in \mathbb{R}^{L \times M}$ ранга d называется разложение

$$\mathbf{X} = \mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{V}^T = \sum_{i=1}^d \sqrt{\lambda_i} U_i V_i^T,$$

где $\mathbf{U} \in \mathcal{O}_{L,d}$, $\mathbf{V} \in \mathcal{O}_{M,d}$, $\mathbf{\Lambda} \in \mathcal{D}_d$, $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d > 0$.

MSSA: $I = \{i_1, \dots, i_n\}$, $\mathbf{U}^{(r)} = [U_1 : \dots : U_r]$, $\mathbf{U}_I^{(r)} = [U_{i_1} : \dots : U_{i_n}]$.

Определение (Преимственность задачи)

Рассмотрим некоторую задачу поиска ортонормированной системы $\mathbf{U}^{(r)} \in \mathcal{O}_{L,r}$, зависящую от параметра $r \leq L$. Будем говорить, что задача обладает **преимственностью**, если для любых r_1 и r_2 , $r_2 > r_1$ существуют решения такие, что

$$\text{colspan}(\mathbf{U}^{(r_1)}) \subset \text{colspan}(\mathbf{U}^{(r_2)}).$$

Таким образом, задача нахождения $\mathbf{U}^{(r)}$ в методе MSSA обладает преимственностью.

$\mathbf{X}_k = \mathcal{T}(\mathbf{X}_k)$, $\mathbf{S}_k = \mathbf{X}_k \mathbf{X}_k^T$, $k = 1, \dots, m$, $\mathbf{X} = [\mathbf{X}_1 : \dots : \mathbf{X}_m]$.
 $\mathbf{X} = \mathbf{U} \mathbf{\Lambda}^{1/2} \mathbf{V}^T$ - сингулярное разложение, $\mathbf{U}^{(r)} = [U_1 : \dots : U_r]$.

Оптимальные свойства:

$$\sum_{k=1}^m \|\mathbf{X}_k - \mathbf{U} \mathbf{D}_k \mathbf{V}_k^T\|^2 \rightarrow \min_{\mathbf{U}, \mathbf{D}_k, \mathbf{V}_k}, \quad (1)$$

$$\left\| \sum_{k=1}^m \mathbf{S}_k - \mathbf{U} \mathbf{D} \mathbf{U}^T \right\|^2 \rightarrow \min_{\mathbf{U}, \mathbf{D}}, \quad (2)$$

где $\mathbf{U} \in \mathcal{O}_{L,r}$, $\mathbf{D}, \mathbf{D}_k \in \mathcal{D}_r$, $\mathbf{V} = [\mathbf{V}_1^T : \dots : \mathbf{V}_m^T]^T \in \mathcal{O}_{mK,r}$.

Свойства:

- Минимум функционалов для рядов конечного ранга при $r \geq \text{rk}[\mathbf{X}_1, \dots, \mathbf{X}_m]$ равен нулю.
- Минимум функционалов равен нулю тогда и только тогда, когда восстановление точное, т.е. $\mathbf{P}^{(r)} \mathbf{X}_k = \mathbf{X}_k$, где $\mathbf{P}^{(r)} = \mathbf{U}^{(r)} \left(\mathbf{U}^{(r)} \right)^T$;
- Преемственность.
- При $m = 1$ совпадает с методом SSA.

$$\mathbf{X}_k = \mathcal{T}(\mathbf{X}_k), \mathbf{S}_k = \mathbf{X}_k \mathbf{X}_k^T, k = 1, \dots, m.$$

$$\sum_{k=1}^m \|\mathbf{X}_k - \mathbf{Z}_k\|^2 \rightarrow \min_{\mathbf{Z}_k \in \Psi}, \sum_{k=1}^m \|\mathbf{S}_k - \Sigma_k\|^2 \rightarrow \min_{\Sigma_k \in \Phi}.$$

- MSSA:

$$\Psi_{\text{MSSA}} : \mathbf{Z}_k = \mathbf{U} \mathbf{D}_k \mathbf{V}_k^T,$$

где $\mathbf{U} \in \mathcal{O}_{L,r}$, $\mathbf{D}_k \in \mathcal{D}_r$, $\mathbf{V} = [\mathbf{V}_1^T : \dots : \mathbf{V}_m^T]^T \in \mathcal{O}_{mK,r}$.

- Covariance Orthogonal Dimensionality Reduction (CODR):

$$\Phi_{\text{CODR}} : \Sigma_k = \mathbf{U} \mathbf{T}_k \mathbf{U}^T,$$

где $\mathbf{U} \in \mathcal{O}_{L,r}$, $\mathbf{T}_k \in \mathbb{R}^{r \times r}$ — неотрицательно определенные, $\mathbf{T}_k = \mathbf{T}_k^T$.

- Diagonal Orthogonal Dimensionality Reduction (DODR):

$$\Phi_{\text{DODR}} : \Sigma_k = \mathbf{U} \mathbf{D}_k \mathbf{U}^T,$$

где $\mathbf{U} \in \mathcal{O}_{L,r}$, $\mathbf{D}_k \in \mathcal{D}_r$.

- Orthogonal Simultaneous Components Analysis (O-SCA):

$$\Psi_{\text{O-SCA}} : \mathbf{Z}_k = \mathbf{U} \mathbf{D}_k \mathbf{V}_k^T,$$

где $\mathbf{U} \in \mathcal{O}_{L,r}$, $\mathbf{D}_k \in \mathcal{D}_r$, $\mathbf{V}_k \in \mathcal{O}_{K,r}$.

Группировка:

- Выбор $I = \{i_1, \dots, i_n\} \subset \{1, \dots, r\}$, $n \leq r$.
- Проектирование и переход к рядам:

$$\mathbf{U}^{(r)} = [U_1^{(r)} : \dots : U_r^{(r)}] \mapsto \mathbf{U}_I^{(r)} = [U_{i_1}^{(r)} : \dots : U_{i_n}^{(r)}] \mapsto \mathbf{P}_I^{(r)} = \mathbf{U}_I^{(r)} \left(\mathbf{U}_I^{(r)} \right)^T,$$

$$\hat{\mathbf{F}}_k = \mathcal{T}^{-1} \mathcal{H} \mathbf{P}_I^{(r)} \mathcal{T} \mathbf{X}_k.$$

Для выполнения группировки нужно упорядочить столбцы $\mathbf{U}^{(r)}$.

Если $\mathbf{U} = [U_1 : \dots : U_r] \in \mathcal{O}_{L,r}$, то выполнено

$$\|\mathbf{X}\|^2 = \sum_{i=1}^r \|U_i^T \mathbf{X}\|^2 + \|\left(\mathbf{U}^\perp\right)^T \mathbf{X}\|^2.$$

Пусть $\mathbf{X} = [\mathbf{X}_1 : \dots : \mathbf{X}_m]$, где $\mathbf{X}_k = \mathcal{T}(\mathbf{X}_k)$. Введем упорядоченность:

$$\left\| U_i^T \mathbf{X} \right\|^2 \geq \left\| U_j^T \mathbf{X} \right\|^2 \iff i \leq j.$$

В MSSA $\lambda_i = \left\| U_i^T \mathbf{X} \right\|^2$, т.е. $\lambda_i \geq \lambda_j \iff i \leq j$.

Модель: $\Sigma_k = \mathbf{U}\mathbf{T}_k\mathbf{U}^T$, $k = 1, \dots, m$:

$$\sum_{k=1}^m \|\mathbf{S}_k - \mathbf{U}\mathbf{T}_k\mathbf{U}^T\|^2 \rightarrow \min_{\mathbf{U}}, \mathbf{U} \in \mathcal{O}_{L,r}, \quad (3)$$

\mathbf{T}_k — симметрическая неотрицательно определенная матрица.

Совпадает с задачей Common Components Analysis (CCA) [Wang, 2011].

Свойства:

- Минимум функционала равен нулю тогда и только тогда, когда восстановление точное.
- Минимум функционала для рядов конечного ранга при $r \geq \text{rk}[\mathbf{X}_1, \dots, \mathbf{X}_m]$ равен нулю.
- При $m = 1$ совпадает с методом SSA.
- Если $\mathbf{U} \in \mathcal{O}_{L,r}$ — решение задачи, то для любого $\mathbf{\Pi} \in \mathcal{O}_{r,r}$ матрица $\mathbf{U}\mathbf{\Pi}$ — решение.
- Отсутствие преимущества.

Модель: $\Sigma_k = \mathbf{U}\mathbf{D}_k\mathbf{U}^T$, $k = 1, \dots, m$:

$$\sum_{k=1}^m \|\mathbf{S}_k - \mathbf{U}\mathbf{D}_k\mathbf{U}^T\|^2 \rightarrow \min_{\mathbf{U}, \mathbf{D}_k}, \mathbf{U} \in \mathcal{O}_{L,r}, \mathbf{D}_k \in \mathcal{D}_r. \quad (4)$$

Совпадает с задачей O-INDSCAL [Trendafilov, 2004].

Свойства:

- Если минимум функционала равен нулю, то восстановление точное. Обратное, вообще говоря, неверно.
- При выполнении некоторых условий минимум функционала для рядов конечного ранга при $r \geq \text{rk}[\mathbf{X}_1, \dots, \mathbf{X}_m]$ равен нулю.
- При $m = 1$ решение совпадает с решением SSA.
- **Преимственность?**

Модель: $\mathbf{Z}_k = \mathbf{U}\mathbf{D}_k\mathbf{V}_k^T$, $k = 1, \dots, m$.

$$\sum_{k=1}^m \|\mathbf{X}_k - \mathbf{U}\mathbf{D}_k\mathbf{V}_k^T\|^2 \rightarrow \min_{\mathbf{U}, \mathbf{D}_k, \mathbf{V}_k}, \mathbf{U} \in \mathcal{O}_{L,r}, \mathbf{D}_k \in \mathcal{D}_r, \mathbf{V}_k \in \mathcal{O}_{K,r} \quad (5)$$

Утверждение

Фиксируем длину окна L . Рассмотрим систему рядов длины N :

$$\mathbf{X}_k: x_n^{(k)} = \sum_{i=1}^{d_k} A_i^{(k)} \cos(2\pi n \omega_i^{(k)}), \quad k = 1, \dots, m,$$

где $\omega_i^{(k)} \in (0; 1/2]$, $i = 1, \dots, d_k$, $k = 1, \dots, m$. Если $L\omega_i^{(k)}$, $K\omega_i^{(k)}$ целые, $i = 1, \dots, d_k$, $k = 1, \dots, m$, $K = N - L + 1$, то минимум функционала равен 0 при $r \geq \text{rk}[\mathbf{X}_1, \dots, \mathbf{X}_m]$. При этом столбцы решения $\mathbf{U}^{(r)}$ являются первыми r левыми сингулярными векторами матрицы $\mathbf{X} = [\mathbf{X}_1 : \dots : \mathbf{X}_m]$.

Модель: $\mathbf{Z}_k = \mathbf{U}\mathbf{D}_k\mathbf{V}_k^T$, $k = 1, \dots, m$.

$$\sum_{k=1}^m \|\mathbf{X}_k - \mathbf{U}\mathbf{D}_k\mathbf{V}_k^T\|^2 \rightarrow \min_{\mathbf{U}, \mathbf{D}_k, \mathbf{V}_k},$$

$$\mathbf{U} \in \mathcal{O}_{L,r}, \mathbf{D}_k \in \mathcal{D}_r, \mathbf{V}_k \in \mathcal{O}_{K,r}, k = 1, \dots, m.$$

Свойства:

- Если минимум функционала равен нулю, то восстановление точное. Обратное, вообще говоря, неверно.
- При выполнении некоторых условий минимум функционала для рядов конечного ранга при $r \geq \text{rk}[\mathbf{X}_1, \dots, \mathbf{X}_m]$ равен нулю.
- При $m = 1$ решение совпадает с решением SSA.
- **Отсутствие преимственности.**

Метод ALS: найти решение задачи

$$\|\mathbf{X} - \mathbf{A}\mathbf{B}^T\|^2 \rightarrow \min_{\mathbf{A} \in \mathcal{A}, \mathbf{B} \in \mathcal{B}}. \quad (6)$$

Вход: \mathbf{A}_0 — начальная матрица.

Алгоритм:

$$\mathbf{B}_i = \operatorname{argmin}_{\mathbf{B} \in \mathcal{B}} \|\mathbf{X} - \mathbf{A}_i \mathbf{B}^T\|^2,$$

$$\mathbf{A}_{i+1} = \operatorname{argmin}_{\mathbf{A} \in \mathcal{A}} \|\mathbf{X} - \mathbf{A} \mathbf{B}_i^T\|^2$$

Используемые пакеты и функции:

- SSA, MSSA: пакет Rssa

```
ssa(f, kind = 'mssa', L)  
ssa(f, L)
```

- CODR: реализация Iterative EVD Algorithm for CCA [Wang, 2011]

```
cca(S, r = r, eps, maxiter)
```

Особенность реализации: решает проблему упорядоченности.

- DODR: пакет PTak

```
CANDPARA(S, dim = r, test = eps, Maxiter = maxiter)
```

Особенность реализации: преемственность решения алгоритма.

- O-SCA: собственная реализация алгоритма с помощью ALS

```
sca(X, r = r, eps, maxiter)
```

X_1, \dots, X_m — система временных рядов, $X_k = F_k + R_k$, F_k — сигнал, R_k — шум.

Напоминание: **MSSA** при $m = 1$ называется **SSA**.

- Ряды разной структуры: $\text{rk}[F_1, \dots, F_m] = \text{rk} F_1 + \text{rk} F_2 + \dots + \text{rk} F_m$.
- Ряды одинаковой структуры: $\text{rk}[F_1, \dots, F_m] = \max_k \text{rk} F_k$.

Для рядов разной структуры лучше применять **SSA** к каждому ряду по отдельности, для рядов одинаковой структуры — **MSSA** к системе рядов.

Задача сравнения: найти модификацию **MSSA**, которая будет не хуже в общем случае, а для рядов с разной структуры будет лучше.

- Ряды одинаковой структуры:

$$X^{(1)} : x_n^{(1)} = 1.8 \cos(2\pi n/8) + \cos(2\pi n/12) + \varepsilon_n^{(1)},$$

$$X^{(2)} : x_n^{(2)} = 2 \cos(2\pi n/8) + 1.5 \cos(2\pi n/12) + \varepsilon_n^{(2)},$$

$N = 71$, $\varepsilon_n^{(k)} \sim N(0, 0.25)$, $\text{rk}[F^{(1)}, F^{(2)}] = 4$, число итераций = 30.

Возрастание ошибок:

- $MSSA \approx CODR \approx DODR \approx O\text{-}SCA < SSA$.

- Ряды различной структуры:

$$X^{(1)} : x_n^{(1)} = 2.5 \cos(2\pi n/8) \quad +0 \quad +\varepsilon_n^{(1)},$$

$$X^{(2)} : x_n^{(2)} = 0 \quad +2 \cos(2\pi n/6) \quad +\varepsilon_n^{(2)},$$

$N = 71$, $\varepsilon_n^{(k)} \sim N(0, 0.25)$, $\text{rk}[F^{(1)}, F^{(2)}] = 4$, число итераций = 30.

Возрастание ошибок:

- $SSA < MSSA \approx CODR \approx DODR \approx O\text{-}SCA$.

$$X^{(1)}: x_n^{(1)} = 1.8 \cos(2\pi n/8) + \cos(2\pi n/12) + \varepsilon_n^{(1)},$$

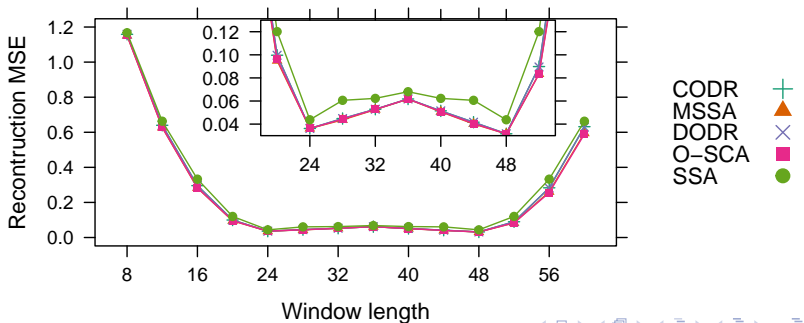
$$X^{(2)}: x_n^{(2)} = 2 \cos(2\pi n/8) + 1.5 \cos(2\pi n/12) + \varepsilon_n^{(2)},$$

$N = 71$, $\varepsilon_n^{(k)} \sim N(0, 0.25)$, $\text{rk}[F^{(1)}, F^{(2)}] = 2$, число итераций = 30.

Результаты по возрастанию ошибок:

MSSA \approx CODR \approx DODR \approx O-SCA $<$ SSA.

MSE \sim L. $r = 4$. $l = \{1, 2\}$



$$X^{(1)}: x_n^{(1)} = 2.5 \cos(2\pi n/8) + 0 + \varepsilon_n^{(1)},$$

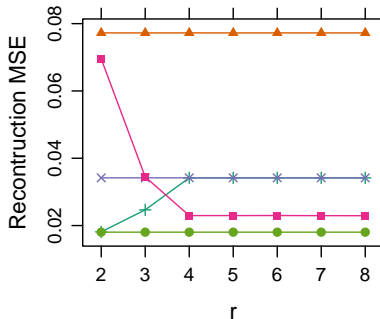
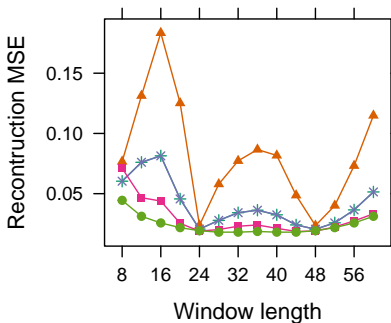
$$X^{(2)}: x_n^{(2)} = 0 + 2 \cos(2\pi n/6) + \varepsilon_n^{(2)},$$

$N = 71$, $\varepsilon_n^{(k)} \sim N(0, 0.25)$, $\text{rk}[F^{(1)}, F^{(2)}] = 2$, число итераций = 30.

CODR + MSSA ▲ DODR × O-SCA ■ SSA ●

MSE ~ L. r = 4. I = {1,2}

MSE ~ r. L = 32. I = {1,2}



Результаты:

- Изучены методы SSA, MSSA, их основные свойства;
- Предложены модификации метода MSSA;
- Проведено сравнение различных свойств этих модификаций;
- Исследованы свойства модификаций метода MSSA, связанные с рядами конечного ранга;
- Построен метод SCA-SSA, отделяющий компоненты сигнала в общем случае не хуже, чем MSSA, а в случае рядов различной структуры лучше метода MSSA;
- Реализованы и разработаны различные алгоритмы на языке R для решения поставленных задач.