

# Модели обнаружения зависимостей во временных рядах (проекции в латентные пространства)

## 1 Цель

Работа посвящена обнаружению причинно-следственных связей между разнородными временными рядами. Примеры зависимых разнородных временных рядов:

1. Эконометрические временные ряды.
2. Связь показателей ЭКГ и пульса (<http://smartlab.ws/component/content/article?id=60>)

Для обнаружения зависимостей между рядами в работе применяется два подхода: тест Гренджера и метод снижения размерности PLS. В обоих случаях для установления зависимости ряда  $\mathbf{x}$  от ряда  $\mathbf{y}$  сравнивается качество прогноза ряда  $\mathbf{x}$  с и без использования ряда  $\mathbf{y}$ . В случае увеличения качества прогноза утверждается, что ряд  $\mathbf{x}$  зависит от ряда  $\mathbf{y}$ .

## 2 Постановка задачи прогнозирования

Поставим задачу прогноза многомерного временного ряда.

Обозначим  $\mathbf{X} = (\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(s)})^\top$  – заданный  $s$ -мерный временной ряд. Построим матрицу плана из сегментов ряда:

$$\begin{pmatrix} x_0^{(1)} & \dots & x_{n-1}^{(1)} \\ \vdots & & \\ x_0^{(s)} & \dots & x_{n-1}^{(s)} \end{pmatrix} = \mathbf{X}_{0:(n-1)}. \quad (1)$$

Пусть  $\mathbf{x}_n = (x_n^{(1)}, \dots, x_n^{(s)})^\top$  – значение ряда  $\mathbf{X}$  в момент времени  $n$ . Построим прогноз  $\hat{\mathbf{x}}$  ряда  $\mathbf{X}$  в точке  $\mathbf{x}_n$ . Прделаем это  $k$  раз для различных обучающих выборок  $\mathbf{X}_{\text{train}}^i = \mathbf{X}_{i:(n+i-1)}$ ,  $i = 0, \dots, (k-1)$ . Получим  $k$  прогнозов  $\hat{\mathbf{X}} = (\hat{\mathbf{x}}_n, \hat{\mathbf{x}}_{n+1}, \dots, \hat{\mathbf{x}}_{n+k-1})$  ряда  $\mathbf{X}$  в точках  $\mathbf{x}_n, \mathbf{x}_{n+1}, \dots, \mathbf{x}_{n+k-1}$ .

Прогностическая модель имеет вид

$$\begin{aligned} \hat{\mathbf{x}}_{t+1} &= \mathbf{f}(\hat{\mathbf{w}}, \mathbf{x}_t, \mathbf{x}_{t-1}, \dots, \mathbf{x}_{t-L+2}), \\ \hat{\mathbf{w}} &= \arg \min_{\mathbf{w}} S(\mathbf{w}, \mathbf{X}, \hat{\mathbf{x}}_n, \hat{\mathbf{x}}_{n+1}, \dots, \hat{\mathbf{x}}_{n+k-1}) = S(\mathbf{w}, \mathbf{X}, \hat{\mathbf{X}}), \end{aligned}$$

где функция потерь

$$S(\mathbf{w}, \mathbf{X}, \hat{\mathbf{X}}) = \sum_{i=0}^{k-1} \mathcal{L}(\mathbf{x}_{n+i}^{(1)}, \hat{\mathbf{x}}_{n+i}^{(1)}).$$

В данной работе в качестве прогностической модели  $\mathbf{f}$  используется алгоритм многомерной гусеницы (MSSA-L). Функция  $\mathbf{f}$  имеет вид:

$$\mathbf{f}(\hat{\mathbf{w}}, \mathbf{x}_t, \mathbf{x}_{t-1}, \dots, \mathbf{x}_{t-L+2}) = \begin{pmatrix} x_{t-L+2}^{(1)} & \dots & x_t^{(1)} \\ x_{t-L+2}^{(2)} & \dots & x_t^{(2)} \\ \vdots & & \\ x_{t-L+2}^{(s)} & \dots & x_t^{(s)} \end{pmatrix} \cdot \mathbf{p}.$$

вектор коэффициентов  $\mathbf{p}$  определяется алгоритмом многомерной гусеницы MSSA-L. Алгоритм MSSA-L подробнее описан в следующем разделе.

### 3 Алгоритм многомерной гусеницы (MSSA-L)

Алгоритм MSSA-L является обобщением на многомерный случай алгоритма гусеницы (SSA). Задача алгоритма MSSA-L состоит в представлении временного ряда в виде суммы интерпретируемых компонент. Это осуществляется в четыре шага: запись ряда в виде траекторной матрицы, сингулярное разложение этой матрицы, группировка компонент, полученных при сингулярном разложении, в интерпретируемые компоненты и восстановление временного ряда по каждой из интерпретируемых компонент.

По ряду (1) построим матрицу Ганкеля  $\mathbf{H} \in \mathbb{R}^{L \times sK}$ ,  $K = N - L + 1$ :

$$\mathbf{H} = [\mathbf{H}_1, \mathbf{H}_2, \dots, \mathbf{H}_s],$$

где  $L$  – ширина окна,  $\mathbf{H}_i \in \mathbb{R}^{L \times K}$  – матрица Ганкеля для ряда  $\mathbf{x}^{(i)}$ ,

$$\mathbf{H}^{(i)} = \begin{pmatrix} x_0^{(i)} & x_1^{(i)} & \dots & x_{N-L}^{(i)} \\ x_1^{(i)} & x_2^{(i)} & \dots & x_{N-L+1}^{(i)} \\ & & \vdots & \\ x_{L-1}^{(i)} & x_L^{(i)} & \dots & x_{N-1}^{(i)} \end{pmatrix}.$$

По матрице Ганкеля  $\mathbf{H}$  восстановим временной ряд  $\mathbf{X}$ . Метод многомерной гусеницы строит приближение  $\hat{\mathbf{H}}$  матрицы  $\mathbf{H}$  меньшего ранга с помощью сингулярного разложения этой матрицы и восстанавливает ряд по матрице  $\hat{\mathbf{H}}$ . Сингулярное разложение матрицы  $\mathbf{H}$  имеет вид

$$\mathbf{H} = \mathbf{U} \mathbf{\Lambda} \mathbf{V} = \sum_{i=1}^d \lambda_i \mathbf{u}_i \mathbf{v}_i^T$$

где  $\lambda_1, \dots, \lambda_d > 0$  – сингулярные числа матрицы  $\mathbf{H}$ ,  $\mathbf{u}_i$  и  $\mathbf{v}_i$  – столбцы матриц  $\mathbf{U}$  и  $\mathbf{V}$ . Тогда наилучшее приближение матрицы  $\mathbf{H}$  матрицей ранга  $r < d$  имеет вид :

$$\hat{\mathbf{H}} = \sum_{i=1}^r \lambda_i \mathbf{u}_i \mathbf{v}_i^T$$

По матрице  $\hat{\mathbf{H}}$  восстанавливается временной ряд  $\mathbf{X}$  путем усреднения элементов, стоящих на антидиагонали.

Алгоритм многомерной гусеницы также позволяет построить прогноз временного ряда в момент  $N$  по  $(L - 1)$  предыдущим значениям ряда. Алгоритм находит такой вектор коэффициентов  $\mathbf{p} \in \mathbb{R}^{(L-1)}$ , что значения ряда  $\mathbf{X}$  в момент  $N$ :

$$\mathbf{x}_N = \begin{pmatrix} x_{N-L+1}^{(1)} & \dots & x_{N-1}^{(1)} \\ x_{N-L+1}^{(2)} & \dots & x_{N-1}^{(2)} \\ & \vdots & \\ x_{N-L+1}^{(s)} & \dots & x_{N-1}^{(s)} \end{pmatrix} \cdot \mathbf{p} = \mathbf{Y} \cdot \mathbf{p} \quad (2)$$

Заметим, что коэффициенты  $\mathbf{p}$  оказываются общими для всех компонент ряда  $\mathbf{X}$ .

Для каждого  $i \in [1, r]$  обозначим  $\tilde{\mathbf{u}}_i$  первые  $(L - 1)$  компонент столбца  $\mathbf{u}_i$ ,  $\pi_i$  – последнюю компоненту столбца  $\mathbf{u}_i$  и  $\nu = \sum_{i=1}^r \pi_i^2$ . Тогда вектор коэффициентов  $\mathbf{p}$  вычисляется по формуле:

$$\mathbf{p} = \frac{1}{1 - \nu^2} \sum_{i=1}^r \pi_i \tilde{\mathbf{u}}_i \quad (3)$$

Заметим, что для одномерного временного ряда справедливы все приведенные соотношения при  $s = 1$ .

## 4 Тест Гренджера

В работе для установления причинно-следственных связей предлагается использовать статистический тест Гренджера. Ниже приведен алгоритм теста Гренджера для проверки наличия зависимости одного временного ряда от другого. Пусть требуется проверить, зависит ли ряд  $\mathbf{x}$  от ряда  $\mathbf{y}$ . Выдвинем гипотезу о независимости ряда  $\mathbf{x}$  от ряда  $\mathbf{y}$  и проверим ее. Делает это следующим образом.

1. Строим прогноз ряда  $\mathbf{x}$  без использования ряда  $\mathbf{y}$  и находим значение функции потерь

$$S_{\mathbf{x}} = \sum_{i=1}^n \mathcal{L}(\mathbf{x}_i, \hat{\mathbf{x}}_i),$$

где  $n$  – длина тестовой выборки.

Функцию  $\mathcal{L}(\mathbf{x}, \hat{\mathbf{x}})$  выбираем в зависимости от распределения ошибок прогноза на тестовой выборке (??).

2. Строим прогноз ряда  $\mathbf{x}$  с использованием ряда  $\mathbf{y}$ . Вычисляем для него значение функции потерь

$$S_{\mathbf{xy}} = \sum_{i=1}^n \mathcal{L}(\mathbf{x}_i, \hat{\mathbf{x}}_i).$$

3. Рассмотрим статистику

$$T(\mathbf{x}, \mathbf{y}) = \frac{N - 2k}{k} \cdot \frac{S_{\mathbf{x}} - S_{\mathbf{xy}}}{S_{\mathbf{xy}}},$$

где  $N$  – длина обучающей выборки,  $k$  – размерность регрессионной модели. Статистика  $T$  имеет распределение  $F(k, N - 2k)$  (распределение Фишера с параметрами  $(k, N - 2k)$ ).

4. Если ряд  $\mathbf{x}$  не зависит от ряда  $\mathbf{y}$ , то значения  $S_{\mathbf{x}}$  и  $S_{\mathbf{xy}}$  будут близки, а статистика  $T(\mathbf{x}, \mathbf{y})$  – незначима. Поэтому в случае больших значений статистики  $T(\mathbf{x}, \mathbf{y})$  отвергаем гипотезу о независимости ряда  $\mathbf{x}$  от  $\mathbf{y}$ . Выберем некоторое критическое значение  $t$  статистики  $T(\mathbf{x}, \mathbf{y})$ . Тогда критерий зависимости ряда  $\mathbf{x}$  от ряда  $\mathbf{y}$  выглядит следующим образом:

Из  $T(\mathbf{x}, \mathbf{y}) > t$  следует, что ряд  $\mathbf{x}$  зависит от ряда  $\mathbf{y}$

5. Аналогично проверим зависимость ряда  $\mathbf{x}$  от восстановленного (с помощью алгоритма MSSA-L) ряда  $\hat{\mathbf{y}}$ . Для этого используем статистику

$$T(\mathbf{x}, \hat{\mathbf{y}}) = \frac{N - 2k}{k} \cdot \frac{S_{\mathbf{x}} - S_{\mathbf{x}\hat{\mathbf{y}}}}{S_{\mathbf{x}\hat{\mathbf{y}}}}.$$

Для более подробного изучения связи между временными рядами  $\mathbf{x}$  и  $\mathbf{y}$  вычисляем кросс-корреляционную функцию  $\gamma_{\mathbf{xy}}(h)$

$$\gamma_{\mathbf{xy}}(h) = \frac{\mathbb{E}[(\mathbf{x}_t - \mu_{\mathbf{x}})(\mathbf{y}_{t+h} - \mu_{\mathbf{y}})]}{\sigma_{\mathbf{x}}\sigma_{\mathbf{y}}},$$

где  $\mathbb{E}$  – математическое ожидание,  $\mu$  – выборочное среднее,  $\sigma$  – выборочная дисперсия.

Если  $h^*$  соответствует максимальному значению кросс-корреляции, то говорят, что ряд  $\mathbf{y}$  сдвинут на  $h^*$  относительно  $\mathbf{x}$ . Заметим, что если ряд  $\mathbf{x}$  сдвинут на  $h_1$  относительно ряда  $\mathbf{y}$ , а ряд  $\mathbf{y}$  сдвинут на  $h_2$  относительно ряда  $\mathbf{z}$ . То ряд  $\mathbf{x}$  сдвинут на  $h_3 = h_1 + h_2$  относительно ряда  $\mathbf{z}$ .

Пусть прогноз ряда  $\mathbf{x}$  строится с использованием истории ряда  $\mathbf{y}$  и пусть с помощью вычисления кросс-корреляции рядов  $\mathbf{x}$  и  $\mathbf{y}$  получено, что ряд  $\mathbf{x}$  отстает от ряда  $\mathbf{y}$  на  $h$  отсчетов времени. Тогда использование при прогнозе ряда  $\mathbf{y}$ , сдвинутого на  $h$  отсчетов назад, может повысить качество прогноза.

## 5 PLS

В этом подходе предлагается строить прогноз по некоторой истории ряда сразу в несколько последующих моментов времени.

Пусть  $\mathbf{X} \in \mathbb{R}^{m \times n}$  – история временного ряда,  $\mathbf{Y} \in \mathbb{R}^{m \times r}$  – значения ряда в последующие моменты времени. Предполагается, что между строками матриц  $\mathbf{X}$  и  $\mathbf{Y}$  существует линейная зависимость:

$$\mathbf{y} = \mathbf{x} \cdot \boldsymbol{\Theta} + \boldsymbol{\varepsilon}, \quad \mathbf{x} \in \mathbb{R}^n, \mathbf{y} \in \mathbb{R}^r,$$

где  $\boldsymbol{\Theta}$  – матрица параметров модели,  $\boldsymbol{\varepsilon}$  – вектор ошибок прогноза.

Ошибка прогноза вычисляется следующим образом:

$$S(\boldsymbol{\Theta}, \mathbf{X}, \mathbf{Y}) = \|\mathbf{Y} - \mathbf{X} \cdot \boldsymbol{\Theta}\|_2^2 = \sum_{i=1}^m \|\mathbf{y}_i - \mathbf{x}_i \cdot \boldsymbol{\Theta}\|_2^2$$

Для нахождения параметров модели  $\boldsymbol{\Theta}$  предлагается использовать метод частных наименьших квадратов PLS. Алгоритм PLS находит в латентном пространстве матрицу  $\mathbf{T} \in \mathbb{R}^{m \times l}$ , наилучшим образом описывающую матрицы  $\mathbf{X}$  и  $\mathbf{Y}$ . Матрицы  $\mathbf{X}$  и  $\mathbf{Y}$  проецируются в латентное пространство следующим образом:

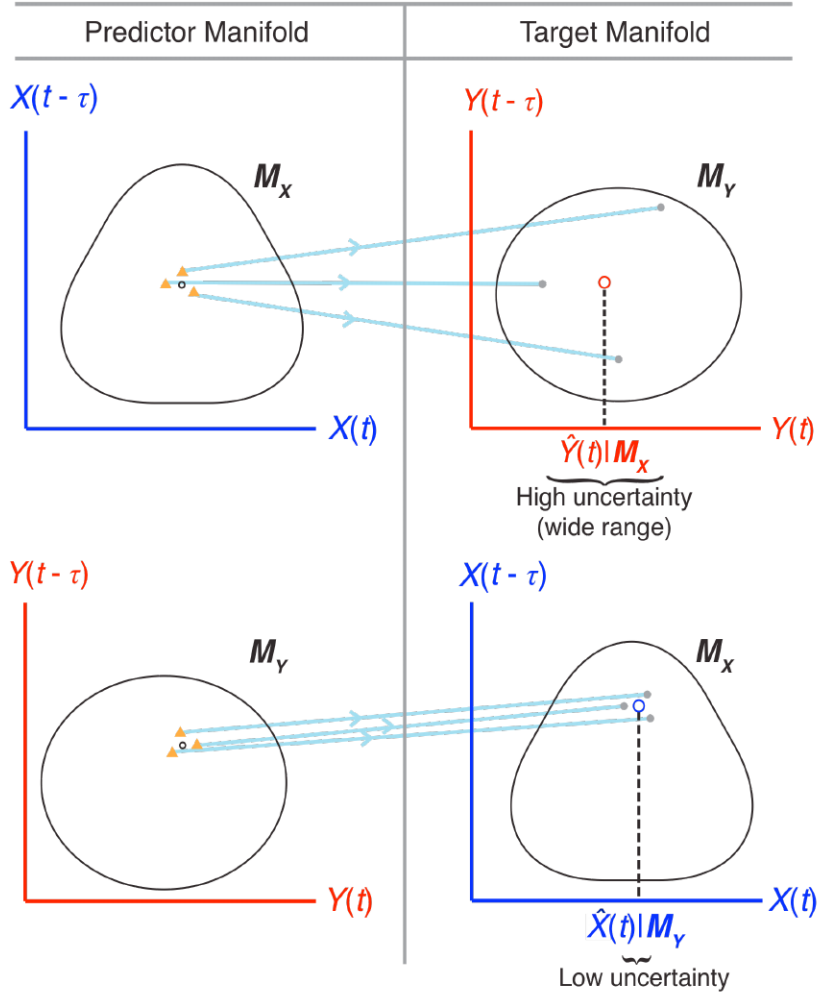
$$\begin{aligned} \mathbf{X} &= \mathbf{T} \cdot \mathbf{P}^T + \mathbf{F} = \sum_{k=1}^l \mathbf{t}_k \cdot \mathbf{p}_k^T + \mathbf{F} \\ \mathbf{Y} &= \mathbf{T} \cdot \mathbf{Q}^T + \mathbf{E} = \sum_{k=1}^l \mathbf{t}_k \cdot \mathbf{q}_k^T + \mathbf{E} \end{aligned}$$

где  $\mathbf{T}$  — матрица совместного описания объектов и ответов в латентном пространстве, причём столбцы матрицы  $\mathbf{T}$  ортогональны;  $\mathbf{P}$ ,  $\mathbf{Q}$  — матрицы перехода из латентного пространства в исходные пространства;  $\mathbf{E}$ ,  $\mathbf{F}$  — матрицы невязок.

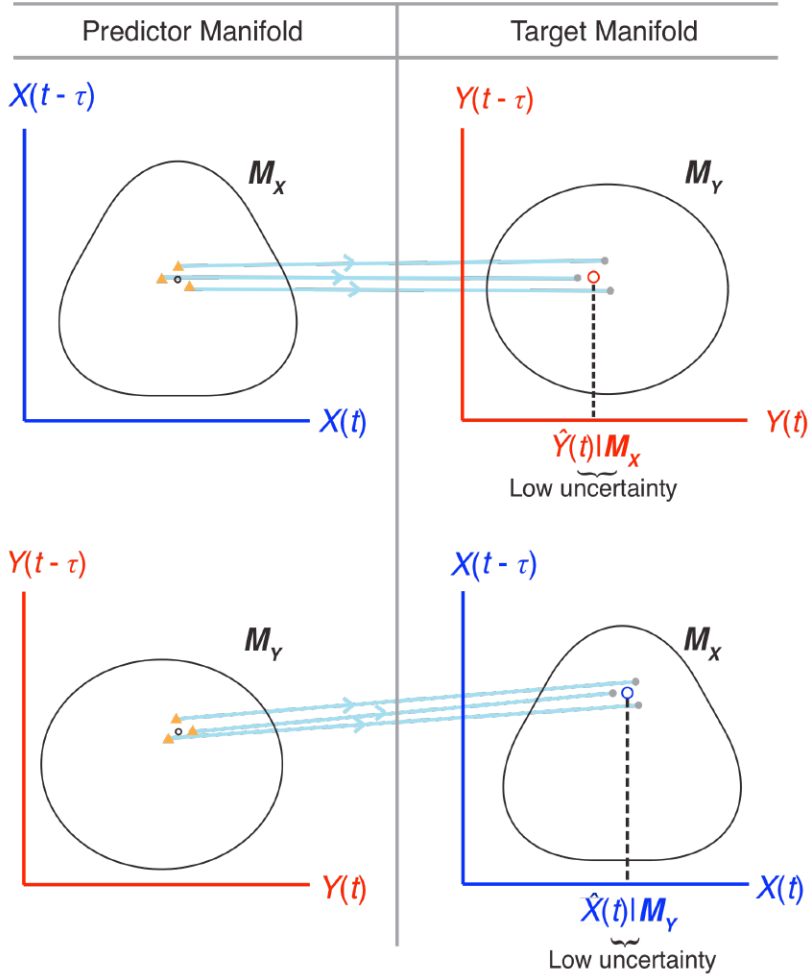
Алгоритм PLS находит матрицы  $\mathbf{T}$ ,  $\mathbf{P}$ ,  $\mathbf{Q}$ , а также такую матрицу  $\mathbf{W}$ , что параметры модели можно вычислить по формуле

$$\boldsymbol{\Theta} = \mathbf{W}(\mathbf{P}^T \mathbf{W})^{-1} \mathbf{Q}^T$$

**B** Asymmetric Causality,  $X \Rightarrow Y$



## A Bidirectional Causality (generic case), $X \Leftrightarrow Y$



Пусть  $\mathbf{X}$  и  $\mathbf{Y}$  –  $s$ -мерные временные ряды длины  $N$ .

Прогноз ряда  $\mathbf{Y}$  с помощью ряда  $\mathbf{X}$  строится следующим образом.

Строим матрицу Ганкеля

$$\mathbf{H} = \begin{pmatrix} \mathbf{X}_1 & \mathbf{X}_2 & \dots & \mathbf{X}_{L-1} & \mathbf{X}_L \\ \mathbf{X}_2 & \mathbf{X}_3 & \dots & \mathbf{X}_L & \mathbf{X}_{L+1} \\ \vdots & \vdots & & \vdots & \vdots \\ \mathbf{X}_{N-L+1} & \mathbf{X}_{N-L+2} & \dots & \mathbf{X}_{N-1} & \mathbf{X}_N \end{pmatrix},$$

где  $L$  – ширина окна, длина истории ряда, используемая при нахождении главных компонент.

Находим  $(L + 1)$  ближайших соседей вектора  $\mathbf{x}_t$ . Обозначим их индексы через  $t_1, \dots, t_{L+1}$ . Тогда ближайшие соседи  $\mathbf{x}_t$  это строки матрицы  $\mathbf{H}$  с номерами  $t_1, \dots, t_{L+1}$ .

$$\mathbf{x}_{t_i} = (\mathbf{x}_{t_i}, \mathbf{x}_{t_i-1}, \dots, \mathbf{x}_{t_i-(L-1)}), \quad i = 1, \dots, (L + 1)$$

Прогноз  $\hat{\mathbf{Y}}_t$  строится следующим образом

$$\hat{\mathbf{Y}}_t = \sum_{i=1}^{L+1} w_i Y_{t_i}, \quad \text{где } t_i \text{ – индексы ближайших соседей } \mathbf{x}_t$$

$$w_i = \frac{u_i}{\sum_i u_i}, \quad u_i = \exp - \left( \frac{\|\mathbf{x}_t - \mathbf{x}_{t_i}\|_2}{\|\mathbf{x}_t - \mathbf{x}_{t_{L+1}}\|_2} \right)$$

Аналогично строится прогноз ряда  $\mathbf{X}$  с использованием ряда  $\mathbf{Y}$ .

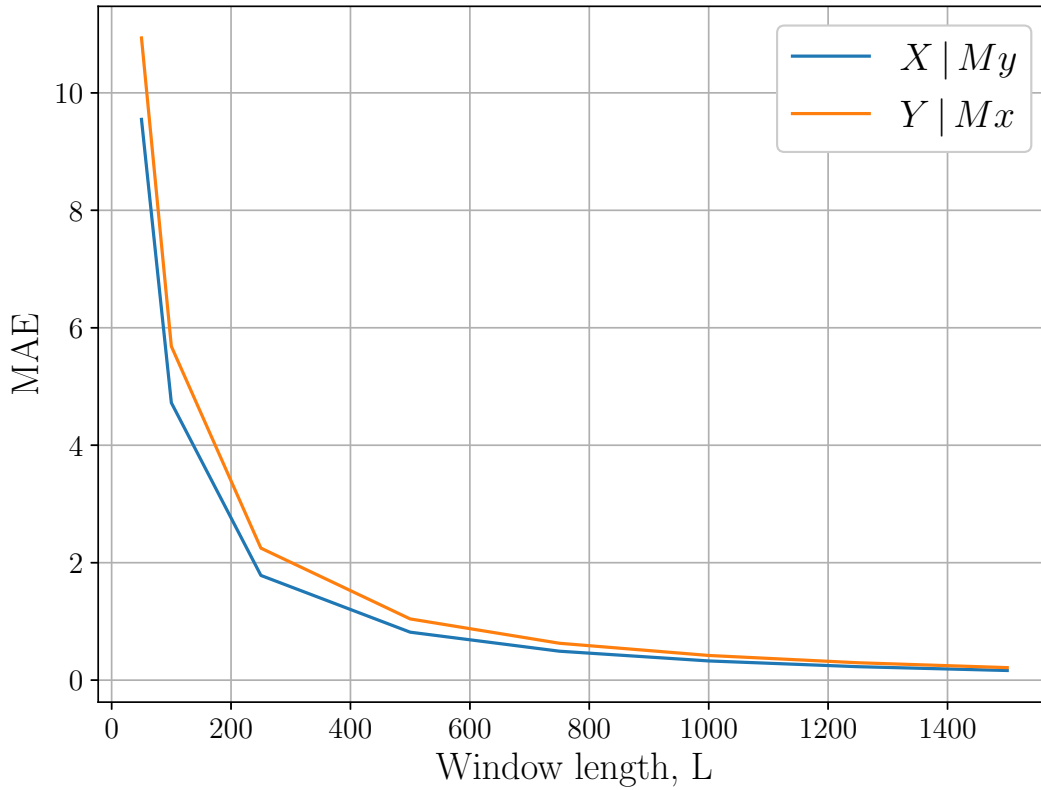
Эксперимент проводился на временных рядах  $\mathbf{X}$  и  $\mathbf{Y}$  длины  $L$ , следующего вида

$$\mathbf{X}_{t+1} = \mathbf{X}_t \cdot [r_x(1 - \mathbf{X}_t) - \beta_{x,y}\mathbf{Y}_t]$$

$$\mathbf{Y}_{t+1} = \mathbf{Y}_t [r_y(1 - \mathbf{Y}_t) - \beta_{y,x}\mathbf{X}_t]$$

При каждом построении прогноза коэффициенты  $r_x, r_y, \beta_{x,y}, \beta_{y,x}$  выбирались случайно. Коэффициенты  $\beta_{x,y}, \beta_{y,x}$  – из равномерного распределения на отрезке  $[3.6, 4.0]$ , коэффициенты  $r_x, r_y$  – из равномерного распределения на отрезке  $[0, 0.5]$ . При каждой длине рядов  $L$  ошибка прогноза усреднялась по всем наборам значений коэффициентов  $r_x, r_y, \beta_{x,y}, \beta_{y,x}$ .

На рисунке показана зависимость средней абсолютной ошибки прогноза в зависимости от длины рядов  $L$ . Видно, что ряд  $\mathbf{X}$  лучше восстанавливается по ряду  $\mathbf{Y}$ , чем ряд  $\mathbf{Y}$  по ряду  $\mathbf{X}$ .

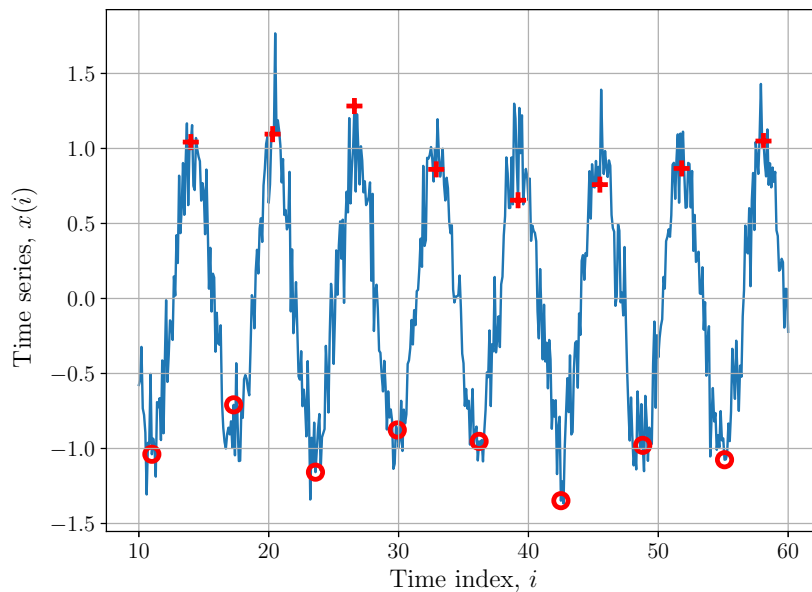


## 7 Фазовые траектории

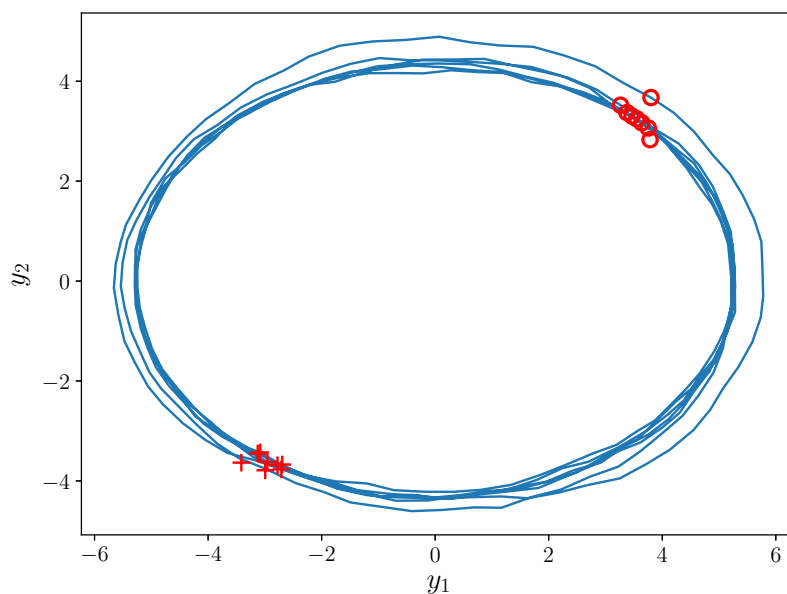
### 7.1 Сгенерированные данные

Эксперимент проводился на сгенерированном ряде  $\mathbf{X}$

$$\mathbf{X}_t = \sin t + \sigma^2 \boldsymbol{\varepsilon}_i, \quad \boldsymbol{\varepsilon}_i \in \mathcal{N}(0, 1)$$



**Figure 1:** Ряд  $\mathbf{X}$ . Красным отмечены точки, расположенные на расстоянии в пол периода



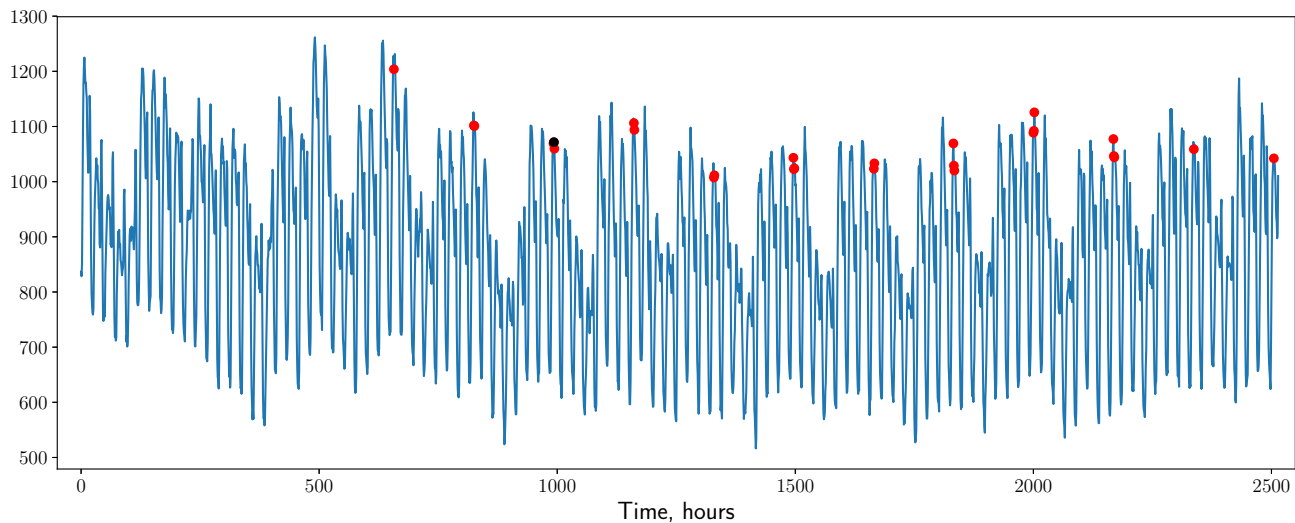
**Figure 2:** Проекция фазовых траекторий на пространство, натянутое на первые две главные компоненты ряда  $\mathbf{X}$



## 7.2 Реальные данные

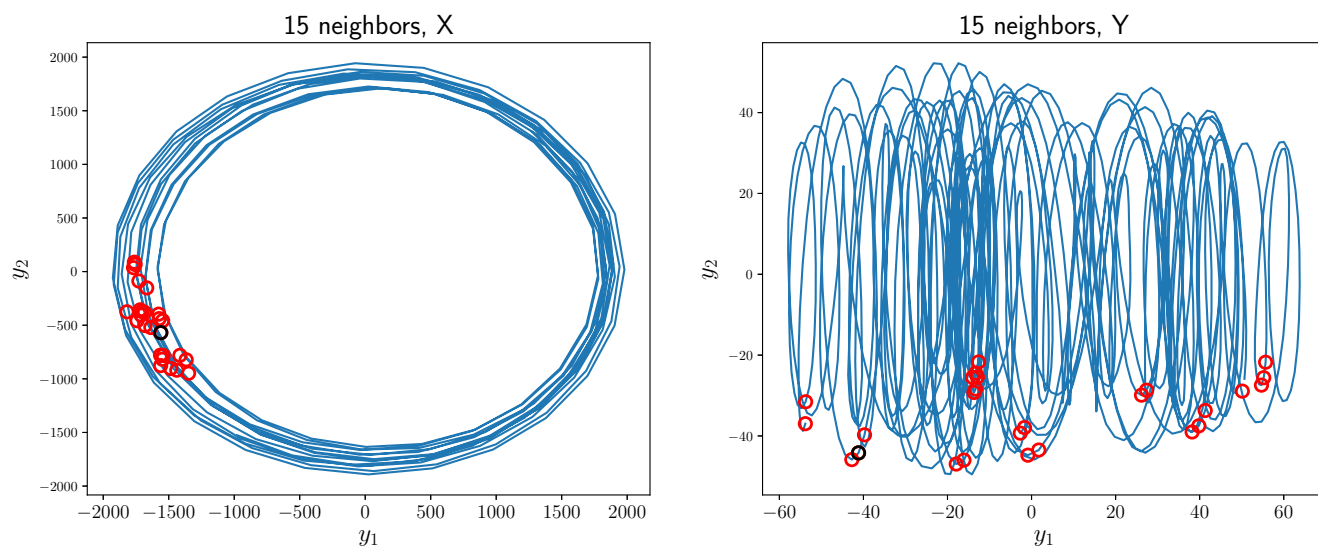
Эксперимент проводился на ряде объемов потребления электроэнергии  $\mathbf{X}$  и ряде значений температуры  $\mathbf{Y}$ .

Для  $t^* = 1000$  найдем  $k$  ближайших соседей вектора  $\mathbf{X}_{t^*}$ . Обозначим, соответствующие им индексы  $t_1, \dots, t_k$ . На рисунке изображен ряд потребления электроэнергии  $\mathbf{X}$ , точки  $\mathbf{X}_{t_1}, \dots, \mathbf{X}_{t_k}$  отмеченными красным, точка  $\mathbf{X}_{t^*}$  – черным.



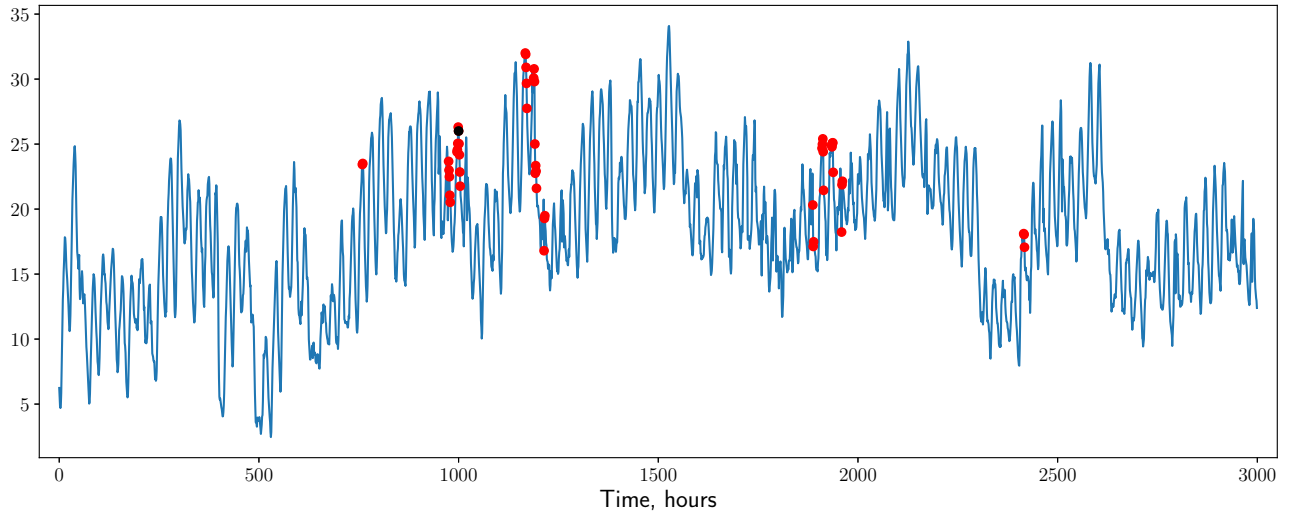
**Figure 3:** Ближайшие соседи точки  $\mathbf{X}_{t^*}$

Изобразим ближайших соседей  $\mathbf{X}_{t_1}, \dots, \mathbf{X}_{t_k}$  на фазовых траекториях ряда  $\mathbf{X}$  и соответствующие им точки  $\mathbf{Y}_{t_1}, \dots, \mathbf{Y}_{t_k}$  на фазовых траекториях ряда  $\mathbf{Y}$ .



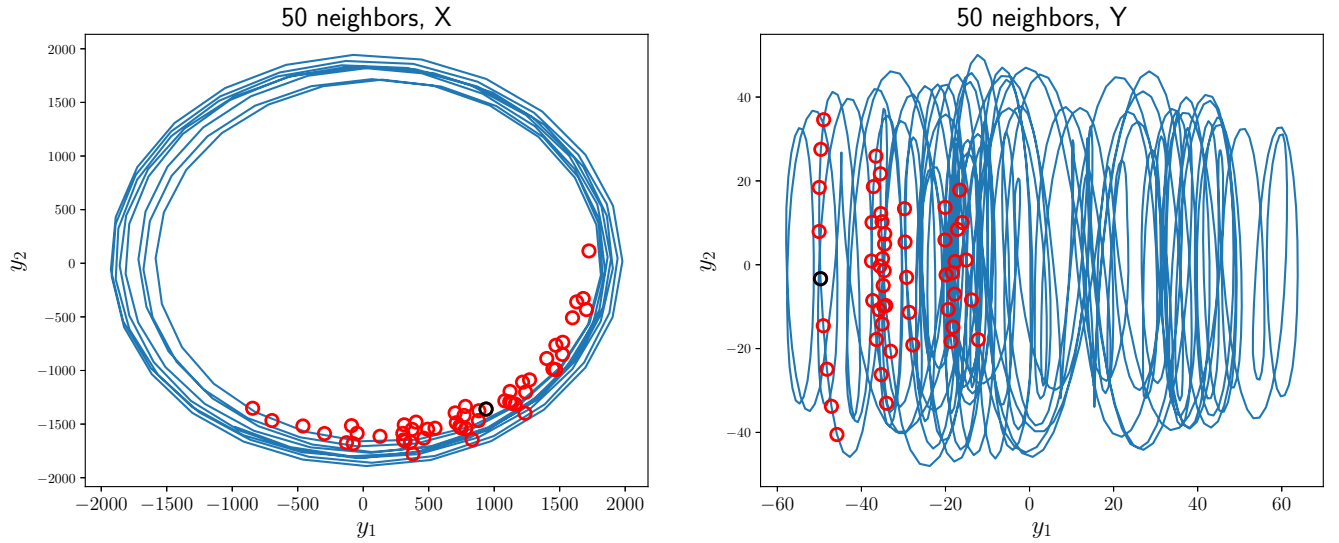
**Figure 4:** Точки на фазовых траекториях рядов  $\mathbf{X}$  и  $\mathbf{Y}$ , соответствующие ближайшим соседям  $\mathbf{X}_{t^*}$

Аналогично для  $t^*$  находим ближайших соседей вектора  $\mathbf{Y}_{t^*}$ . Обозначим их  $\mathbf{Y}_{t_1}, \dots, \mathbf{Y}_{t_k}$ .



**Figure 5:** Ближайшие соседи точки  $\mathbf{Y}_{t^*}$

Изобразим ближайших соседей  $\mathbf{Y}_{t_1}, \dots, \mathbf{Y}_{t_k}$  на фазовых траекториях ряда  $\mathbf{Y}$  и соответствующие им точки  $\mathbf{X}_{t_1}, \dots, \mathbf{X}_{t_k}$  на фазовых траекториях ряда  $\mathbf{X}$ .



**Figure 6:** Точки на фазовых траекториях рядов  $\mathbf{X}$  и  $\mathbf{Y}$ , соответствующие ближайшим соседям  $\mathbf{Y}_{t^*}$