

# Tensor decomposition and forecast for multidimensional time series

Semkin Kirill<sup>1</sup> and Strijov Vadim<sup>2</sup>

<sup>1</sup>Moscow Institute of Physics and Technology, semkin.ki@phystech.edu

<sup>2</sup>Moscow Institute of Physics and Technology, strijov@ccas.ru

## Abstract

Processing of multidimensional time series is associated with additional task of modeling interconnections between signals. Its ignorance may cause incorrectness of any method's outcomes. On the other hand, taking this dependencies into account makes models larger and less interpretable as well as series analysis more narrow. A non-parametric method based on tensor data representation and SSA approach is proposed in the paper. Then, series decomposition technique and forecast type and formula were derived. Finally, elaborated theory was applied to electricity consumption and weather data. Obtained results were compared with mSSA, VAR, RNN models.

**Key words:** time series, decomposition, forecast, SSA, CPD.

## 1 Introduction

The main object of the paper is multidimensional time series — a set of observed signals  $\{x_i(t)\}_{i=1}^m$  where  $t \in 1 \dots N$ .

Decomposition [1, 2, 3] and forecast [4, 1, 5] methods for sole signal can not be directly transferred to set of signals if their behavior is interconnected. For instance, let's take predator-prey model [6]. The size of one population is dependent on another's and vice-versa. We are further assuming this connection to exist and will give formal definitions in the scope of particular models.

First, RNN model [7, 8] links series with each other and their past values through composition of many nonlinear transformations. This information is encapsulated in hidden vector at each step. With that, one is able to forecast future values.

Second, vector autoregression (VAR) [9, 10] is a linear stochastic model for multidimensional signals. Denoting vector  $\mathbf{x}_t = (x_1(t) \dots x_m(t))^\top$  as a series realizations at time  $t$ , the further dynamics is following:

$$\mathbf{x}_t = \boldsymbol{\mu} + \sum_{i=1}^p A_i \mathbf{x}_{t-i} + \mathbf{u}_t$$

Here  $\boldsymbol{\mu}$  — some constant vector,  $A_i$  are matrices  $m \times m$ ,  $\mathbf{u}_t$  is a random vector (e.g. white noise  $WN(t)$ ). Signal's connection is defined by  $A_i$  transformations, so again each series is dependent on each other and the whole past *linearly*.

Revised solutions allow to make predictions, but have quite a number of hyperparameters to be tuned. They also require exhaustive learning procedures. Furthermore, their structure does not contain explicit way of making series decomposition.

To tackle listed issues, we developed a new method called tSSA which has only two adjustable variables. Computing tensor Canonical polyadic decomposition (CPD) is the only requirement for forecast and decomposition. Next, the approach is an extension of SSA method [11] for multidimensional time series. Due to that, its theory will be based on dynamical systems subtopic — *signal subspace* [12]. The main idea here is to build a phase representation of the observed signals (fig. 1).

In the end, we should mention another SSA modification — mSSA [13]. In contrast, it operates with data only in a way of matrices and matrix decomposition. Nonetheless, both methods share same ideas in forecast and decomposition techniques.

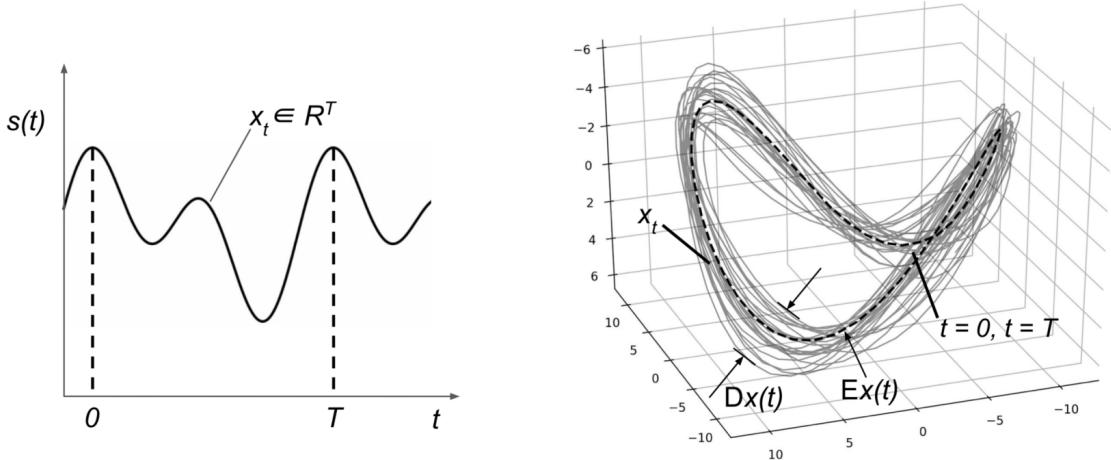


Figure 1: Time series's phase trajectory visualization

The rest of the paper covers the theory and application of our method on the real data. Firstly, the mathematical model will be introduced and basis search in the signal space problem will be stated and resolved. These results will enable us to propose a way of decomposing the signals and make a forecast. Simultaneously, features of obtained solutions will be examined and result in two theorems. Finally, tSSA and mentioned methods will be applied on two datasets: electricity consumption and meteorology observations. We will obtain signal's predictions and additive decompositions. For the latter a special metric will be formulated. All results will be backed with figures and discussion.

## 2 Problem statement

Let us have observed multidimensional time series  $\{x_i(t)\}_{i=1}^m$  generated by some *dynamical system*  $f$ . By that we mean a law of coordinates  $\mathbf{y} \in X$  evolution in discrete time:

$$\begin{aligned}\mathbf{y}(t+1) &= f(\mathbf{y}(t)), \quad t \in \mathbb{N} \\ \mathbf{y}(0) &= \mathbf{y}_0\end{aligned}$$

Here  $X$  is generally a high dimensional smooth manifold. Then, each trajectory of the system spawns our signals through unknown mapping  $\varphi : X \rightarrow \mathbb{R}^m$ :

$$\varphi(\mathbf{y}(t)) = \mathbf{x}_t \Leftrightarrow \begin{cases} \varphi_1(\mathbf{y}(t)) = x_1(t) \\ \dots \\ \varphi_m(\mathbf{y}(t)) = x_m(t) \end{cases}$$

After that, it is hypothesized that trajectories  $\mathbf{y}(t)$  actually lie in a smaller dimension manifold  $M \subset X$ . So now we should find an embedding of  $M$  into  $\mathbb{R}^L$  for some  $L$  and discover a basis in the image of the embedding. Having done that, we will obtain characterization of the initial system in terms of the standard linear space. The same will be equally applied to all  $x_i(t)$ .

## One signal case

Our future approach roots from SSA method and Takens's theorem [14]. Let's describe it in short. The theorem provides a solution for the stated problem in case of single series: any point  $\mathbf{y}(t) \in M$  we associate with the following vector:

$$(\varphi \circ f^{t-L+1}(\mathbf{y}(t)), \dots, \varphi \circ f(\mathbf{y}(t)), \varphi \circ \mathbf{y}(t))^T = (x(t-L+1), \dots, x(t-1), x(t))^T$$

It is called *delay vector* in time  $t$  and denoted as  $\overleftarrow{\mathbf{x}}_t$ . Their dimensionality  $L$  should satisfy  $L > 2 \cdot \dim(M)$ . Function  $\varphi(\cdot)$  should also meet some regularity conditions which are assumed to be fulfilled.

So, with signal  $x(t)$  of length  $N$  we build  $N - L + 1$  delay vectors. Corresponding embedding space  $\text{Lin}(\{\overleftarrow{\mathbf{x}}_t\})$ , so called *signal subspace*, must be low dimensional, meaning  $\text{Lin}(\{\overleftarrow{\mathbf{x}}_t\}) \subset \mathbb{R}^L$ . Orthonormed basis here is a SVD's  $U$ -component of *trajectory matrix*  $\mathbf{H}_x$  which is build from delay vectors:

$$\mathbf{H}_x = [\overleftarrow{\mathbf{x}}_1 \dots \overleftarrow{\mathbf{x}}_{N-L+1}]$$

## Series interconnection and tSSA method

Теперь обобщим данный подход на несколько временных рядов.

Now let's generalize previous results for the multiple series.

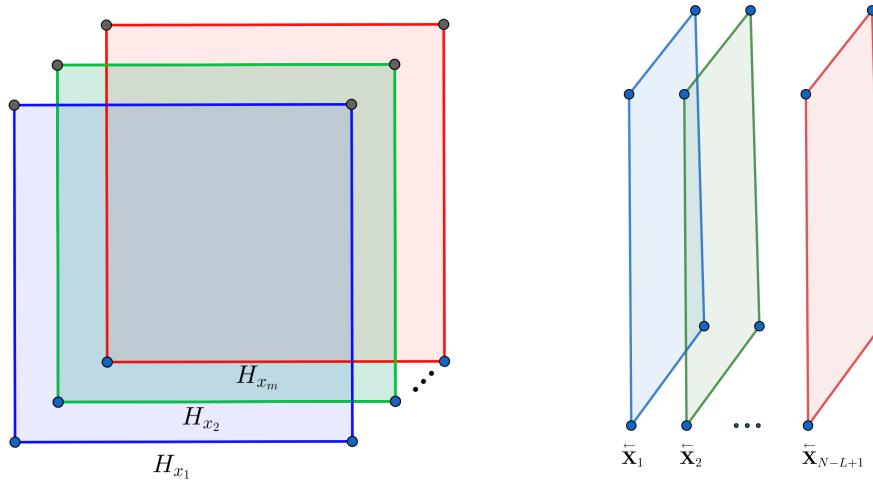


Figure 2: Two views on trajectory tensor. The left is in terms of signals' trajectory matrices  $\{x_i(t)\}_{i=1}^m$ . The right is in terms of delay matrices.

In this case  $\varphi(\cdot)$  is multidimensional. So instead of the single delay vector we should consider their set for all  $m$  signals in time  $t$ . That means the embedding image now contains so called *delay matrices*  $(\overleftarrow{\mathbf{x}}_{1t} \dots \overleftarrow{\mathbf{x}}_{mt}) := \overleftarrow{\mathbf{X}}_t$ . Next, instead of trajectory matrix we should introduce *trajectory tensor*  $\mathbf{T}$ . It is constructed by putting delay matrices along the second dimension

of the tensor, fig. 2 (left). This procedure is equivalent to the one for  $\mathbf{H}_x$ . But our tensor has an important property:  $\mathbf{T}$  can also be built as an alignment of signals' trajectory matrices  $\mathbf{H}_{x_i}$  along the third dimension, fig. 2 (right).

Then, reasoning analogues to the one signal case we obtain *signals' set subspace* as a linear span of delay matrices. But we were supposed to built the subspace for each signal individually.

It's time to give a definition for signals' interrelation in our model's scope.

**Definition.** We call a set of time series *interconnected* if they all share the common subspace with the same basis.

Now we are ready to find this shared elements for signals. Note that to do so we just need to decompose each signal's trajectory matrix through the same set of factors. Finally, let's apply CPD to our trajectory tensor and have a look at its slices by third dimension.

$$\mathbf{T} = \sum_{i=1}^r \mathbf{a}_i \otimes \mathbf{b}_i \otimes \mathbf{c}_i \Leftrightarrow \begin{cases} \mathbf{H}_{x_1} = \sum_{i=1}^r \boldsymbol{\sigma}_{x_1}(i) \cdot \mathbf{a}_i \mathbf{b}_i^\top \\ \mathbf{H}_{x_2} = \sum_{i=1}^r \boldsymbol{\sigma}_{x_2}(i) \cdot \mathbf{a}_i \mathbf{b}_i^\top \\ \dots \\ \mathbf{H}_{x_m} = \sum_{i=1}^r \boldsymbol{\sigma}_{x_m}(i) \cdot \mathbf{a}_i \mathbf{b}_i^\top \end{cases} \quad (1)$$

What have we got? CPD is defined by tensor rank  $r$  and a set of vectors which we can pack in matrices:  $A = [\mathbf{a}_1 \dots \mathbf{a}_r]; B = [\mathbf{b}_1 \dots \mathbf{b}_r]; C = [\mathbf{c}_1 \dots \mathbf{c}_r]$ . Row  $k$  of matrix  $C$  is denoted as  $\boldsymbol{\sigma}_{x_k}$ . It has the sense of singular values but can be negative.

After, from CPD definition as a (1) kind decomposition with minimal  $r$  we can derive that  $A, B, C$  are full-rank matrices. So, signals' shared subspace is  $\text{Lin}(\{\mathbf{a}_i\})$  with unified basis  $\{\mathbf{a}_i\}_{i=1}^r$  which is not orthonormed in general. Then, back in the problem statement we assumed dimensionality of the embedding to be rather small. Therefore relation  $r \ll L$  should be true.

Lastly, let's note that rows of  $\mathbf{H}_{x_k}$  are also delay vectors but have another dimension of  $N - L + 1$ . So the shared subspace for them would be  $\text{Lin}(\{\mathbf{b}_i\})$ . This result proves our approach to be consistent with choice of  $\overleftarrow{\mathbf{x}}_{k,t}$ . As a remark, mSSA lacks such feature.

What to the signals' link, it is regulated by  $\boldsymbol{\sigma}_{x_k}$ . For instance, if all the rows have zero elements in non-overlapping positions then our built subspace breaks down into a direct sum of individual signal's subspaces. It is total link absence case. The exact opposite situation is when  $\boldsymbol{\sigma}_{x_k}$  are non-zero for all series.

## Time series decomposition

We can break series into several components using the conception: *decomposition of trajectory matrix  $\mathbf{H}_{x_k}$  defines decomposition of related signal*. Its first part implies representation of  $\mathbf{H}_{x_k}$  as a sum of factors (1). This technique comes from SSA. Because all matrices are similarly decomposed, we can focus on the task for a single series.

Besides, we should point out the main feature of trajectory matrices — *hankelness*. That simply means equal elements along each anti-diagonal. Every series of length  $N$  bijectionally corresponds to hankel matrix of size  $L \times (N - L + 1)$ .

Now, decomposition procedure is following: we assume factors of  $\mathbf{H}_{x_k}$  to be able to pack in  $s$  groups in a special way. That means if we sum all factors within each group the result matrices  $C_1, \dots, C_s$  will be hankel. This situation would perfectly correspond to representation of  $x_k(t)$  as a sum of  $s$  signals. Unfortunately, it is almost infeasible even with trivial series [11], so every  $C_i$  should be additionally *hankelized*. This operation average each matrix's anti-diagonal so it becomes hankel. Let's denote the operator as  $\text{Hankel}(\cdot)$ .

The whole algorithm can be written as a chain of identical expressions:

$$\begin{aligned}
\mathbf{H}_{x_k} &\stackrel{1}{=} \sum_{i=1}^r \boldsymbol{\sigma}_{x_k}(i) \cdot \mathbf{a}_i \mathbf{b}_i^\top \stackrel{2}{=} \sum_{i \in \mathbb{I}_1} \boldsymbol{\sigma}_{x_k}(i) \cdot \mathbf{a}_i \mathbf{b}_i^\top + \dots + \sum_{i \in \mathbb{I}_s} \boldsymbol{\sigma}_{x_k}(i) \cdot \mathbf{a}_i \mathbf{b}_i^\top \stackrel{3}{=} \\
&\stackrel{3}{=} C_1 + \dots + C_s \stackrel{4}{=} \text{Hankel}(C_1) + \dots + \text{Hankel}(C_s) \Leftrightarrow x_k(t) = c_1(t) + \dots + c_s(t) \quad (2)
\end{aligned}$$

Here  $\mathbb{I}_1 \sqcup \dots \sqcup \mathbb{I}_s = \{1, \dots, r\}$  — chosen groups of indices. Actually, the forth equality needs some extra validation. To make it, first consider hankel operator's property:

**Lemma 1.** *Hankel( $\cdot$ ) operator is linear.*

*Proof.* Having matrices  $A$  and  $B$  with the same size, let's look at their elements from one anti-diagonal  $a_1, \dots, a_n$  and  $b_1, \dots, b_n$ . Then this anti-diagonal in  $\text{Hankel}(A + B)$  is written as  $\frac{1}{n} \sum_{i=1}^n (a_i + b_i) = \frac{1}{n} \sum_{i=1}^n a_i + \frac{1}{n} \sum_{i=1}^n b_i$ , resulting in a sum of  $\text{Hankel}(A)$  and  $\text{Hankel}(B)$  anti-diagonals.

Secondly, having a scalar  $\alpha$  it is also trivial that  $\frac{1}{n} \sum_{i=1}^n \alpha \cdot a_i = \alpha \frac{1}{n} \sum_{i=1}^n a_i$ . Then we finally derive  $\text{Hankel}(\alpha A) = \alpha \text{Hankel}(A)$ .  $\square$

Returning to forth equality in (2), we can combine this lemma and the fact that  $\mathbf{H}_{x_k}$  is itself hankel to get a complete justification. This outputs in the correctness of series decomposition procedure for any chosen grouping of factors.

## Optimal decomposition and its computational complexity

Несмотря на корректность имеющегося алгоритма, хочется группировать факторы так, что каждая матрица  $C_i$  была как можно более "*<ганселевой>*". Это соответствует сформулированному принципу о связи разложения матриц и сигналов, а также повышает интерпретируемость компонент. Опишем задачу формально и исследуем её.

Рассмотрим первое равенство в (2), а также применим к нему оператор ганселизации:

$$\begin{cases} \mathbf{H}_{x_k} = \sum_{i=1}^r \boldsymbol{\sigma}_{x_k}(i) \cdot \mathbf{a}_i \mathbf{b}_i^\top \\ \mathbf{H}_{x_k} = \sum_{i=1}^r \text{Hankel}(\boldsymbol{\sigma}_{x_k}(i) \cdot \mathbf{a}_i \mathbf{b}_i^\top) \end{cases}$$

Ещё раз напомним, что  $\mathbf{H}_{x_k}$  ганселева по построению. Вычтем из первого второе и обозначим  $H_i = \boldsymbol{\sigma}_{x_k}(i) \cdot \mathbf{a}_i \mathbf{b}_i^\top - \text{Hankel}(\boldsymbol{\sigma}_{x_k}(i) \cdot \mathbf{a}_i \mathbf{b}_i^\top)$ , что соответствует невязке между исходной матрицей и её ганселизацией. Т.о. получим:

$$H_1 + \dots + H_r = 0 \Leftrightarrow H_r = - \sum_{j=1}^{r-1} H_j \quad (3)$$

Теперь группировка факторов и ганселевость итоговых матриц  $C_i$  (см. предыдущий подраздел) становится равносильна следующему:  $\sum_{k \in \mathbb{I}_i} H_k = 0, \forall i \in 1, \dots, s$ . Т.е. матрицы невязок суммируются в ноль внутри каждой группы. Далее, несложно заметить, что если существует разложение на  $s$  групп, то к нему можно прийти, разделяя имеющиеся группы факторов на две последовательно. Поэтому остановимся на этой упрощённой задаче.

Из (3) видно, что все матрицы невязок уже суммируются в 0, т.е. группу для  $H_r$  можно определить, сгруппировав оставшиеся невязки. Введём для каждой из них переменную-индикатор  $\beta_j \in \{0, 1\}$ , показывающую, к какой из двух групп относится невязка.

Итак, ещё раз перечислим все условия.  $H_i$  в одной группе суммируются в 0; также считаем, что все они ненулевые (иначе мы уже имеем искомую группировку), т.е. в группе должно быть хотя бы две невязки. Т.о. задача поиска группировки свелась к следующей:

$$\begin{cases} \sum_{j=1}^{r-1} \beta_j H_j = 0 \\ \beta_j \in \{0, 1\}, \forall j \in 1, \dots, r \\ \sum_{i=1}^{r-1} \beta_i \geq 2 \end{cases} \quad (4)$$

Для придания задаче более знакомого вида, во-первых, векторизуем каждую  $H_i$ , и соберём все эти векторы в одну матрицу  $\Lambda$ , а также введём  $\boldsymbol{\beta} = (\beta_1 \dots \beta_{r-1})$ . Вспомним, что на практике мы не ждём идеальной ситуации суммирования в 0 матриц невязок (ганселевости  $C_i$ , см. предыдущий подраздел), нам достаточно сделать сумму как можно более близкой к нулю. Это приводит нас к конечной формулировке:

$$\begin{cases} \|\Lambda \boldsymbol{\beta}\| \rightarrow \min_{\boldsymbol{\beta}} \\ \beta_j \in \{0, 1\}, \forall j \in 1, \dots, r \\ \sum_{i=1}^{r-1} \beta_i \geq 2 \end{cases} \quad (5)$$

Назовём решение этой системы *оптимальной декомпозицией ряда*. Она представляет из себя задачу наименьших квадратов с целочисленными ограничениями (ILS), которая, как показано в [15], является NP-сложной. Все проведённые рассуждения приведём в виде:

**Theorem 1.** *Оптимальная декомпозиция ряда, в смысле поиска группировки факторов разложения траекторной матрицы ряда, является NP-сложной задачей.*

Несмотря на полученный результат, существуют методы и эвристики для сведения ILS к эффективно вычислим задачам, см. [16].

## Прогноз для набора рядов

Получение предсказания нашей модели сводится к восстановлению компонент векторов задержек каждого входного ряда. Т.к. мы построили общий для всех сигналов базис  $\{\mathbf{a}_i\}_{i=1}^r$ , то каждый сигнал может обрабатываться независимо.

Если бы  $\{\mathbf{a}_i\}_{i=1}^r$  был ортогональным, то мы бы могли применить технику прогнозирования SSA [11]. Мы не имеем такой гарантии, поэтому процедуре необходимо модифицировать.

Рассмотрим входной ряд  $x(t)$  и построим для него прогноз на один шаг вперёд. Вектор задержки  $\overleftarrow{\mathbf{x}}_{N+1} = (x(N-L) \dots x(N+1))^T$ , в котором последняя компонента неизвестна, лежит в пространстве  $\text{Lin}(\{\mathbf{a}_i\})$ . В терминах вводимой ранее матрицы  $A \in \mathbb{R}^{L \times r}$  полного ранга (см. раздел [Взаимосвязь рядов и метод tSSA](#)) это записывается как:

$$\begin{aligned} \overleftarrow{\mathbf{x}}_{N+1} = A\boldsymbol{\lambda} &\Leftrightarrow \begin{cases} \mathbf{x}_{kn} = A_{kn}\boldsymbol{\lambda} \\ x(N+1) = \mathbf{a}_{pr}^T \boldsymbol{\lambda} \end{cases}, \text{ где} \\ A &= \left( \frac{A_{kn}}{\mathbf{a}_{pr}^T} \right) \\ \overleftarrow{\mathbf{x}}_{N+1} &= (\mathbf{x}_{kn} \ x(N+1))^T \end{aligned} \quad (6)$$

Здесь  $\boldsymbol{\lambda} \in \mathbb{R}^r$ ; вектор задержки и матрица представлены в блочном виде, разделяющим известную часть сигнала (known) и его будущее значение (predicted). Для построения прогноза необходимо определить  $\boldsymbol{\lambda}$ .

Матрица имеющейся СЛАУ  $A_{kn} \in \mathbb{R}^{(L-1) \times r}$  имеет ранг не менее  $r-1$ , но мы предположим, что удаление одной строчки оставит ранг полным. Как показывает практика, это необременительное

условие. Сама система, в силу предположения модели  $r \ll L$ , является переопределённой. Значит её решение разумно искать только в смысле наименьших квадратов:  $\lambda = (A_{kn}^T A_{kn})^{-1} A_{kn}^T \mathbf{x}_{kn}$ . В итоге прогноз модели даётся:

$$x(N+1) = \mathbf{a}_{pr}^T (A_{kn}^T A_{kn})^{-1} A_{kn}^T \mathbf{x}_{kn} \quad (7)$$

Здесь  $\mathbf{a}_{pr}^T (A_{kn}^T A_{kn})^{-1} A_{kn}^T = \mathbf{d}^T$  — вектор-строка, и, вычислив её единожды, можно последовательно строить предсказания на произвольный горизонт вперёд.

Теперь перепишем (7) в терминах  $x(t)$ :

$$x(t) = \sum_{i=1}^{L-1} d_i \cdot x(t-i)$$

Это выражение приводит нас к окончательному выводу о характере прогноза:

**Theorem 2.** *Модель прогнозируемого методом tSSA ряда — авторегрессионная AR(L-1). Поведение прогноза полностью определяется коэффициентами авторегрессии  $d_i$ , найденными в (7).*

### 3 Вычислительный эксперимент

#### Цели и постановка

В данном разделе мы рассмотрим применение метода tSSA для декомпозиции и прогноза нескольких многомерных временных рядов, а также сравним его с другими моделями, упомянутыми во введении: mSSA, VAR, RNN. Упомянем, что mSSA по теории и применению во многом схожа с нашим методом, с её помощью можно решать обе упомянутые задачи. Две другие модели способны только к предсказанию.

Для измерения **качества декомпозиции** введём метрики:

**Definition.** *Абсолютной ошибкой ганкелизации* матрицы  $M$  назовём

$$\text{AHE} = \|M - \text{Hankel}(M)\|_2$$

**Definition.** *Относительной ошибкой ганкелизации* матрицы  $M$  назовём

$$\text{RHE} = \frac{\text{AHE}}{\|M\|_2}$$

AHE можно представлять как суммарное стандартное отклонение антидиагоналей матрицы. RHE её более интерпретируемая модификация, которая и будет использоваться далее.

Напомним, что разложение ряда на компоненты эквивалентно группировке факторов разложения траекторной матрицы сигнала, т.е. получению матриц  $C_i$  (см. раздел [Декомпозиция набора рядов](#)). Именно для них мы и будем рассчитывать RHE, а также среднее RHE по одному ряду —  $\overline{\text{RHE}}_{ts}$ , и среднее RHE всех рядов датасета —  $\overline{\text{RHE}}$ .

Для измерения **качества прогноза** ряда будем использовать стандартные метрики: среднеквадратическую ошибку  $\text{MSE}_{ts}$  и среднюю абсолютную процентную ошибку  $\text{MAPE}_{ts}$ , а также их усреднения по всему датасету с обозначениями  $\overline{\text{MSE}}$  и  $\overline{\text{MAPE}}$ .

В качестве сигналов будут рассмотрены два набора данных: потребление электроэнергии и её цена за единицу мощности (рис. 3); а также три ряда, характеризующие погоду г. Берлин в период с 1980 по 1990 год: средние за день температура, осадки и атмосферное давление (рис. 4). Мы предполагаем наличие связи рядов в каждом наборе (на основе экономических/физических причин), т.к. это важное условие применения нашего метода

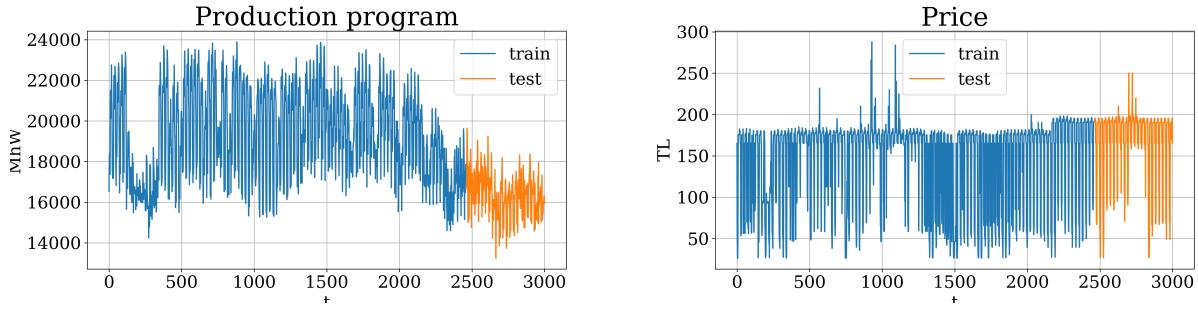


Figure 3: Многомерный ряд потребления электроэнергии и его цены за единицу мощности

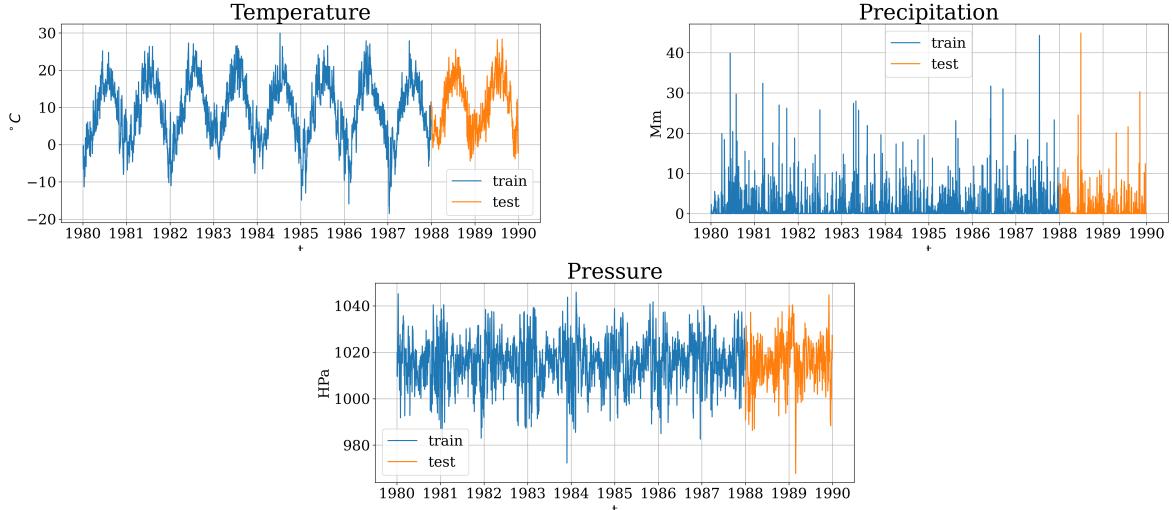


Figure 4: Многомерный ряд физических величин: температуры, осадков и давления, измеренных на метеостанции г. Берлин

(см. Взаимосвязь рядов и метод tSSA). Каждый ряд имеет длину  $\approx 3 \cdot 10^3$ , из которых 20% — отложенная выборка для оценки качества прогноза.

Опишем нюансы применения метода tSSA. Во-первых, вычисление канонического ранга тензора — NP-трудная задача [17], поэтому каждый раз ранг разложения будет указываться явно. Во-вторых, используемый алгоритм CP-разложения — ALS [18], имеет ошибку аппроксимации, из-за чего возникает неточность разложения траекторных матриц. Относительные величины этих невязок будут указаны ниже. В-третьих, для поиска декомпозиции рядов будет решаться задача ILS (5), в которой матрица имеет высокую строковую размерность. Для ускорения мы сократим её до нескольких сотен, что всё ещё будет много больше размерности искомого вектора параметров. Используемый солвер — SCIP [19].

Наконец, мы зафиксируем единый для всех моделей параметр, отвечающий за длину предуславливания методов на историю рядов, который обозначался за  $L$ . Для tSSA и mSSA он был описан в начале раздела 2; в RNN это можно считать длиной энкодера; в VAR это максимальная длина истории ряда, на которую производится авторегрессия. Везде ниже  $L = 500$ .

## Результаты и обсуждение

Начнём с задачи прогнозирования. На рис. 5 и 6 отображена зависимость метрик качества прогноза tSSA в зависимости от ранга CP-разложения. Точка минимума и характер изменения MSE и MAPE совпадают. Наблюдается эффект переобучения при повышении значения ранга, особенно резкое ухудшение метрик у данных погоды. При этом ошибка CP-разложения монотонно убывает, см. рис. 7.

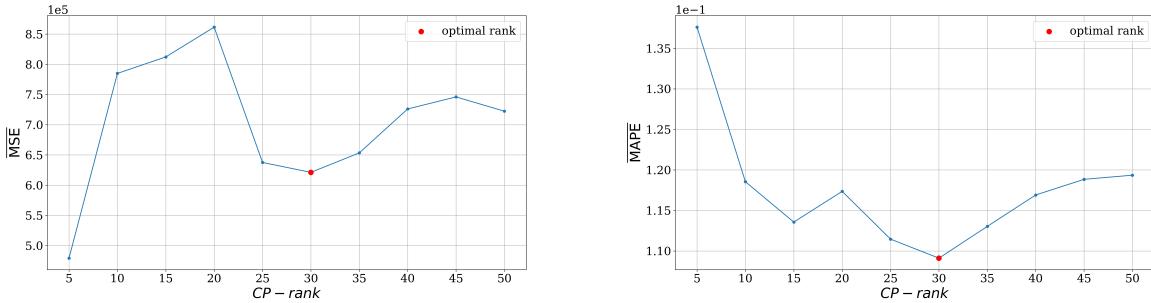


Figure 5: Данные электроэнергии. Значения метрик  $\overline{\text{MSE}}$  и  $\overline{\text{MAPE}}$  предсказания tSSA для разных рангов СР-разложения. Отдельно выделен оптимальный ранг.

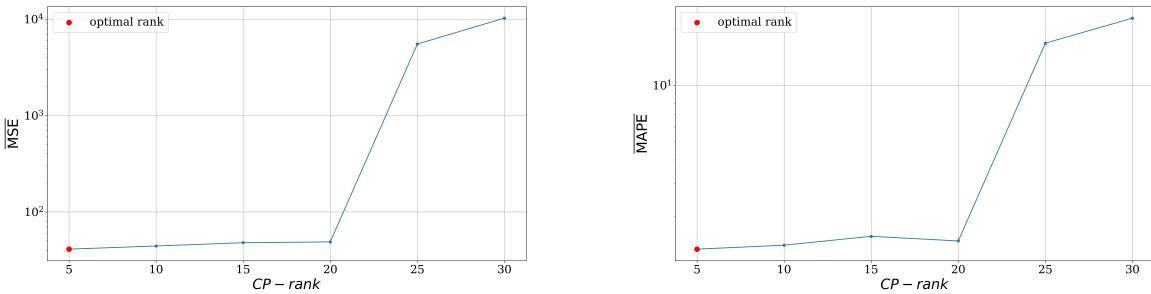


Figure 6: Данные погодных условий. Значения метрик  $\overline{\text{MSE}}$  и  $\overline{\text{MAPE}}$  предсказания tSSA для разных рангов СР-разложения. Отдельно выделен оптимальный ранг.

Графики наилучшего прогноза tSSA приведены на рис. 8 и 9, где наглядно видны различия метода с малым и большим рангом. В первом случае авторегрессионная модель прогноза большего порядка (см. теорему раздела [Прогноз для набора рядов](#)), что позволяет точно аппроксимировать ряд. В случае же погоды большой порядок приводил к нестабильности и неограниченному росту, но при уменьшении степеней свободы возможно аппроксимировать сигналы только в среднем.

Итоговые метрики качества прогноза всех моделей — в табл. 1 и 2. Наш метод показал наилучшие результаты, хотя mSSA был близок по точности. Метод VAR оказался нестабилен на выбранном горизонте прогнозирования, а RNN обучался в константную функцию.

Table 1: Метрики моделей на прогнозировании данных электроэнергии

	tSSA	mSSA	VAR	RNN
$\overline{\text{MSE}}_{\text{Production}}, 10^6$	1.24	1.51	7.81	2.70
$\overline{\text{MSE}}_{\text{Price}}, 10^3$	0.88	1.03	4.85	30.0
$\overline{\text{MSE}}, 10^6$	<b>0.62</b>	0.75	3.91	135.00
$\overline{\text{MAPE}}_{\text{Production}}$	0.054	0.060	0.137	0.999
$\overline{\text{MAPE}}_{\text{Price}}$	0.164	0.170	0.360	1.004
$\overline{\text{MAPE}}$	<b>0.109</b>	0.115	0.249	1.002

Теперь рассмотрим задачу декомпозиции. На рис. 10 представлены зависимости качества разложения от значения СР-ранга. В обоих случаях  $\overline{\text{RHE}}$  достаточно быстро достигает минимума, после чего не меняется или растёт. Вспоминая аналогичный результат для прогнозирования, можно сделать вывод, что метод tSSA достигает наибольшей обобщающей способности на небольших канонических рангах траекторных тензоров данных. В свою очередь, этот ранг пропорционален размерности собственного пространства сигналов (см.

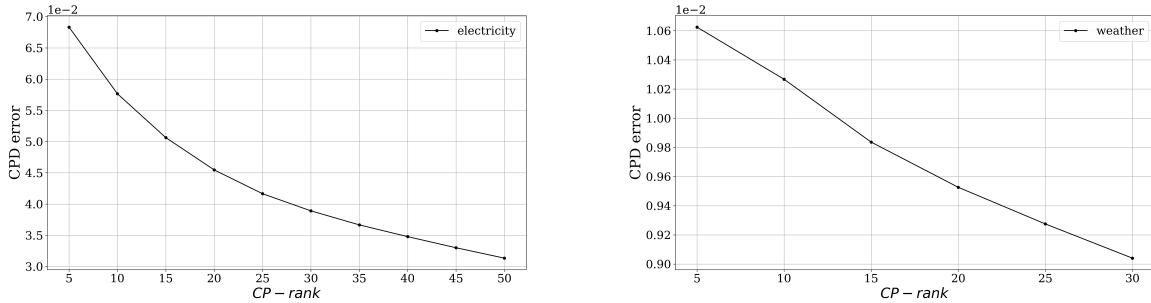


Figure 7: Относительные ошибки CPD-аппроксимации траекторных тензоров в зависимости от ранга разложения. Слева — электроэнергии, справа — погоды.

Table 2: Метрики моделей на прогнозировании данных погоды

	<i>tSSA</i>	<i>mSSA</i>	<i>VAR</i>	<i>RNN</i>
$\overline{\text{MSE}}_{\text{Temp}}$	14.16	20.12	40.67	-
$\overline{\text{MSE}}_{\text{Prec}}$	11.02	11.23	22.84	-
$\overline{\text{MSE}}_{\text{Pres}}$	98.02	101.23	156.75	-
$\overline{\text{MSE}}$	<b>41.067</b>	44.193	73.420	-
$\overline{\text{MAPE}}_{\text{Temp}}$	0.850	0.626	0.468	-
$\overline{\text{MAPE}}_{\text{Prec}}$	3.151	2.834	4.637	-
$\overline{\text{MAPE}}_{\text{Pres}}$	0.007	1.254	3.983	-
$\overline{\text{MAPE}}$	<b>1.336</b>	1.571	3.029	-

начало раздела 2). Из этого заключаем о выполнении главной теоретической предпосылки для применения нашего метода.

На рис. 11 и 12 представлены результаты наилучшей найденной декомпозиции рядов на две компоненты методом tSSA. В силу вычислительной сложности задачи ILS, разложение на большее количество компонент не производится. Из графиков видно, как сильно общий базис сигналов влияет на их декомпозицию: для генерации электроэнергии и её цены компоненты получились почти идентичными, с точностью до смещения и масштаба. Аналогичная ситуация для температуры и осадков в данных погоды, а для атмосферного давления разложение получилось другим. Далее, на рис. 12 метод выделил для каждого сигнала по две периодические составляющие: одна из них большой амплитуды, другая много меньшей. Небольшой шум остался в каждой компоненте. На рис. 11 метод выделил по два высокочастотных колебания разной амплитуды и смещения.

Для сравнения на рис. 13 и 14 приведены результаты декомпозиции методом mSSA на пять компонент. Для их получения не было необходимости в решении ILS или настраивания СР-ранка, но они подбирались вручную на основе близости сингулярных чисел некоторой матрицы, см. [11]. Тем не менее компоненты получились более интерпретируемыми и подробными: метод выделил несколько трендов и гармоник. Также, как видно из табл. 3 и 4, mSSA немного выиграл по точности. Т.о. вычислительная сложность декомпозиции tSSA является главной проблемой метода и требует разработки собственных эвристик, которые, к слову, у mSSA уже имеются [11].

## 4 Заключение

Тензорный метод tSSA был разработан для прогноза и декомпозиции набора временных рядов, обработка которых требует учёта фактора взаимосвязанности. Его главное достоинство

Table 3: Метрики моделей на декомпозиции данных электроэнергии

	tSSA	mSSA
$\overline{RHE}_{Production}$	0.507	0.308
$\overline{RHE}_{Price}$	0.511	0.31
$\overline{RHE}$	0.509	<b>0.309</b>

Table 4: Метрики моделей на декомпозиции данных погоды

	tSSA	mSSA
$\overline{RHE}_{Temp}$	0.512	0.467
$\overline{RHE}_{Prec}$	0.508	0.538
$\overline{RHE}_{Pres}$	0.542	0.502
$\overline{RHE}$	0.521	<b>0.502</b>

— проста в использовании: имеется всего два настраиваемых параметра, для которых не требуется изощрённых процедур обучения. Между тем, теория метода опирается на весьма общую модель динамических систем и предъявляет к данным лишь возможность маломерного представления. При применении на реальных датасетах удалось установить справедливость теоретических требований алгоритма, а также показать качество выше нейросетевых и статистических подходов. Главным вызовом для tSSA является результат о NP-сложности поиска оптимальной декомпозиции рядов. Нахождение способов решения этой проблемы или поиск другого пути для построения разложения — основные направления дальнейшего развития метода.

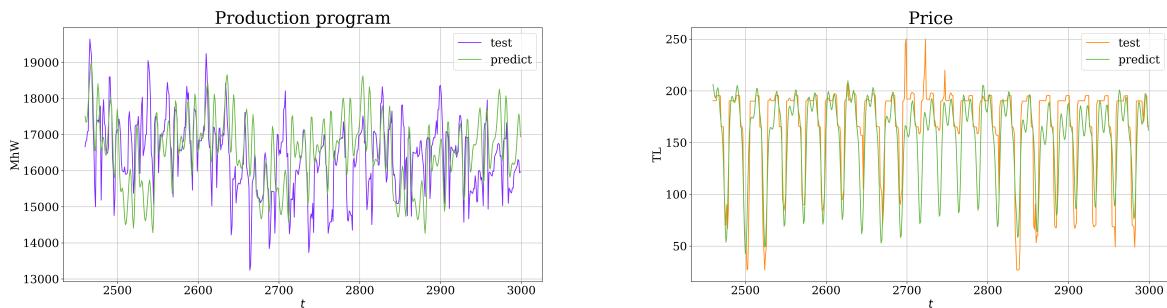


Figure 8: Предсказание данных электроэнергии методом tSSA. CPD-ранг = 30

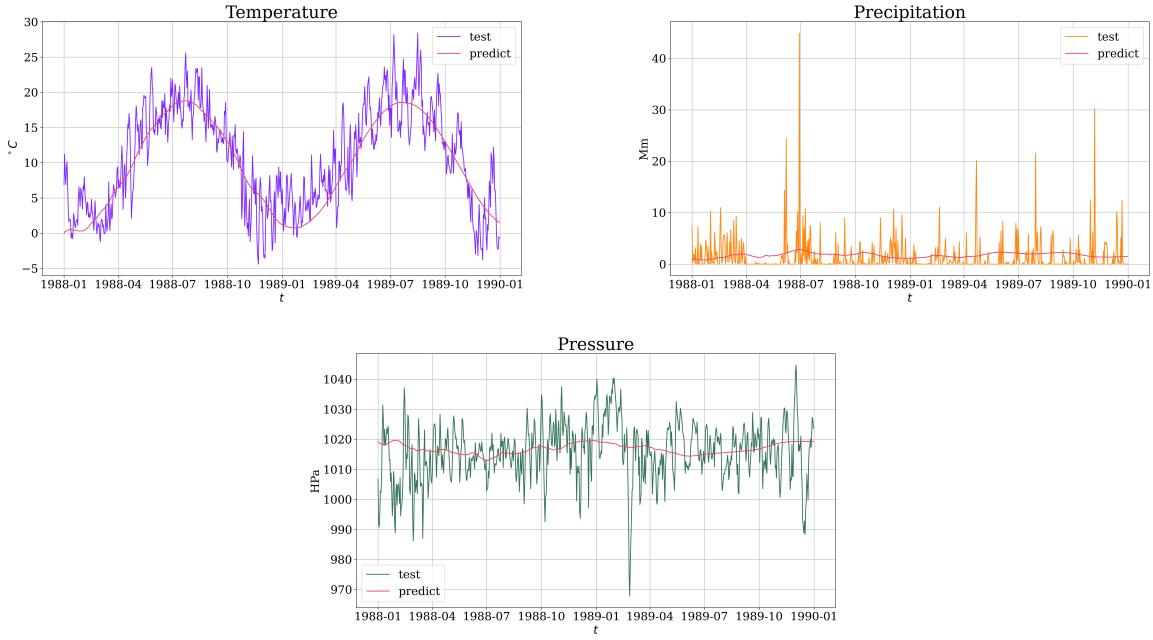


Figure 9: Предсказание данных погоды методом tSSA. CPD-ранг = 5

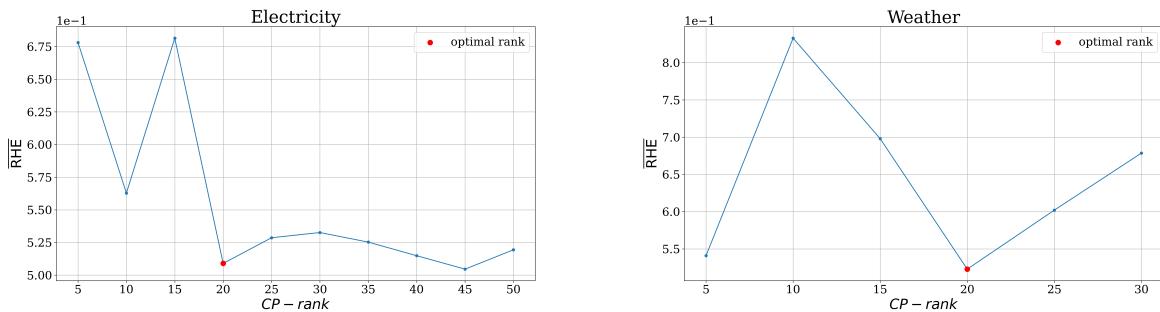


Figure 10: Зависимость метрики  $\overline{RHE}$  от ранга CP-разложения. Слева — для данных электроэнергии, справа — погоды. Отдельно выделен оптимальный ранг.

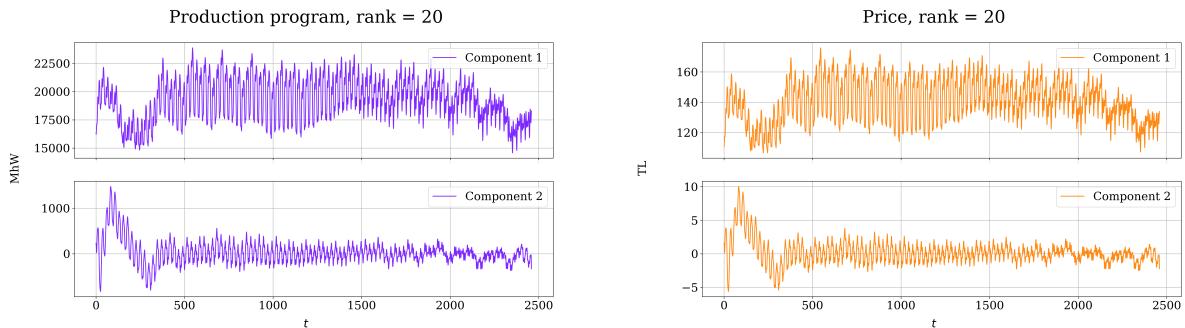


Figure 11: Разложение рядов на две компоненты методом tSSA. Данные электроэнергии. CPD-ранг = 20

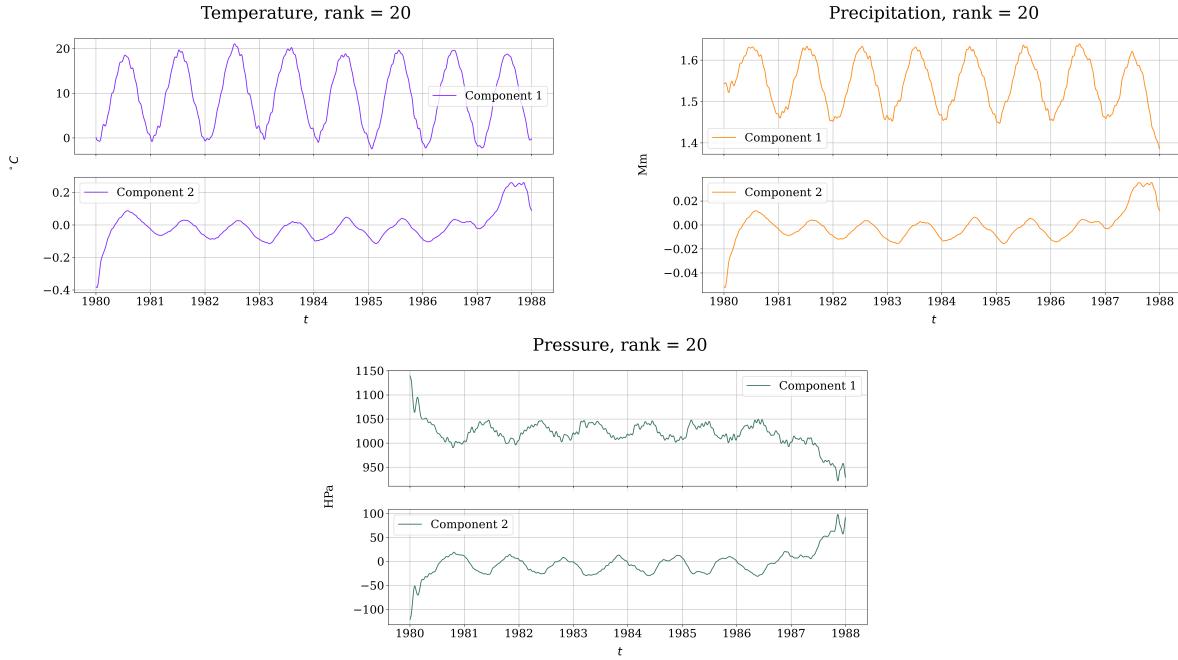


Figure 12: Разложение рядов на две компоненты методом tSSA. Данные погоды. CPD-ранг = 20

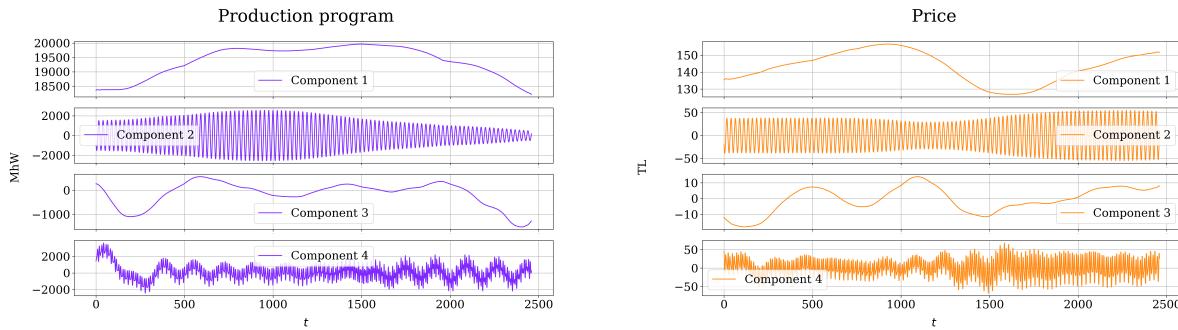


Figure 13: Разложение рядов на компоненты методом mSSA. Данные электроэнергии.

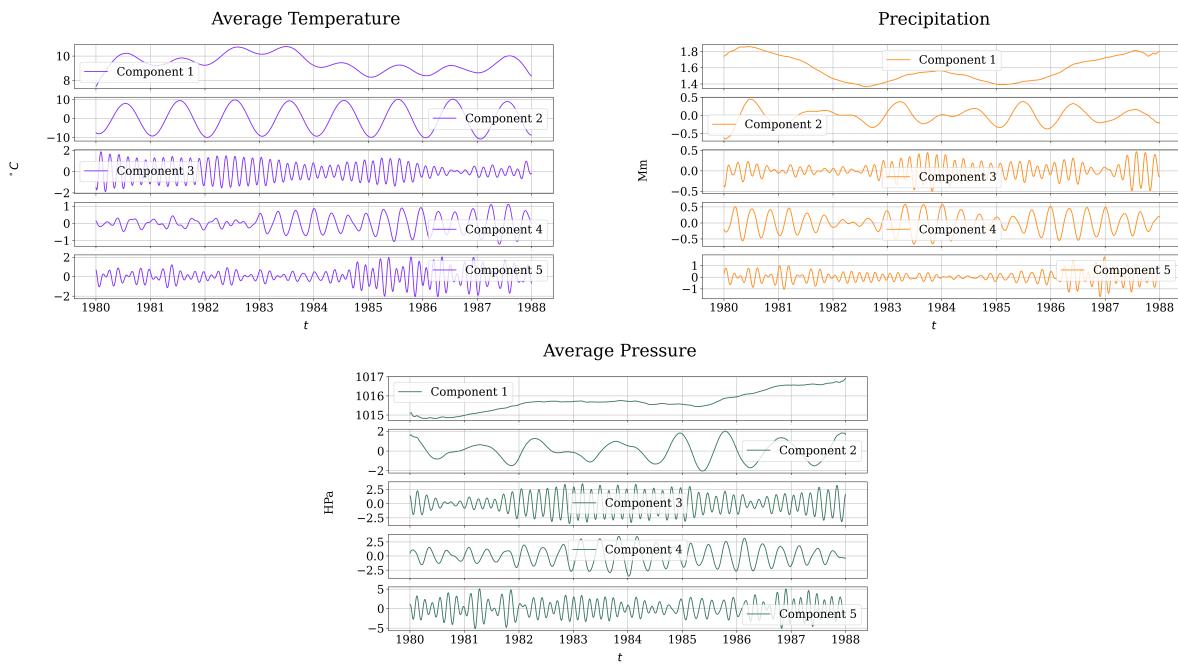


Figure 14: Разложение рядов на компоненты методом mSSA. Данные погоды.

## References

- [1] W. Enders and John Wiley & Sons. *Applied Econometric Time Series*. Wiley series in probability and statistics. Wiley, 2010. ISBN: 9788126543915. URL: <https://books.google.ru/books?id=vzJ0CgAAQBAJ>.
- [2] Estela Dagum and Silvia Bianconcini. *Seasonal Adjustment Methods and Real Time Trend-Cycle Estimation*. Aug. 2016. ISBN: 978-3-319-31820-2.
- [3] Robert B. Cleveland et al. “STL: A Seasonal-Trend Decomposition Procedure Based on Loess (with Discussion)”. In: *Journal of Official Statistics* 6 (1990), pp. 3–73.
- [4] Robin John Hyndman and George Athanasopoulos. *Forecasting: Principles and Practice*. English. 2nd. Australia: OTexts, 2018.
- [5] Granville Tunnicliffe Wilson. “Time Series Analysis: Forecasting and Control,5th Edition, by George E. P. Box, Gwilym M. Jenkins, Gregory C. Reinsel and Greta M. Ljung, 2015. Published by John Wiley and Sons Inc., Hoboken, New Jersey, pp. 712. ISBN: 978-1-118-67502-1”. In: *Journal of Time Series Analysis* 37 (Mar. 2016), n/a–n/a. DOI: [10.1111/jtsa.12194](https://doi.org/10.1111/jtsa.12194).
- [6] Vito Volterra. “Математическая теория борьбы за существование”. In: *UFN* 8.1 (1928), pp. 13–34.
- [7] Sepp Hochreiter and Jürgen Schmidhuber. “Long Short-Term Memory”. In: *Neural Computation* 9.8 (Nov. 1997), pp. 1735–1780. ISSN: 0899-7667. DOI: [10.1162/neco.1997.9.8.1735](https://doi.org/10.1162/neco.1997.9.8.1735). eprint: <https://direct.mit.edu/neco/article-pdf/9/8/1735/813796/neco.1997.9.8.1735.pdf>. URL: <https://doi.org/10.1162/neco.1997.9.8.1735>.
- [8] Ahmed Tealab. “Time series forecasting using artificial neural networks methodologies: A systematic review”. In: *Future Computing and Informatics Journal* 3.2 (2018), pp. 334–340. ISSN: 2314-7288. DOI: <https://doi.org/10.1016/j.fcij.2018.10.003>. URL: <https://www.sciencedirect.com/science/article/pii/S2314728817300715>.
- [9] Dimitrios Asteriou and Stephen Hall. “Vector Autoregressive (VAR) Models and Causality Tests”. In: Jan. 2016, pp. 333–346. ISBN: 978-1-137-41546-2. DOI: [10.1057/978-1-137-41547-9\\_15](https://doi.org/10.1057/978-1-137-41547-9_15).
- [10] Arnold Zellner. “An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias”. In: *Journal of the American Statistical Association* 57.298 (1962), pp. 348–368. DOI: [10.1080/01621459.1962.10480664](https://doi.org/10.1080/01621459.1962.10480664). eprint: <https://www.tandfonline.com/doi/pdf/10.1080/01621459.1962.10480664>. URL: <https://www.tandfonline.com/doi/abs/10.1080/01621459.1962.10480664>.
- [11] D. Stepanov and N. Golyandina. “SSA-based approaches to analysis and forecast of multi-dimensional time series”. In: *Proceedings of the 5th St.Petersburg Workshop on Simulation*. 2005, pp. 293–298.
- [12] Hamid KRIM. “Two Decades of Array Signal Processing Research”. In: *IEEE Signal Processing Magazine* 0 (1994), pp. 67–94. URL: <https://cir.nii.ac.jp/crid/1572261550523548160>.
- [13] Hossein Hassani and Rahim Mahmoudvand. “Multivariate singular spectrum analysis: A general view and new vector forecasting approach”. In: *International Journal of Energy and Statistics* 01 (Apr. 2013). DOI: [10.1142/S2335680413500051](https://doi.org/10.1142/S2335680413500051).
- [14] Floris Takens. “Detecting Strange Attractors in Turbulence”. In: *Dynamical Systems and Turbulence, Warwick 1980*. Ed. by David Rand and Lai-Sang Young. Vol. 898. Lecture Notes in Mathematics. Berlin: Springer, 1981. Chap. 21, pp. 366–381. ISBN: 978-3-540-11171-9. DOI: [10.1007/bfb0091924](https://doi.org/10.1007/bfb0091924). URL: <http://dx.doi.org/10.1007/bfb0091924>.

- [15] P. van Emde-Boas. *Another NP-complete partition problem and the complexity of computing short vectors in a lattice*. Report. Department of Mathematics. University of Amsterdam. Department, Univ., 1981.
- [16] Erik Grafarend, Silvelyn Zwanzig, and Joseph Awange. “Integer Least Squares”. In: *Applications of Linear and Nonlinear Models: Fixed Effects, Random Effects, and Total Least Squares*. Cham: Springer International Publishing, 2022, pp. 579–603. ISBN: 978-3-030-94598-5. DOI: [10.1007/978-3-030-94598-5\\_16](https://doi.org/10.1007/978-3-030-94598-5_16). URL: [https://doi.org/10.1007/978-3-030-94598-5\\_16](https://doi.org/10.1007/978-3-030-94598-5_16).
- [17] Johan Håstad. “Tensor rank is NP-complete”. In: *Journal of Algorithms* 11.4 (1990), pp. 644–654. ISSN: 0196-6774. DOI: [https://doi.org/10.1016/0196-6774\(90\)90014-6](https://doi.org/10.1016/0196-6774(90)90014-6). URL: <https://www.sciencedirect.com/science/article/pii/0196677490900146>.
- [18] Tamara Kolda and Brett Bader. “Tensor Decompositions and Applications”. In: *SIAM Review* 51 (Aug. 2009), pp. 455–500. DOI: [10.1137/07070111X](https://doi.org/10.1137/07070111X).
- [19] Suresh Bolusani et al. *The SCIP Optimization Suite 9.0*. ZIB-Report 24-02-29. Zuse Institute Berlin, Feb. 2024. URL: <https://nbn-resolving.org/urn:nbn:de:0297-zib-95528>.