

Содержание

1	Введение	4
2	Постановка задачи	4
2.1	Матрицы ковариации набора временных рядов	4
2.2	Обнаружение взаимосвязи между рядами	5
2.3	Анализ матриц ковариации набора временных рядов	7
2.4	Метод ковариационного, перекрестного отображения	9
2.5	Выбор прогностической модели	10
2.5.1	Модель улучшения базового прогноза	10
3	Вычислительный эксперимент	11
3.1	Цели и постановка	11
3.2	Результаты и обсуждение	12
4	Заключение	19
	Список литературы	21

Аннотация

Матрицы ковариации многомерных временных рядов лежат в римановом пространстве SPD матриц. Свойства этого пространства используются для решения задач классификации многомерных временных рядов. Проблема с использованием этого же подхода для прогнозирования многомерных временных рядов заключается в том, что в отличие от задач классификации, нет возможности решать задачу в пространстве SPD матриц. Предлагаются прогностические модели, использующие риманову геометрию SPD матриц, и исследуется качество их прогноза в зависимости от свойств прогнозируемых временных рядов.

Ключевые слова: набор временных рядов, прогноз, матрицы ковариации, пространство SPD матриц.

1 Введение

Матрицы ковариаций многомерных временных рядов принадлежат гладкому многообразию SPD матриц. Свойства этого пространства [1] могут быть применены для решения различных задач, связанных с многомерными временными рядами. В частности, были предложены методы классификации EEG сигналов, основанные на римановой геометрии пространства SPD матриц [2], [3], [4]. Главная идея состоит в том, чтобы перейти в пространство SPD матриц и использовать риманову метрику для решения задачи. Также был разработан метод, в которых матрицы ковариации отображались в касательное пространство, являющееся евклидовым, и в нем классифицировались с помощью известных методов классификации, например SVM. [5].

В отличие от классификации, для прогнозирования многомерных временных рядов не разработаны прогностические модели, использующие риманову геометрию матриц ковариации. В первую очередь, это связано с тем, что нельзя решить задачу в пространстве матриц ковариации, т.к. модель должна прогнозировать исходные временные ряды, а не метку, как в задаче классификации. Однако, можно рассматривать векторное представление матриц ковариации как описание взаимосвязи фазовых траекторий многомерных временных рядов и использовать эту информацию для улучшения качества прогноза.

2 Постановка задачи

2.1 Матрицы ковариации набора временных рядов

Набор временных рядов в каждый момент времени задается n -мерным вектором: $\{\mathbf{x}_t\}_{t=1}^T = \{[x_{1t}, \dots, x_{nt}]^T\}_{t=1}^T$. Каждому ряду набора можно поставить в соответствие фазовое пространство векторов задержек $\overleftarrow{\mathbf{x}}_{it} = [x_{i(t-L+1)}, \dots, x_{it}]^T$ размерности L . Для точек фазовых траекторий, отвечающих одному и тому же моменту времени поставим в соответствие матрицу ковариации:

$$\mathbf{C}_t = \frac{1}{L-1} \mathbf{X}_t \mathbf{X}_t^T, \quad (1)$$

где $\mathbf{X}_t = [\overleftarrow{\mathbf{x}}_{1t} \dots \overleftarrow{\mathbf{x}}_{nt}]^T$. Таким образом, \mathbf{C}_t описывает взаимосвязь между точками фазовых пространств различных временных рядов набора.

Матрицы ковариации, при условии, что они не вырождены, являются симмет-

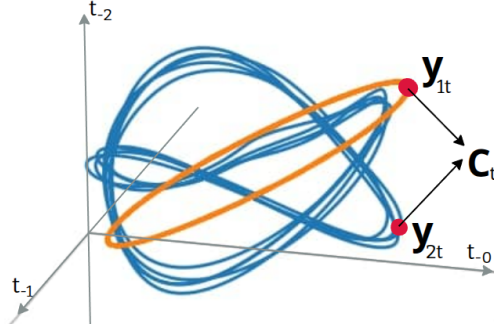


Рис. 1: Точкам фазовых траекторий $\overleftarrow{\mathbf{x}}_{1t}, \overleftarrow{\mathbf{x}}_{2t}$ ставится в соответствие матрица ковариации \mathbf{C}_t .

ричными положительно определенными (SPD). Пространство SPD – риманово многообразие M , на котором вводится метрика, геометрическое среднее его элементов и определяется касательное пространство в каждой точке многообразия [1]. Касательное пространство в каждой точке – пространство симметричных матриц, являющееся векторным. Следовательно, с элементами касательного пространства можно работать как с векторами. Перевод SPD матриц в касательное пространство осуществляется при помощи логарифмического отображения [1]:

$$\mathbf{S}_t = \text{Log}_{\mathbf{C}}(\mathbf{C}_t) = \mathbf{C}^{\frac{1}{2}} \log(\mathbf{C}^{-\frac{1}{2}} \mathbf{C}_t \mathbf{C}^{-\frac{1}{2}}) \mathbf{C}^{\frac{1}{2}}, \quad (2)$$

где \mathbf{C}_t – матрица ковариации, \mathbf{S}_t – ее образ в касательном к \mathbf{C} пространстве. Обратное преобразование осуществляется с помощью экспоненциального отображения [1]:

$$\mathbf{C}_t = \text{Exp}_{\mathbf{C}}(\mathbf{S}_t) = \mathbf{C}^{\frac{1}{2}} \exp(\mathbf{C}^{-\frac{1}{2}} \mathbf{S}_t \mathbf{C}^{-\frac{1}{2}}) \mathbf{C}^{\frac{1}{2}}. \quad (3)$$

В качестве точки построения касательного пространства выбирается геометрическое среднее матриц ковариации $\{\mathbf{C}_t\}_{t=L}^T$ [5] из тех соображений, что образ матрицы в касательном пространстве тем ближе к исходной матрице, чем ближе точка построения касательного пространства. На Рис.2 представлена иллюстрация перехода в касательное пространство и обратно.

2.2 Обнаружение взаимосвязи между рядами

При прогнозировании набора временных рядов одним из основных вопросов является вопрос о том, имеет ли смысл прогнозировать временные ряды совместно или по-отдельности. Классическим методом определения наличия причинных связей меж-

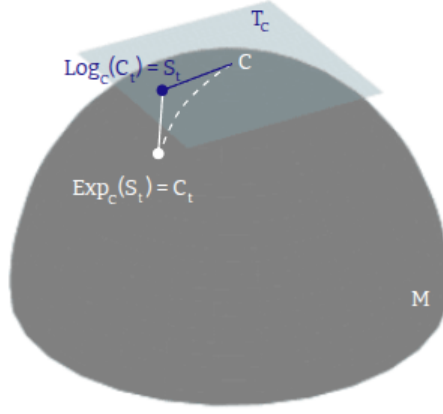


Рис. 2: Перевод C_t в пространство T_C и обратно.

ду временными рядами является тест Грейнджера [6]. Вывод о наличии причинной по Грейнджеру связи между рядами X, Y делается если удастся улучшить прогноз X с помощью VAR модели [7] при добавлении значений Y . В этом случае говорят, что X является причиной по Грейнджеру Y . Одной из проблем этого метода является то, что он предполагает отсутствие информации о Y в X . Другими словами, если информация о втором временном ряде фундаментально заложена в первом, то тест Грейнджера не подойдет для обнаружения причинных связей. Для решения этой проблемы был предложен метод сходящегося перекрестного отображения (CCM) [8].

Метод сходящегося перекрестного отображения основан на том, что временные ряды X, Y причинно связаны если они порождены одной и той же динамической системой. По теореме Такенса [9], фазовое пространство векторов задержек, при достаточно большом L и при выполнении условий регулярности функции преобразования обобщенных координат в значения временного ряда, является вложением многообразия обобщенных координат динамической системы в линейное пространство \mathbb{R}^L . Таким образом, фазовое пространство временного ряда является ни чем иным, как реконструкцией динамической системы, его порождающей. Неформально, метод проверяет, что по состояниям динамической системы, порождающей X можно восстановить состояния динамической системы Y . Заметим, что эти отношения не обязательно симметричные. Возможно, что в X хранится достаточно информации для восстановления Y , однако обратное неверно.

Для того, чтобы проверить является ли X причиной Y с помощью метода перекрестного отображения, строится фазовое пространство векторов задержек ряда X . Далее, для каждого момента времени t находим $L + 1$ ближайших точек фазового пространства к точке $\overleftarrow{\mathbf{x}}_{Xt} : \overleftarrow{\mathbf{x}}_{Xt_1}, \dots, \overleftarrow{\mathbf{x}}_{Xt_{L+1}}$. Получив времена ближайших соседей, строим

прогноз:

$$\hat{Y}_{t+1} = \sum_{i=1}^{L+1} d_i Y_{t_i}, \quad (4)$$

$$\text{где } d_i = \frac{\exp(-\rho(\bar{\mathbf{x}}_{Xt}, \bar{\mathbf{x}}_{Xt_i})/\rho(\bar{\mathbf{x}}_{Xt}, \bar{\mathbf{x}}_{Xt_1}))}{\sum_{j=1}^{L+1} \exp(-\rho(\bar{\mathbf{x}}_{Xt}, \bar{\mathbf{x}}_{Xt_j})/\rho(\bar{\mathbf{x}}_{Xt}, \bar{\mathbf{x}}_{Xt_1}))}, \quad \rho - \text{евклидова метрика. Таким образом, вес значения тем меньше, чем меньше расстояние между точками в фазовом пространстве ряда } X.$$

Вычислив прогноз для каждого момента времени, посчитаем корреляцию прогноза \hat{Y} и исходного ряда Y . При наличии причинных связей, присутствует корреляция значений. Однако, помимо этого, требуется, что при увеличении истории временных рядов X , Y значение корреляции увеличиваясь, сходилось к некоторому значению. Это требование сходимости отвечает тому, что при наличии информации о Y в X ее должно становиться больше при увеличении истории или, другими словами, фазовые траектории временных рядов становятся все плотнее и ближе к аттракторам.

2.3 Анализ матриц ковариации набора временных рядов

Зададимся вопросом, когда информация о матрицах ковариации, соответствующих точкам фазовых пространств временных рядов набора, может помочь улучшить качество прогноза. Рассмотрим, для простоты, случай набора из двух временных рядов $\mathbf{x}_t = (x_{1t}, x_{2t})$. Заметим, что самом простом случае, если временные ряды незашумлены, скоррелированы и матрица ковариации \mathbf{C}_t константна, то не получится улучшить прогноз с ее помощью. Можно сказать, что матрица ковариации, в этом случае, описывает нулевую дисперсию в данных и поэтому улучшение прогноза невозможно. С другой стороны, если шум в данных будет слишком большим, то и матрицы ковариации будут очень зашумлены, а следовательно они будут плохо описывать реальные ковариации временных рядов. Проведем анализ изменения матриц ковариации при добавлении шума в исходные временные ряды:

Теорема 1. Пусть $\mathbf{x}_t = (x_{1t}, x_{2t})$ и $x_{it} = \tilde{x}_{it} + \varepsilon_{it}$, где $i \in \{1, 2\}$, $\varepsilon_{it} \sim \mathcal{N}(0, \sigma_i^2)$. Пусть $\tilde{\mathbf{C}}_t$ – матрица ковариации, вычисленная по формуле (1) для набора временных рядов $\tilde{\mathbf{x}}_t = (\tilde{x}_{1t}, \tilde{x}_{2t})$. Тогда матрица ковариации \mathbf{C}_t для набора \mathbf{x}_t равна:

$$\mathbf{C}_t = \tilde{\mathbf{C}}_t + \delta \mathbf{C}_t, \quad (5)$$

где матрица среднеквадратичных отклонений элементов \mathbf{C}_t :

$$\Delta \mathbf{C}_t = \frac{1}{L-1} \begin{pmatrix} \sqrt{4\|\tilde{\mathbf{x}}_{1t}\|^2\sigma_1^2 + 2L\sigma_1^4} & \sqrt{\sigma_2^2\|\tilde{\mathbf{x}}_{1t}\|^2 + \sigma_1^2\|\tilde{\mathbf{x}}_{2t}\|^2 + L\sigma_1^2\sigma_2^2} \\ \sqrt{\sigma_2^2\|\tilde{\mathbf{x}}_{1t}\|^2 + \sigma_1^2\|\tilde{\mathbf{x}}_{2t}\|^2 + L\sigma_1^2\sigma_2^2} & \sqrt{4\|\tilde{\mathbf{x}}_{2t}\|^2\sigma_2^2 + 2L\sigma_2^4} \end{pmatrix} \quad (6)$$

Как можем заметить из (6), отклонения тем больше, чем больше дисперсии шумов σ_1^2, σ_2^2 . Теперь исследуем вопрос о том, насколько могут отдалиться от друг друга одинаковые матрицы ковариации $\tilde{\mathbf{C}}_{t_1} = \tilde{\mathbf{C}}_{t_2} = \tilde{\mathbf{C}}$ (будем считать, что они построены на одинаковых векторах задержек) при добавлении шума, в терминах метрики в римановом пространстве SPD матриц. Для этого вспомним определение метрики в пространстве SPD матриц [1]:

$$d(\mathbf{C}_{t_1}, \mathbf{C}_{t_2}) = \|\text{Log}(\mathbf{C}_{t_1}^{-1}\mathbf{C}_{t_2})\|_F = \left[\sum_{i=1}^L \ln^2 \lambda_i \right]^{\frac{1}{2}}, \quad (7)$$

где λ_i – i -е собственное число матрицы $\mathbf{C}_{t_1}^{-1}\mathbf{C}_{t_2}$. В невозмущенном случае, очевидно, собственные числа произведения матрицы $\tilde{\mathbf{C}}_{t_1}^{-1}\tilde{\mathbf{C}}_{t_2} = \mathbf{E}$ равны $\tilde{\lambda}_1 = \tilde{\lambda}_2 = 1$. Посмотрим, как они изменятся при добавлении шума. Считаем, что отклонения $\delta \mathbf{C}_{t_i}$ – небольшие поправки к исходным матрицам. В этом случае, мы можем приближенно посчитать отклонение произведения матриц $\mathbf{C}_{t_1}^{-1}\mathbf{C}_{t_2}$ от единичной матрицы, а дальше и отклонение собственных чисел λ_i от единицы, воспользовавшись первым порядком теории возмущений. Единственное, заметим, что корректность определения римановой метрики обеспечивается тем, что собственные числа матрицы, равной произведению SPD матриц, (в общем случае не симметричной матрицы) вещественны и положительны. Таким образом, нужно потребовать, чтобы матрица $\tilde{\mathbf{C}}^{-1} - \tilde{\mathbf{C}}^{-1}\delta \mathbf{C}_{t_1}\tilde{\mathbf{C}}^{-1}$ была положительно определена. С учетом этого условия рассмотрим приближенное вычисление произведения матриц :

$$\mathbf{C}_{t_1}^{-1}\mathbf{C}_{t_2} \approx (\tilde{\mathbf{C}}^{-1} - \tilde{\mathbf{C}}^{-1}\delta \mathbf{C}_{t_1}\tilde{\mathbf{C}}^{-1})(\tilde{\mathbf{C}} + \delta \mathbf{C}_{t_2}) \approx \mathbf{E} + \tilde{\mathbf{C}}^{-1}(\delta \mathbf{C}_{t_2} - \delta \mathbf{C}_{t_1}) = \mathbf{E} + \mathbf{M}. \quad (8)$$

Лемма 1. Пусть $\mathbf{x}_t = (x_{1t}, x_{2t})$ и $x_{it} = \tilde{x}_{it} + \varepsilon_{it}$, где $i \in \{1, 2\}, \varepsilon_{it} \sim \mathcal{N}(0, \sigma_i^2)$. Пусть $\mathbf{C}_{t_1} = \tilde{\mathbf{C}} + \delta \mathbf{C}_{t_1}, \mathbf{C}_{t_2} = \tilde{\mathbf{C}} + \delta \mathbf{C}_{t_2}$. Тогда, при условии приближения (8), в первом порядке теории возмущений, собственные числа $\mathbf{C}_{t_1}^{-1}\mathbf{C}_{t_2}$ равны:

$$\lambda_1 = 1 + \mathbf{M}_{11}, \quad (9)$$

$$\lambda_2 = 1 + \mathbf{M}_{22}. \quad (10)$$

Теперь, когда знаем приращения к собственным числам матрицы $\mathbf{C}_{t_1}^{-1}\mathbf{C}_{t_2}$, посчитаем приближенно отклонение расстояния в римановом пространстве между матрицами $\mathbf{C}_{t_1}, \mathbf{C}_{t_2}$. Для этого будем считать:

$$\ln \lambda_i = \ln(1 + \mathbf{M}_{ii}) \approx \mathbf{M}_{ii}. \quad (11)$$

Подставим эти значения собственных чисел в (7) и получим оценку того, насколько одинаковые матрицы могут отклониться от друг друга, в терминах метрики в римановом пространстве SPD матриц, при добавлении шума в исходные временные ряды:

$$d(\mathbf{C}_{t_1}, \mathbf{C}_{t_2}) = \sqrt{\mathbf{M}_{11} + \mathbf{M}_{22}}. \quad (12)$$

2.4 Метод ковариационного, перекрестного отображения

Оценки выше – приближенные и были получены в предположениях того, что шум в данных не очень большой. Однако, хотелось бы иметь способ по набору из двух временных рядов, без априорных знаний об их разделении на сигнал и шум, определять насколько использование матриц ковариации в прогнозировании целесообразно. Эту задачу можно переформулировать следующим образом: сколько информации о временных рядах содержится в ряде из матриц ковариаций. Таким образом, для ее решения используется идея метода перекрестного, сходящегося отображения 1. Аналогично случаю набора из двух временных рядов, для данного момента времени t , определяются $L+1$ ближайший сосед матрицы \mathbf{C}_t по римановой метрике (7): $\mathbf{C}_{t_1}, \mathbf{C}_{t_2}, \dots, \mathbf{C}_{t_{L+1}}$. Далее, строится прогноз i -го временного ряда:

$$x_{i(t+1)} = \sum_{k=1}^{L+1} \alpha_k x_{it_k}, \quad (13)$$

где вес α_k пропорционален расстоянию $d(\mathbf{C}_t, \mathbf{C}_{t_k})$. Для преобразования массива корреляций в число $\bar{\rho}$, используется среднее арифметическое элементов массива:

$$\bar{\rho} = \frac{1}{n} \sum_{i=1}^n \rho_i. \quad (14)$$

Algorithm 1 Метод ковариационного, перекрестного отображения

Require: Временной ряд $\mathbf{x}_t = (x_{1t}, \dots, x_{nt})^T$, матрицы ковариации \mathbf{C}_t ; $t = 1, \dots, N$

Ensure: $[\rho_1, \dots, \rho_n]$

for $t \in [L+2, N-1]$ **do**

$\mathbf{C}_{t_1}, \dots, \mathbf{C}_{t_{L+1}} \leftarrow \text{nn}(\mathbf{C}_t)$

for $i \in [1, n]$ **do**

$$[\alpha_1, \dots, \alpha_{L+1}] \leftarrow \frac{\exp(-\rho(\mathbf{C}_t, \mathbf{C}_{t_j})/\rho(\mathbf{C}_t, \mathbf{C}_{t_1}))}{\sum_{l=1}^{L+1} \exp(-\rho(\mathbf{C}_t, \mathbf{C}_{t_l})/\rho(\mathbf{C}_t, \mathbf{C}_{t_1}))}$$

$$\hat{x}_{i(t+1)} \leftarrow \sum_{k=1}^{L+1} d_k x_{it_k}$$

end for

end for

for $i \in [1, n]$ **do**

$$\rho_i \leftarrow \rho([x_{i(L+2)}, \dots, x_{iN}], [\hat{x}_{i(L+2)}, \dots, \hat{x}_{iN}])$$

end for

2.5 Выбор прогностической модели

Требуется выбрать прогностическую модель, которая при построении прогноза будет учитывать матрицы ковариации или их образы в касательном пространстве. Т.к. ставится вопрос об улучшении качества прогноза, при добавлении в модель матриц ковариации, то прогностическая модель должна состоять из двух частей: *базовая* прогностическая модель f и *ковариационная* модель F . Под базовой прогностической моделью понимается модель, с помощью которой производится прогноз набора временных рядов с учетом, или без учета матриц ковариации. Под ковариационной моделью имеется ввиду модель, с помощью которой информация о матрицах ковариации учитывается в прогнозе. Наиболее простым случаем модели является отсутствие базовой модели, т.е. прогноз осуществляется только при помощи матриц ковариации.

2.5.1 Модель улучшения базового прогноза

Рассмотрим более сложный способ построения модели. При таком способе построения прогностической модели, требуется выбрать f, F :

$$\mathbf{x}_t \xrightarrow{f} \hat{\mathbf{x}}_{t+1}^f, \quad (15)$$

$$(\hat{\mathbf{x}}_{t+1}^f, \mathbf{S}_t) \xrightarrow{F} \hat{\mathbf{x}}_{t+1}, \quad (16)$$

где $\hat{\mathbf{x}}_{t+1}^f$ – прогноз базовой модели, \mathbf{S}_t – образ матрицы ковариации \mathbf{C}_t в касательном пространстве, $\hat{\mathbf{x}}_{t+1}$ – прогноз ковариационной модели F по $\hat{\mathbf{x}}_{t+1}^f$ и \mathbf{S}_t . После выбора f, F

решаются следующие оптимизационные задачи:

$$\hat{\mathbf{w}}_f = \arg \min_{\mathbf{w} \in \mathbb{R}^k} \mathcal{L}(\mathbf{w}, \hat{\mathbf{x}}_{t+1} - \mathbf{x}_{t+1}), \quad (17)$$

$$\hat{\mathbf{w}}_F = \arg \min_{\mathbf{w} \in \mathbb{R}^m} \|F(\hat{\mathbf{x}}_{t+1}^f, \mathbf{S}_t | \mathbf{w}) - \mathbf{x}_{t+1}\|_2^2. \quad (18)$$

В качестве критерия качества используется среднеквадратичная ошибка прогноза на тестовой выборке:

$$\text{MSE} = \frac{1}{h} \sum_{i=N+1}^{N+h} \frac{1}{2} \|\hat{\mathbf{x}}_i - \mathbf{x}_i\|_2^2, \quad (19)$$

где h – горизонт прогноза.

3 Вычислительный эксперимент

3.1 Цели и постановка

Вычислительные эксперименты проводились на двух видах данных: синтетический набор временных рядов и набор временных рядов потребления электроэнергии. Каждый из наборов необходимо проанализировать и для каждого из них оценить возможность улучшения прогноза с помощью матриц ковариации.

Данные для эксперимента

В качестве синтетического набора временных рядов рассматривался набор из двух зашумленных синусов различной амплитуды. Исследовалось влияние величины дисперсии нормального шума на результаты метода ковариационного сходящегося отображения, качество прогноза и связь между ними. Помимо этого, были проведены эксперименты для более сложного синтетического набора, когда ковариации временных рядов имели прямое влияние на будущее.

Набор временных рядов потребления электроэнергии был получен разделением одного временного ряда на 3 ряда потребления электроэнергии за одни и те же промежутки времени, но в разные года. Конкретнее за лето 2005-2007 годов. Эксперименты проводились на наборах из всевозможных пар рядов для сравнения результатов в зависимости от свойств каждого набора. В качестве размерности вложения для всех рядов

была выбрана половина недели $L = 84$.

Модель улучшения базового прогноза

Для построения прогностической модели, требуется выбрать f и F . В силу того, что матрицы ковариации \mathbf{C}_t ставятся в соответствие матрице векторов задержек \mathbf{X}_t (точек фазовых пространств), то в качестве f рассматриваются динамические модели прогнозирования. Другими словами, модели опирающиеся на теорию динамических систем Такенса и использующие для прогноза фазовое пространство векторов задержек. Наиболее простой динамической моделью прогнозирования является локальная модель прогнозирования [10]. В этом случае прогноз строится с помощью метода ближайших соседей в фазовом пространстве векторов задержек временного ряда.

В случае построения ковариационной модели F на прогнозах базовой, в качестве F рассматривались простая линейная модели и двухслойная, полносвязная нейронная сеть. Разбиение временных рядов для обучения представлено на Рис.3. Где history ts – используется для обучения f , F train – используется для обучения F , F test – оценка качества F , test ts – сравнение качества F и f . Гиперпараметром обучения является размер выборки для обучения ковариационной модели: F horizon. В ходе экспериментов, обучение F проводилось при различных значениях этого гиперпараметра.

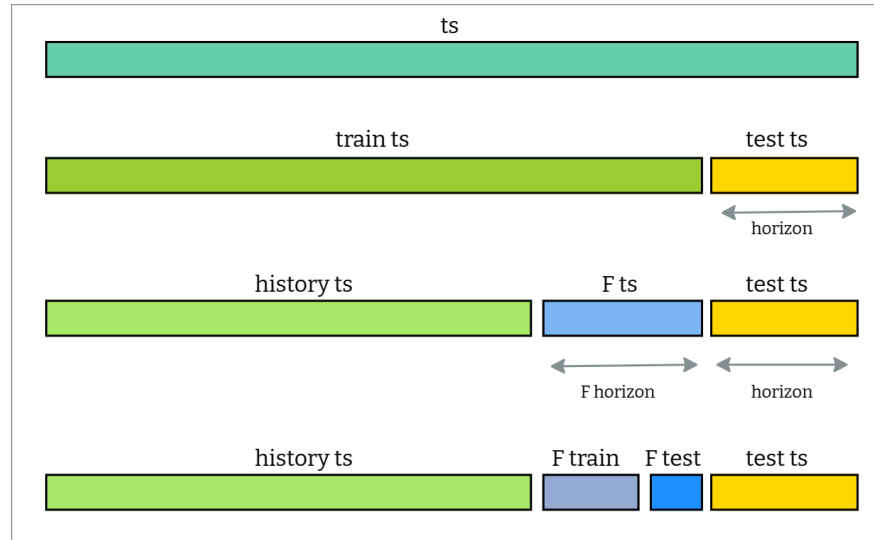


Рис. 3: Схема разделения временного ряда для обучения прогностической модели

3.2 Результаты и обсуждение

Синтетические данные

Исследовался набор из двух временных рядов:

$$x_{1t} = \sin(t) + \varepsilon_{1t}, \quad (20)$$

$$x_{2t} = 3 \sin(t) + \varepsilon_{2t}, \quad (21)$$

где $\varepsilon_{it} \sim \mathcal{N}(0, \sigma^2)$. Рассматривались несколько значений стандартных отклонений шума: $\sigma \in \{0.01, 0.1, 1, 2, 3\}$. Рассмотрим, для начала, результаты применения метода ковариационного, перекрестного отображения. В силу присутствия случайности в данных, значения корреляций для каждого значения истории N были подсчитаны несколько раз, а дальше усреднены. На Рис.4 приведены графики значений корреляций прогноза и фактических значений ряда при различных значениях дисперсии шума.

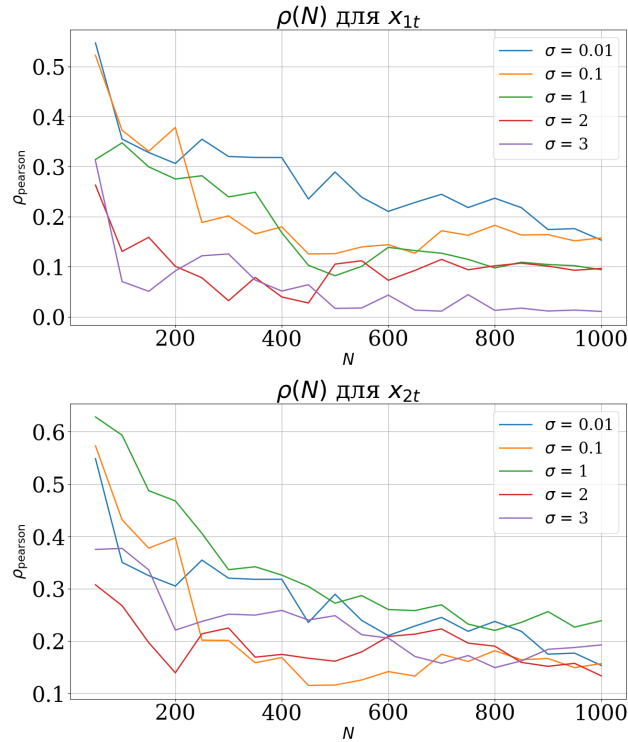


Рис. 4: Результаты метода ковариационного, перекрестного отображения для x_{1t} и x_{2t} при различных длинах истории N и различных значениях дисперсии шума

Низкие значения корреляции связаны с одинаковой периодической структурой временных рядов. Это можно понять, рассмотрев две точки фазовой траектории временного ряда. Для простоты рассмотрим незашумленный временной ряд. На Рис.5 изображены векторы задержек набора временных рядов в моменты времени t_1, t_2 . Нетрудно заметить, что соответствующие матрицы ковариации $\mathbf{C}_{t_1}, \mathbf{C}_{t_2}$ будут практически равны. Следовательно, в прогнозе для сдвинутых на целое число периодов векторов задержек и x_{it_1} , и x_{it_2} будут учитываться с одинаковым весом, что приведет к низкой корреляции прогноза и реальных значений временных рядов.

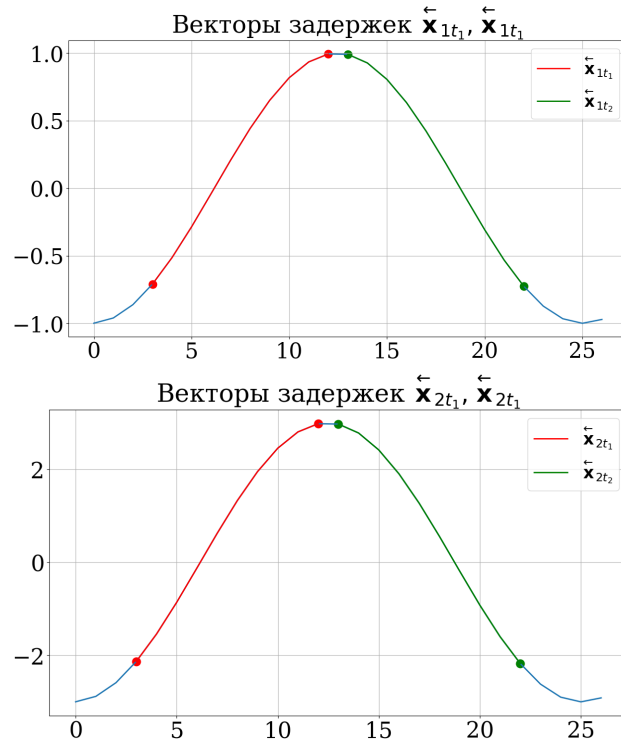


Рис. 5: Две различные точки фазовых траекторий набора временных рядов, имеющие одну и ту же матрицу ковариации

Другими словами, можно сказать, что при подсчете матрицы ковариации теряется информация о фазе сигналов. В силу этого, построение модели, улучшающей прогноз при помощи матрицы ковариации, невозможно в этом случае.

Данные потребления электроэнергии

Данные представляют из себя почасовые значения потребления электроэнергии за лето 2005, 2006, 2007. Временные ряды представлены на Рис.5. Из трех рядов было собрано три набора по два временных ряда в каждом.

Взаимосвязь рядов в каждом наборе была проанализированна с помощью метода перекрестного, сходящегося отображения. На Рис.6 представлены результаты метода в исходных фазовых пространствах рядов. Как можем видеть, несмотря высокую корреляцию перекрестного прогноза, отсутствует сходимость ρ при увеличении длины истории N . Таким образом, мы не можем утверждать о наличии причинной взаимосвязи между временными рядами, в терминах перекрестного, сходящегося отображения.

Тем не менее, используем метод ковариационного перекрестного отображения для оценки соответствия между фазовыми пространствами временных рядов и пространством матриц ковариации. Результаты метода приведены на Рис.7. Как видим, в отличие от синтетического набора рядов, для каждого из трех наборов рядов по-

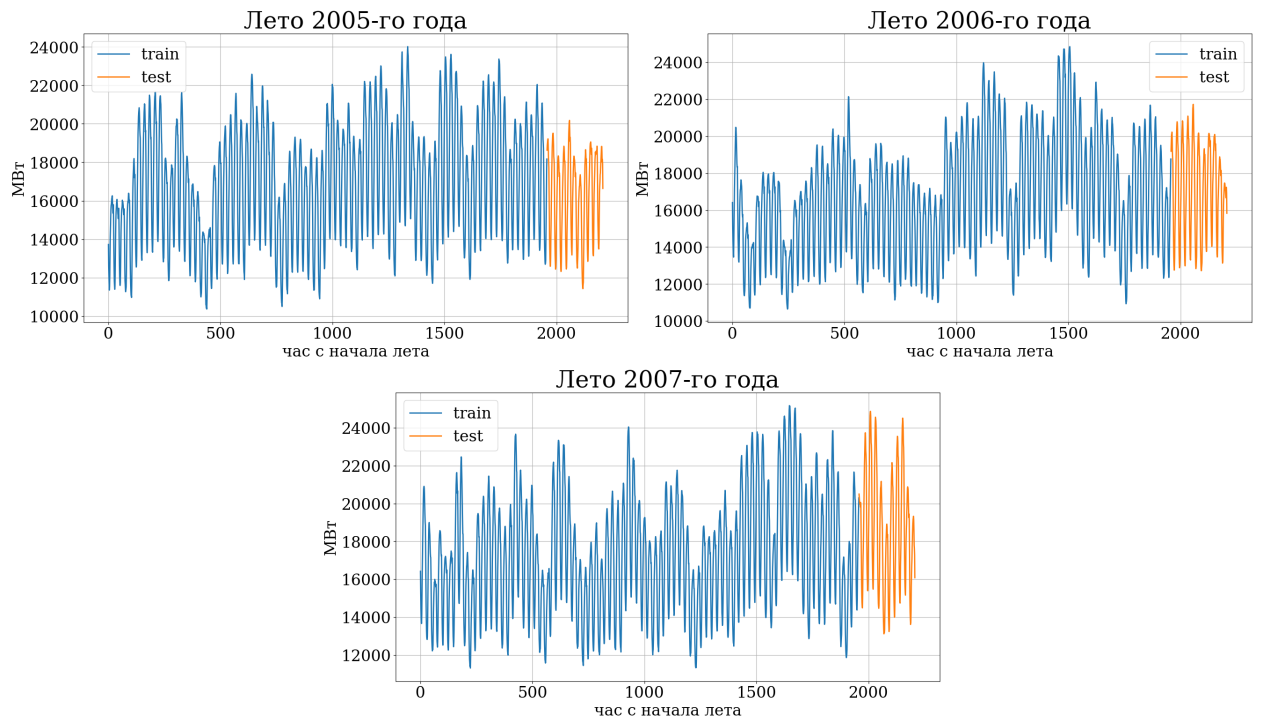


Рис. 6: Ряды потребления электроэнергии и их разделение на обучающую и тестовую выборки.

требления электроэнергии наличие высокой корреляции между ковариационным, перекрестным прогнозом и рядами наборов. Это значит, что потеря информации о фазе временных рядов менее критично и использование ковариаций для улучшения прогноза более обоснованно, чем в предыдущем случае.

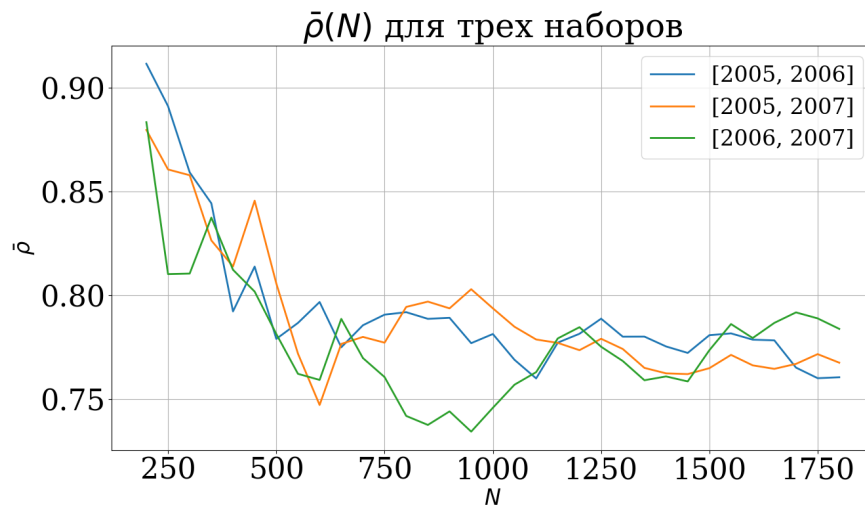


Рис. 8: Результаты метода ковариационного перекрестного отображения для трех наборов рядов

Применим метод улучшения базового прогноза, с локальной моделью прогнозирования в качестве f и линейную модель в качестве ковариационной модели. Модель была обучена для различных значений F horizon. Помимо этого, определим оптималь-

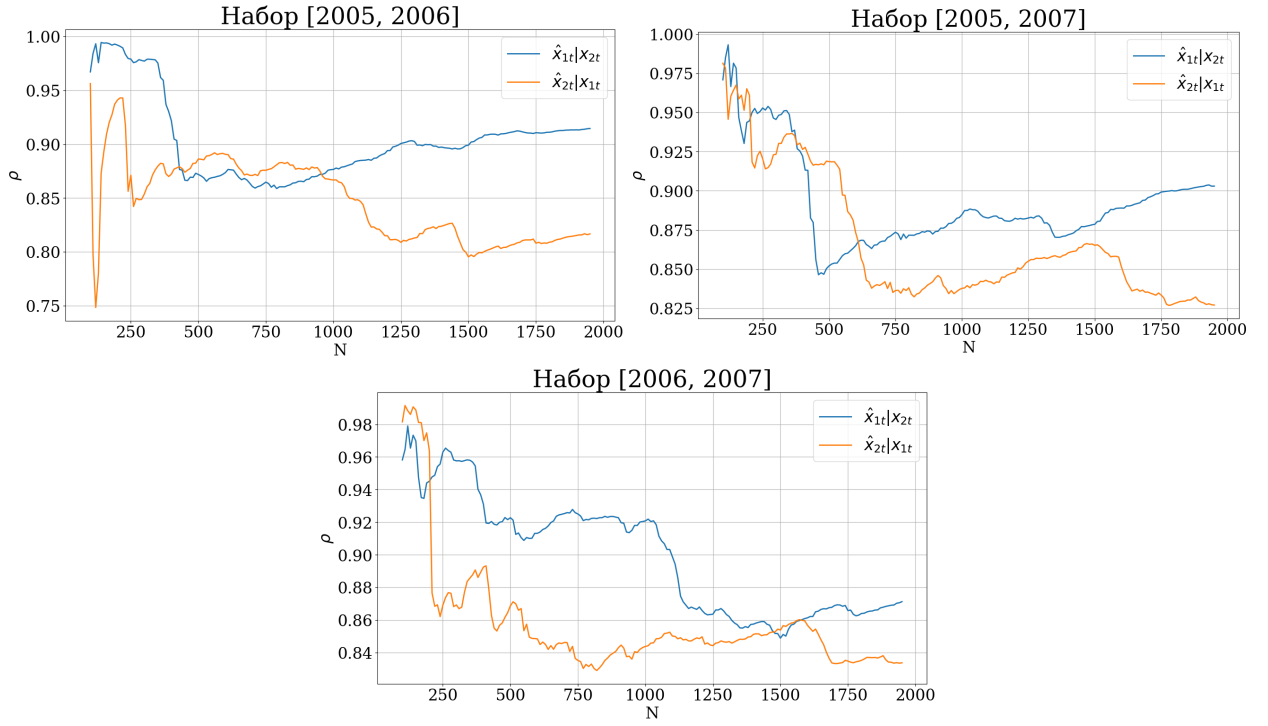


Рис. 7: Результаты метода перекрестного, сходящегося отображения для трех наборов рядов

ный горизонт для каждой модели на валидационной выборке. Заметим, что основной проблемой этого подхода является очень большая чувствительность к нестационарности как в временных рядах набора, так и в ковариациях, т.к. ковариационная модель – линейная модель, которая ожидает данные из одного и того же распределения. Однако, даже при условии стационарности, ковариационная модель имеет тенденцию к переобучению. Гиперпараметры и качество моделей приведены в Таблице 1. Результаты объясняются замечаниями, сделанными выше. Прогноз выполнялся на горизонт, равный $F \text{ horizon} * 0.1$, равный длине валидационной выборки.

	MSE_{loc}	$F \text{ horizon}_{cov}$	MSE_{cov}
[2005, 2006]	2181364	550	2486685
[2005, 2007]	701901	125	1707057
[2006, 2007]	1146901	125	7144845

Таблица 1: Качество базовой модели и модели улучшения базового прогноза

Улучшение прогноза с помощью ковариационного прогноза

Приведем пример набора временных рядов, для которого ковариационный прогноз (прогноз ближайших соседей по Римановой метрике) более точный, чем локальный прогноз в фазовом пространстве. Рассмотрим набор, в котором ковариация между точками фазовых пространств временных рядов напрямую влияет на поведение одного

из временных рядов в будущем. Пусть, x_{1t} – синус, x_{2t} – косинус небольших амплитуд. Пусть, в некоторые случайные моменты времени, на протяжении временного промежутка L , x_{2t} становится сильно антикоррелированным с x_{1t} , т.н. период зарядки системы. Прямо после этого, происходит взрыв: амплитуда и частота первого временного ряда возрастают и остаются такими на временном промежутке длины L . Потом система опять стабилизируется до следующего момента зарядки и взрыва. Приведем синтетические временные ряды на Рис.8.

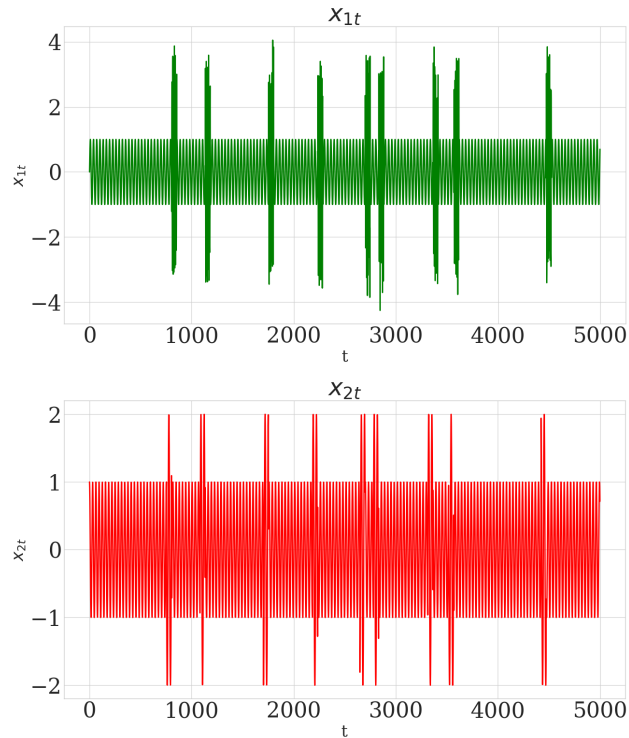


Рис. 9: Временные ряды x_{1t}, x_{2t}

Несмотря на то, что в этом случае поведение одного из временных рядов напрямую зависит от значения кросс-ковариации, прогноз x_{1t} с помощью ковариационного перекрестного отображения не приводит к лучшему результату, чем локальный прогноз. Причина этого состоит в том, что, как и в случае эксперимента с синтетическим набором синусов, потеря информации о фазе временных рядов приводит к тому, что ближайшие матрицы ковариации соответствуют далеким точкам фазовых траекторий.

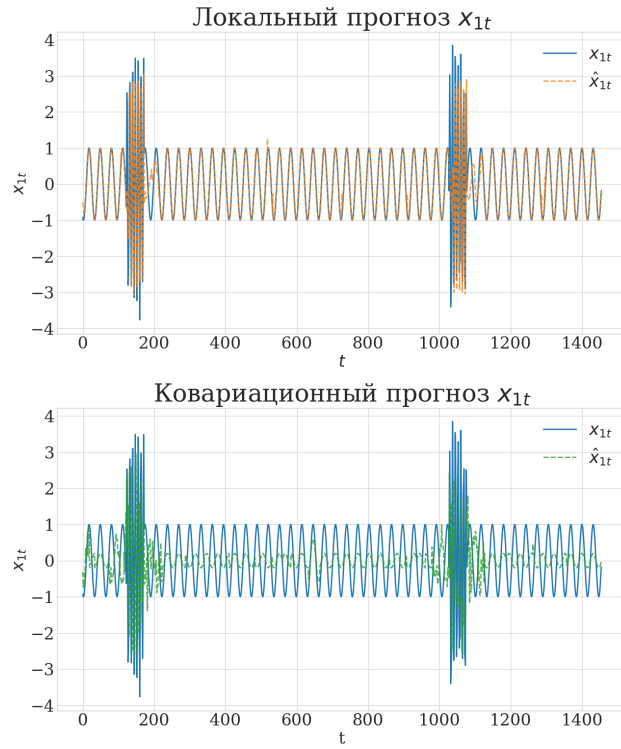


Рис. 10: Локальный и ковариационный прогноз временного ряда x_{1t}

Лучшего качества при ковариационном прогнозе удастся добиться, если прогнозировать не сам временной ряд x_{1t} , а его дисперсию в временном окне $[t + 1, \dots, t + L]$. Основным отличием от прогноза x_{1t} является то, что переходом к дисперсии мы добиваемся того, что фаза x_{1t} теперь не важна для прогноза. Например, точкам фазовых траекторий в моменты времени t_1, t_2 на Рис.5 теперь будет соответствовать одинаковое целевое значение. Ухудшение локального прогноза, в свою очередь, связано с тем, что, т.к. предсказывается дисперсия за следующие L моментов времени, одинаковым точкам фазовых траекторий могут соответствовать совершенно разные значения дисперсии (текущая точка фазовой траектории не пересекается с промежутком подсчета дисперсии). Результаты прогнозов приведены на Рис.11:

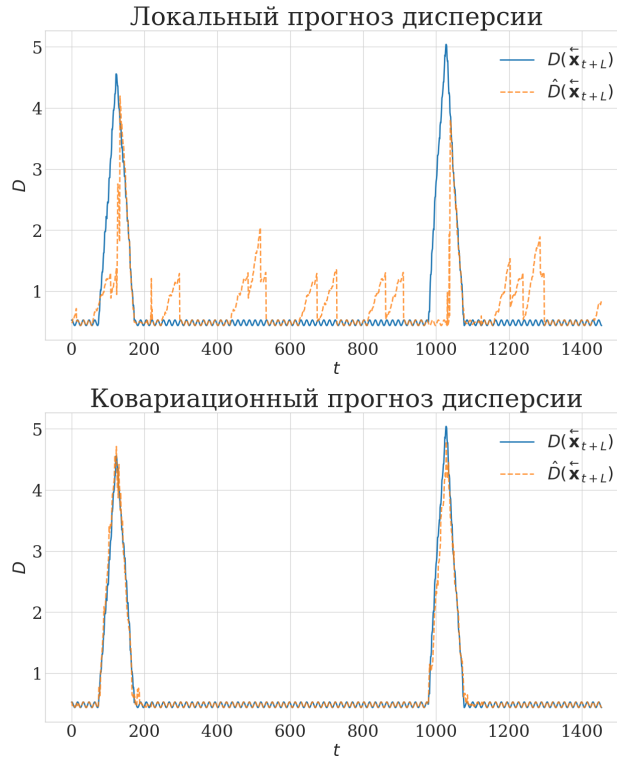


Рис. 11: Локальный и ковариационный прогноз дисперсии x_{1t}

4 Заключение

Главным препятствием к улучшению прогноза набора временных рядов с помощью матриц ковариации является потеря информации о фазе сигналов. Это приводит к тому, что близким, в терминах Римановой метрики, матрицам ковариации могут соответствовать далекие точки фазовых траекторий временных рядов. А т.к. модель не может отличить два разных состояния, то и не может учесть более сложные взаимосвязи между точками фазовых пространств. Помимо этого, даже когда потеря информации о фазе не настолько критична (как в случае наборов потребления электроэнергии), для улучшения прогноза необходимо, чтобы матрица ковариации в данный момент времени имела явное влияние на будущее ряда/рядов. В противном случае, матрицы ковариации, просто напросто, не несут в себе дополнительной информации к описанию временных рядов в фазовых пространствах. Важность этих условий особо подчеркивается последним примером: при переходе к прогнозу дисперсии, качество ковариационного прогноза улучшилось, т.к. информация о фазе перестала иметь значение для прогноза. В то же время, условия нарушились для локального прогноза. Точка фазовой траектории в данный момент перестала иметь влияние на будущее поведение прогнозируемого ряда, т.к. оно определялось только ковариацией рядов набора. Это и привело к пло-

хому качеству прогноза в этом случае. Эксперимент с моделью улучшения базового прогноза на наборе потребления электроэнергии показал, что использование линейной ковариационной модели не позволяет улучшить прогноз и более обоснованным является использование локальных свойств пространства матриц ковариации.

Список литературы

1. *Moakher M.* A differential geometric approach to the geometric mean of symmetric positive-definite matrices // SIAM journal on matrix analysis and applications. — 2005. — Т. 26, № 3. — С. 735—747.
2. Riemannian geometry applied to BCI classification / A. Barachant [и др.] // International conference on latent variable analysis and signal separation. — Springer. 2010. — С. 629—636.
3. Multiclass brain–computer interface classification by Riemannian geometry / A. Barachant [и др.] // IEEE Transactions on Biomedical Engineering. — 2011. — Т. 59, № 4. — С. 920—928.
4. *Congedo M., Barachant A., Bhatia R.* Riemannian geometry for EEG-based brain–computer interfaces; a primer and a review // Brain-Computer Interfaces. — 2017. — Т. 4, № 3. — С. 155—174.
5. Classification of covariance matrices using a Riemannian-based kernel for BCI applications / A. Barachant [и др.] // Neurocomputing. — 2013. — Т. 112. — С. 172—178.
6. *Granger C. W.* Investigating causal relations by econometric models and cross-spectral methods // Econometrica: journal of the Econometric Society. — 1969. — С. 424—438.
7. *Lütkepohl H.* Introduction to multiple time series analysis. — Springer Science & Business Media, 2013.
8. Detecting causality in complex ecosystems / G. Sugihara [и др.] // science. — 2012. — Т. 338, № 6106. — С. 496—500.
9. *Takens F.* Detecting strange attractors in turbulence // Dynamical Systems and Turbulence, Warwick 1980: proceedings of a symposium held at the University of Warwick 1979/80. — Springer. 2006. — С. 366—381.
10. *McNames J. N.* Innovations in local modeling for time series prediction. — stanford university, 1999.