

## ОЦЕНКА ВОЗМОЖНОСТЕЙ МЕТОДА АНАЛОГОВ ДЛЯ ТЕКУЩЕГО ПРОГНОЗА ТЕМПЕРАТУРЫ ВОЗДУХА

Восканян К.Л., Кузнецов А.Д., Сероухова О.С., Симакина Т.Е.  
Российский государственный гидрометеорологический университет,  
г. Санкт-Петербург

---

*Поступила в редакцию 05.03.2019, после переработки 23.03.2019.*

---

Рассматривается алгоритм прогнозирования нестационарных временных рядов, основанный на методе аналогов. Предлагаемый метод тестируется на значениях температуры воздуха по данным автоматической метеорологической станции. Построены шесть моделей прогноза, использующих различные меры сходства, определена ошибка прогноза на независимой выборке. Сравнение с оценкой инерционного прогноза позволила прийти к заключению, что метод аналогов дает удовлетворительные результаты текущего прогнозирования температуры.

**Ключевые слова:** метод аналогов, текущий прогноз температуры, меры близости.

*Вестник ТвГУ. Серия: Прикладная математика. 2019. № 1. С. 101–116. <https://doi.org/10.26456/vtprmk528>*

### Введение

Несмотря на значительное число публикаций по разработке методов и практическому применению прогнозирования временных рядов, задача получения адекватного прогноза временного ряда остается по-прежнему актуальной, поскольку существующие методы часто дают слишком большую ошибку [1,2]. Это особенно справедливо для нестационарных рядов, которыми в большинстве случаев являются временные ряды, содержащие данные измерений метеорологических величин. Одним из возможных способов прогнозирования является метод аналогов. Отметим, что метод аналогов может использоваться в самых различных предметных областях [3,4].

Метод аналогов, иллюстрация которого представлена на Рис. 1, заключается в поиске в предыстории временного ряда длины  $n$  среди всех векторов размерности  $l$ , составленных из всех возможных наборов отрезков этого временного ряда  $(f_i, f_{i+1}, \dots, f_{i+l-1})$ , одного или нескольких векторов, наиболее «похожих» на последний в рассматриваемой выборке вектор  $(f_{n-l+1}, f_{n-l+2}, \dots, f_n)$ . При этом мера сходства («похожести») определяется с помощью задания соответствующей метрики. Продолжение во времени такого отрезка-аналога, т.е. его последующие значения должны быть близки (с точки зрения использованного критерия «похожести») на соответствующее продолжение значений текущего отрезка, а эти значения — это прогностические значения временного ряда [5].

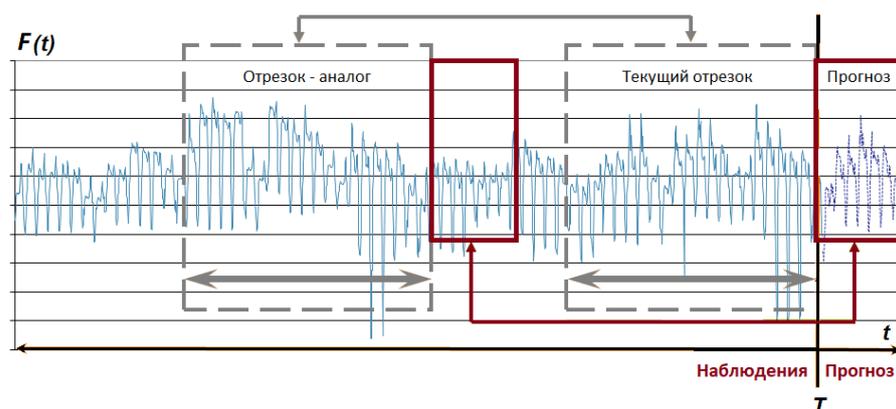


Рис. 1: Иллюстрация выявления во временном ряде отрезка, аналогичного текущему (содержащему последние измеренные значения), и построения прогностического отрезка по его продолжению

Таким образом, основные этапы метода аналогов следующие:

- поиск и выбор аналога, построение модели и её исследование,
- экстраполяция данных с аналога на изучаемый объект,
- верификация экстраполяционных выводов.

## 1. Поиск и выбор аналога

Меры сходства можно подразделить на четыре вида [6–9]:

- коэффициенты корреляции;
- меры расстояния;
- коэффициенты ассоциативности;
- вероятностные коэффициенты сходства.

Наиболее широкое распространение получили меры расстояния и коэффициенты корреляции, поэтому в данной работе рассматриваются именно они. Два отрезка временного ряда идентичны, если описывающие их переменные принимают одинаковые значения. В этом случае расстояние между ними равно нулю, а коэффициент корреляции равен 1.

Рассмотрим наиболее известные методы оценки расстояния между отрезками временного ряда (метрики), заданными векторами  $\mathbf{x}$  и  $\mathbf{y}$ , и методику их применения при реализации метода аналогов.

1. Стандартная Евклидова метрика:

$$\rho_{Ev}(\mathbf{x}, \mathbf{y}, n) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \text{ или } \rho_{Ev}(\mathbf{x}, \mathbf{y}) = \sqrt{(\mathbf{x} - \mathbf{y})^T (\mathbf{x} - \mathbf{y})}, \quad (1)$$

где  $\mathbf{x}$  и  $\mathbf{y}$  — значения временного ряда, исследуемого на принадлежность к аналогу отрезка и контрольного отрезка-эталона соответственно;  $n$  — длина векторов  $\mathbf{x}$  и  $\mathbf{y}$  — аналога и эталона.

Тот отрезок  $\mathbf{x}$ , для которого значение среднего квадратического отклонения окажется минимальным, принимается за аналог контрольному отрезку временного ряда  $\mathbf{y}$ .

2. Квадрат Евклидова расстояния, когда возведение в квадрат стандартного Евклидова расстояния придает большие веса более отдаленным друг от друга значениям:

$$\rho_{Ev}(\mathbf{x}, \mathbf{y}, n) = \sum_{i=1}^n (x_i - y_i)^2 \text{ или } \rho_{Ev}(\mathbf{x}, \mathbf{y}) = (\mathbf{x} - \mathbf{y})^T (\mathbf{x} - \mathbf{y}). \quad (2)$$

3. Манхэттенское расстояние, или «расстояние городских кварталов» (city-block), определяется следующим образом:

$$\rho_M(\mathbf{x}, \mathbf{y}, n) = \sum_{i=1}^n |x_i - y_i|. \quad (3)$$

Манхэттенское расстояние позволяет уменьшать влияние отдельных больших выбросов.

4. Относительное расстояние, при котором метрика определяется относительной разностью совпадения двух векторов (двух отрезков временного ряда):

$$\rho(\mathbf{x}, \mathbf{y}, n) = \frac{1}{q} \sum_{i=1}^n \frac{|x_i - y_i|}{|x_i + y_i|}. \quad (4)$$

5. Коэффициент корреляции, определяющий степень взаимного влияния изменения двух векторов (двух отрезков временного ряда):

$$r(\mathbf{x}, \mathbf{y}, n) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}. \quad (5)$$

Следующие метрики содержат эмпирические параметры, позволяющие придать их использованию более универсальный подход к определению степени близости векторов, однако задание таких параметров является нетривиальной задачей.

6. Степенное расстояние, где  $r$  и  $p$  — параметры, определяемые пользователем:

$$\rho_c(\mathbf{x}, \mathbf{y}, n, p, r) = \sum_{i=1}^n (|x_i - y_i|^p)^{1/r}. \quad (6)$$

Данная метрика позволяет прогрессивно увеличить или уменьшить вес, относящийся к размерности, для которой соответствующие значения сильно отличаются. Параметр  $p$  ответственен за постепенное взвешивание разностей по отдельным координатам, параметр  $r$  — за прогрессивное взвешивание больших расстояний между объектами. Если оба параметра:  $r$  и  $p$  — равны двум, то это расстояние совпадает с расстоянием Евклида.

7. Метрика Минковского:

$$\rho_M(\mathbf{x}, \mathbf{y}, n, p) = \left[ \sum_{i=1}^n |x_i - y_i|^p \right]^{1/p}. \quad (7)$$

В рассмотренных схемах сравнения отрезков временного ряда предполагалась равноценность элементов отрезков с точки зрения определения их близости и, тем самым, их равноценность для последующего прогностического отрезка. В этом случае алгоритмы сравнения упрощаются. Однако такое уравнивание не учитывает уменьшения ценности в нем элементов, содержащихся в эталонном отрезке, по мере удаления от его «правого» края, т. е. по мере удаления от последнего (до начала текущего прогноза) имеющегося во временном ряде элемента.

Частично этот недостаток устраняется при использовании следующей метрики, предполагающей введение весовых коэффициентов для элементов сравниваемых между собой векторов.

8. Диагонально взвешенная Евклидова метрика:

$$\rho_{E\nu W}(\mathbf{x}, \mathbf{y}) = \sqrt{(\mathbf{x} - \mathbf{y})^T \Lambda (\mathbf{x} - \mathbf{y})}, \quad (8)$$

где матрица  $\Lambda = \text{diag}(\lambda)$ .

Для диагональной матрицы последнее соотношение можно записать в следующем виде:

$$\rho_{E\nu W}(\mathbf{x}, \mathbf{y}, n, \lambda_{ii}) = \sqrt{\sum_{i=1}^n \lambda_{ii} (x_i - y_i)^2}. \quad (9)$$

Для диагонально взвешенной Евклидовой метрики общим предположением является то, что конец предыстории для прогноза более важен, чем его начало. Поэтому параметры  $\lambda_{ii}$  увеличиваются с порядковым номером  $i$ . В частности, можно предположить, что последовательность весовых параметров  $\lambda_{ii}$  имеет степенной вид:

$$\lambda_{ii} = \lambda_1^{n-i+1}, \quad 0 < \lambda_1 < 1, \quad i = 1, \dots, n. \quad (10)$$

Тогда для определения матрицы  $\Lambda$  достаточно задать, например,  $\lambda_{11} = \lambda_1$ . Пример задания диагональных элементов матрицы  $\Lambda$  для трех значений параметра  $\lambda_1$  приведен в Таблице 1.

Таблица 1: Диагональные элементы матрицы  $\Lambda$  для трех значений параметра  $\lambda_1$

$\lambda_{ii} =>$	$\lambda_{1,1}$	$\lambda_{2,2}$	$\lambda_{3,3}$	$\lambda_{4,4}$	$\lambda_{5,5}$
$\lambda_1 = 0.2$	0.0003	0.0016	0.008	0.04	0.2
$\lambda_1 = 0.5$	0.0312	0.0625	0.125	0.25	0.5
$\lambda_1 = 0.667$	0.132	0.198	0.296	0.444	0.667

Рассмотренный вариант задания диагональных элементов матрицы  $\Lambda$  не связан со свойствами имеющейся выборки временного ряда и подразумевает эвристический подбор этих параметров. Следовательно, такой подход к определению метрики оставляет неопределенность в задании параметров матрицы  $\Lambda$ . Эту неопределенность в реализации расчетов с использованием диагонально взвешенной Евклидовой метрики можно преодолеть, определив оптимальное значение параметра  $\lambda_1$  на основе проведения серии численных экспериментов на имеющемся архиве экспериментальных данных.

9. Расстояние Махаланобиса, формула которого имеет следующий вид:

$$\rho_{\Sigma}(\mathbf{x}, \mathbf{y}) = (\mathbf{x} - \mathbf{y})^T \Sigma^{-1} (\mathbf{x} - \mathbf{y}), \quad (11)$$

где  $\Sigma$  — общая внутригрупповая дисперсионно-ковариационная матрица. В отличие от метрик Минковского и Евклидовой, эта метрика связана с корреляциями переменных, задаваемых с помощью матрицы дисперсий-ковариаций.

Из представленных выше 9-ти оценок близости аналога к эталонному отрезку временного ряда в работе рассматривались подходы под номерами 1, 3, 4, 5, 7 и 8.

## 2. Прогноз с использованием аналога

После нахождения отрезка-аналога осуществляется этап экстраполяции его продолжения на заданное количество шагов. Рассмотрены два способа такой экстраполяции.

### 2.1 Адаптивная композиция моделей прогнозирования

Если находится только один «ближайший» вектор, то прогноз определяется как следующие за ближайшим отрезком-аналогом на длину заблаговременности прогноза. Если при поиске аналогов находятся  $k$  аналогов, то для прогноза могут использоваться все  $k$  продолжений отрезков-аналогов. При использовании адаптивной композиции моделей (АКМ, гибридная АКМ) прогноз формируется как

взвешенная сумма прогнозов, полученных по альтернативным моделям. Такой подход удобен в использовании для случаев, когда возникают трудности с однозначным выбором одной определенной структуры модели.

Применение АКМ возможно в следующих случаях:

1. последовательное применение одной методической прогностической модели;
2. применение нескольких прогностических моделей;
3. применения методической прогностической модели и инерционного прогноза и т.п.

Поскольку в данной работе рассматривается реализация 6 метрик для поиска аналогов, то модель адаптивной композиции может быть применена независимо для каждой метрики с использованием нескольких прогностических отрезков временного ряда: оптимального и квазиоптимальных. В этом случае прогноз рассматривается как суперпозиция найденных аналогов отдельно по каждой из 6 метрик с весами, рассчитанными с использованием ранжированных расстояний эталона  $\mathbf{y}$  от оптимального ( $i = 1$ ) и квазиоптимальных ( $i = 2, 3, \dots, k$ ) векторов по следующим соотношениям [10]:

$$\mathbf{x}^* = \sum_{j=1}^k w_j \mathbf{x}_j^*, \quad (12)$$

где  $\mathbf{x}^*$  — прогностическое продолжение вектора  $\mathbf{x}$ ,

$$w_i = \frac{W_i}{\sum_{j=1}^k W_j}, \quad W_i = \left\{ 1 - \left[ \frac{\rho(\mathbf{x}_i, \mathbf{y})}{\rho(\mathbf{x}_{k+1}, \mathbf{y})} \right]^2 \right\} \left\{ 1 - \left[ \frac{\rho(\mathbf{x}, \mathbf{y})}{\rho(\mathbf{x}_{k+1}, \mathbf{y})} \right]^2 \right\}. \quad (13)$$

При расчете весовых коэффициентов предполагается, что ранжирование отрезков-аналогов идет в порядке возрастания значений  $\rho(\mathbf{x}, \mathbf{y})$  — чем меньше эта величина, тем «ближе» отрезок-аналог  $\mathbf{x}$  к эталонному отрезку  $\mathbf{y}$ . Такой подход справедлив для всех метрик, кроме коэффициента корреляции: чем больше по абсолютной величине коэффициент корреляции, тем «ближе» отрезки. Поэтому в этом случае при расчете весовых коэффициентов использовался параметр

$$\rho_r(\mathbf{x}, \mathbf{y}) = 1 - |\mathbf{r}(\mathbf{x}, \mathbf{y})|. \quad (14)$$

В формуле (13) для  $W_i$  знаменатель дроби  $\rho(\mathbf{x}_{k+1}, \mathbf{y})$  — это расстояние до аналога  $\mathbf{x}_{k+1}$ , которое определяется, но соответствующий этому аналогу прогностический отрезок в расчете  $\mathbf{x}^*$  не участвует.

## 2.2 Коррекция прогностического отрезка, найденного по методу аналогов

Предположим, что значения эталонного отрезка  $y(i)$  и отрезка-аналога  $x(i)$  связаны следующим соотношением:

$$x(i) = k_1 y(i) + k_0 + \varepsilon, \quad i = 1, 2, \dots, m, \quad (15)$$

где  $\varepsilon$  — невязка (погрешность такого представления), которую при определении коэффициентов  $k_1$  и  $k_0$  нужно минимизировать. Такую минимизацию можно получить, используя метод наименьших квадратов. Тогда, решая уравнение

$$x(i) = k_1 y(i) + k_0 \quad (16)$$

относительно  $y(i)$  и заменяя его на  $x^*(i)_{\text{кор}}$ , для скорректированного прогноза можно записать:

$$x^*(i)_{\text{кор}} = x^*(i)/k_1 - k_0, \quad i = 1, 2, \dots, m. \quad (17)$$

### 3. Иллюстрация возможностей выбора аналогов

Для демонстрации возможностей метода аналогов для метеорологического прогноза среди возможных источников временных рядов метеорологических величин были выбраны данные, получаемые с помощью автоматических метеорологических станций [11,12]. Был использован временной ряд температуры воздуха с автоматической метеорологической станции Санкт-Петербурга за сентябрь 2018 г. (2878 значений) с дискретностью 15 мин, представленный на Рис. 2.



Рис. 2: Временной ряд температуры воздуха в Санкт-Петербурге с 01.09.2018 г. по 30.09.2018 г. с дискретностью 15 мин.

Программа, реализующая прогнозирование по методу аналогов, после задания положения и размера эталонного отрезка позволяет:

- найти 4 отрезка-аналога с использованием 6 метрик и рассчитать статистические характеристики их близости;
- по продолжению отрезков-аналогов найти соответствующие им прогностические отрезки и сравнить точность таких прогнозов с инерционными прогнозами;
- произвести корректировку прогностических отрезков на основе расчета коэффициентов  $k_0$  и  $k_1$  (см. соотношения (15) – (17));

– осуществить адаптивную композицию (см. соотношения (12) – (13)) с использованием 3 из полученных 4 прогностических отрезков для каждой из 6 метрик.

Фрагмент данных, получаемых в процессе работы программы, приведен на Рис. 3.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S			
1	Начальный №	СКО		№	Эталон	№	Аналог №	№	Измерен №	7	1644	Прогноз	№	Измерен №	7	2396	Инерционный	№	Измерен №	7	1644	Скорректиро
2	ряда-аналога	рд-аналог			2370	10,3	1618	10,3	2396	7,1	1645	7,7	2396	7,1	2397	7,6	2396	7,1	1644	7,6		
3		0,19			2371	10,2	1619	10,2	2397	7,1	1645	7,7	2397	7,1	2397	7,6	2397	7,1	1645	7,6		
4		Средняя разность			2372	10,1	1620	10,4	2398	7,1	1646	7,8	2398	7,1	2398	7,6	2398	7,1	1646	7,7		
5		рд-аналог			2373	10,1	1621	10,5	2399	6,8	1647	7,6	2399	6,8	2399	7,6	2399	6,8	1647	7,4		
6		-0,11			2374	10	1622	10,4	2400	6,5	1648	7,7	2400	6,5	2400	7,6	2400	6,5	1648	7,6		
7	№ метрики =>	1			2375	10	1623	10,1	2401	6,4	1649	7,6	2401	6,4	2401	7,6	2401	6,4	1649	7,4		
8	Метрика =>	минимум			2376	10,1	1624	10	2402	6,1	1650	7,7	2402	6,1	2402	7,6	2402	6,1	1650	7,6		
9	абсолют разности				2377	10	1625	9,9	2403	6	1651	8	2403	6	2403	7,6	2403	6	1651	7,9		
10	(Манхэттенская)				2378	9,8	1626	9,8	2404	5,8	1652	8,3	2404	5,8	2404	7,6	2404	5,8	1652	8,2		
11	Используемый	3			2379	9,6	1627	9,7	2405	5,9	1653	8,6	2405	5,9	2405	7,6	2405	5,9	1653	8,6		
12	№ колонки на Листе'				2380	9,4	1628	9,6	2406	5,7	1654	9,1	2406	5,7	2406	7,6	2406	5,7	1654	9,0		
13	Длина				2381	9,4	1629	9,5	2407	5,7	1655	8,9	2407	5,7	2407	7,6	2407	5,7	1655	8,8		
14	отрезка				2382	9,3	1630	9,4	2408	5,8	1656	9	2408	5,8	2408	7,6	2408	5,8	1656	8,9		
15	av=>	26			2383	9,4	1631	9,2	2409	5,6	1657	9,5	2409	5,6	2409	7,6	2409	5,6	1657	9,4		
16	НН для отрезка =>	2370	2395		2384	9	1632	9	2410	5,6	1658	9,8	2410	5,6	2410	7,6	2410	5,6	1658	9,7		
17	Заблаговременность прогноза	18			2385	8,8	1633	8,9	2411	5,6	1659	10,5	2411	5,6	2411	7,6	2411	5,6	1659	10,5		
18	Поиск отрезка-аналога				2386	8,7	1634	8,8	2412	5,7	1660	10,4	2412	5,7	2412	7,6	2412	5,7	1660	10,4		
19	начинается с № =>				2387	8,7	1635	8,5	2413	6	1661	10,7	2413	6	2413	7,6	2413	6	1661	10,7		
20	Суммарное число данных				2388	8,5	1636	8,4	Результ использо аналог для ТП			Результ использо инерция для ТП			Результ примен коррекц для ТП							
21	в колонке 3				2389	9,4	1637	8,4	СКО			СКО			СКО							
22	Оптим №1				2390	7,9	1638	8,4	прогноз 1,46 ошибка			прогноз 0,52 ошибка n			прогноз -1,47							
23	Оптим №2				2391	7,9	1639	8,3														
24	Оптим №3				2392	8,3	1640	8,2														
25	Оптим №4				2393	8,1	1641	8,1														
26	Результ поиска				2394	7,7	1642	8														
27	колонки				2395	7,6	1643	7,9														
28	аналога:				Результ использо аналог для ТП																	
29	колонки																					
30	аналога:																					
31	колонки																					
32	аналога:																					

Рис. 3: Пример результатов расчетов при использовании диагонально взвешенной Евклидовой метрики ( $\lambda_1 = 0,8$ , см. (10))

Первые три столбца несут информацию о задании исходных параметров, следующие четыре содержат данные измерений и соответствующие им прогностические значения, затем — результаты инерционного прогноза и, наконец, последние столбцы — результат скорректированного прогноза. Такое представление данных производится для всех 6 метрик и для 4 последовательно найденных отрезков-аналогов.

На Рис. 4 приведен пример сопоставления эталонного отрезка и двух найденных отрезков-аналогов при использовании диагонально взвешенной Евклидовой метрики с  $\lambda_1 = 0,8$ .

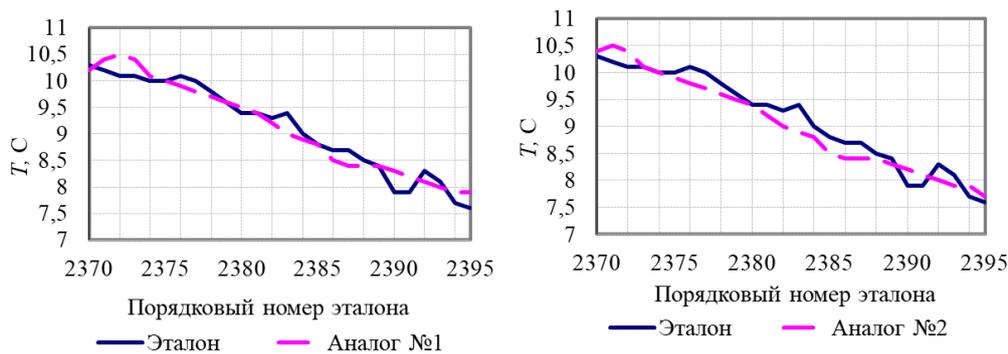


Рис. 4: Сопоставление эталонного отрезка и аналогов №1 и №2

Первый аналог начинается с номера 1618, средняя разность с эталоном составила  $-0.22$  °C, среднеквадратичное отклонение  $0.24$  °C. Второй аналог начинается с номера 1020, средняя разность этих отрезков составляет  $0.21$  °C, среднеквадратичное отклонение  $0.08$  °C.

Рис. 5 иллюстрирует изменчивость отрезков-аналогов при их ранжировании по значению Манхеттенской метрики, т.е. при последовательном увеличении величины метрики от ее минимального значения. На рисунке представлен эталонный отрезок и четыре его аналога.

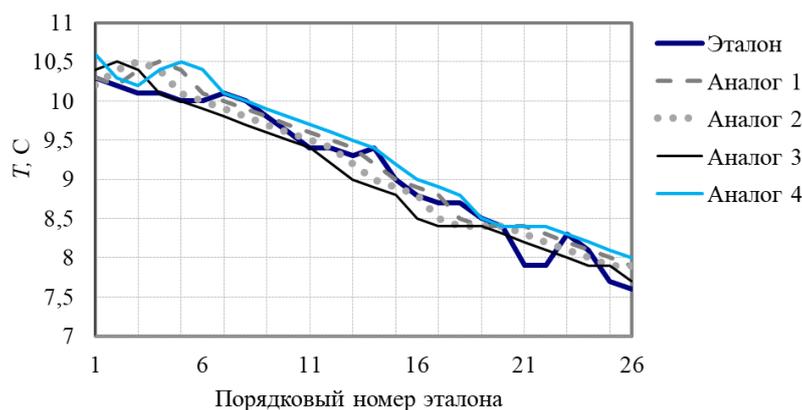


Рис. 5: Сопоставление эталонного и четырех отрезков-аналогов временного ряда температуры воздуха с дискретностью 15 мин при использовании Манхеттенской метрики

Изменение статистических характеристик, полученных при сравнении эталона с аналогами, представлено в Таблице 2.

Таблица 2: Статистические характеристики «близости» эталона и аналогов по Манхеттенской метрике

	Аналог №1	Аналог №2	Аналог №3	Аналог №4
Средняя ошибка, °C	0.05	0.08	0.12	0.16
СКО, °C	0.37	0.44	0.51	0.59
Начальный №	2338	2337	2336	2335

Влияние выбора метрики на результат поиска аналога демонстрирует Рис. 6. Аналог №1 на Рис. 6а получен при использовании стандартной Евклидовой метрики (формула (2)). Средняя разность с эталоном составила  $0.05$  °C, а СКО  $0.37$  °C. Рис. 6б представляет результат использования в качестве метрики коэффициента корреляции. В этом случае средняя разность аналога и эталона составила  $-0.64$  °C, а СКО  $0.15$  °C. На Рис. 6в приведено сопоставление эталонного и аналогового отрезков при использовании диагонально взвешенной Евклидовой метрики с параметром  $\lambda_1 = 0.8$  (формула (9)). В этом случае средняя разность двух отрезков

составила  $-0.64$  °С, а СКО  $0.79$  °С. На Рис. 6г пример метрики Махаланобиса 8 (формула (11)), средняя разность  $-0.52$  °С, а СКО  $0.71$  °С.

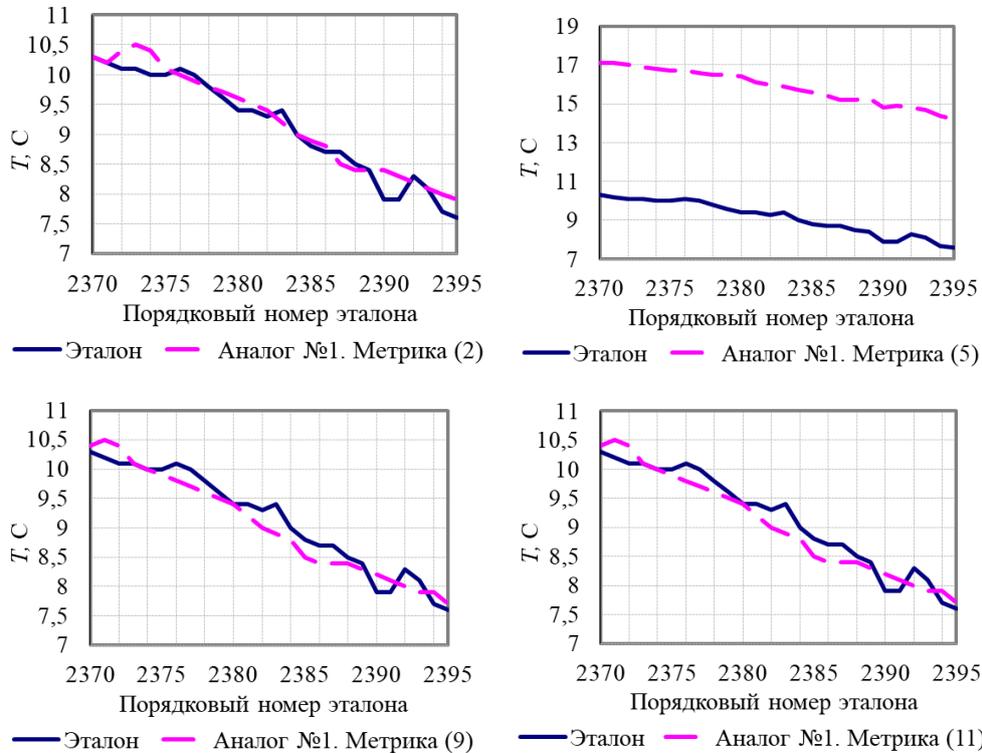


Рис. 6: Сопоставление эталонного отрезка и аналога №1, найденного с помощью различных метрик

#### 4. Верификация прогноза по методу аналогов

Прогноз производился на основе поиска оптимальных аналогов, а также по методу АКМ (формулы (12)-(14)) и экстраполяцией с линейной коррекцией (формулы (15)-(17)).

На Рис. 7 приведен пример сопоставления инерционного и методического прогнозов с данными измерений при использовании диагонально взвешенной Евклидовой метрики с  $\lambda_1 = 0.8$ . Здесь прогнозом является продолжение отрезка-аналога №1. Статистические характеристики точности методического и инерционного прогноза, приведенные в Таблице 3, показывают, что метод аналогов успешен, поскольку значительно превосходит по своей точности инерционный прогноз.

На Рис. 8 приведены графики, иллюстрирующие изменение прогностического отрезка при использовании четырех аналогов. Аналоги найдены с помощью Манхеттенской метрики. Интересной особенностью полученных прогнозов является практическая независимость точности прогноза от заблаговременности. С ростом

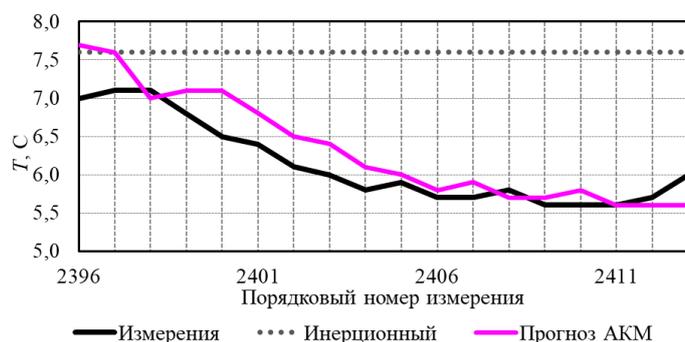


Рис. 7: Сопоставление данных измерений с результатами инерционного прогноза и метода аналогов при использовании диагонально взвешенной Евклидовой метрики с  $\lambda_1 = 0.8$ .

Таблица 3: Статистические характеристики точности методического и инерционного прогнозов

Показатели	Методический прогноз	Инерционный прогноз
Средняя ошибка, °C	-0.20	-1.47
СКО, °C	0.27	0.52

заблаговременности прогноза его погрешность может даже уменьшаться. В Таблице 4 представлены статистические характеристики точности прогноза, усредненной по временному промежутку в 270 мин при использовании отрезков-аналогов, представленных на Рис. 5.

Таблица 4: Статистические характеристики точности прогноза с использованием последовательного нахождения аналогов

Показатели	Прогноз по аналогу №1	Прогноз по аналогу №2	Прогноз по аналогу №3	Прогноз по аналогу №4
Средняя ошибка, °C	-0.17	-0.25	-0.35	-0.43
СКО, °C	0.13	0.16	0.17	0.16
Начальный № прогностического отрезка	2364	2363	2362	2361

Рис. 9 содержит пример использования коррекции прогностического отрезка на основе расчета коэффициентов  $k_1$  и  $k_0$  (см. (15) – (17)). Влияние коррекции на прогностический отрезок заметно, но незначительно: при переходе от прогностического отрезка к скорректированному средняя ошибка уменьшается с  $-0.2$  °C на

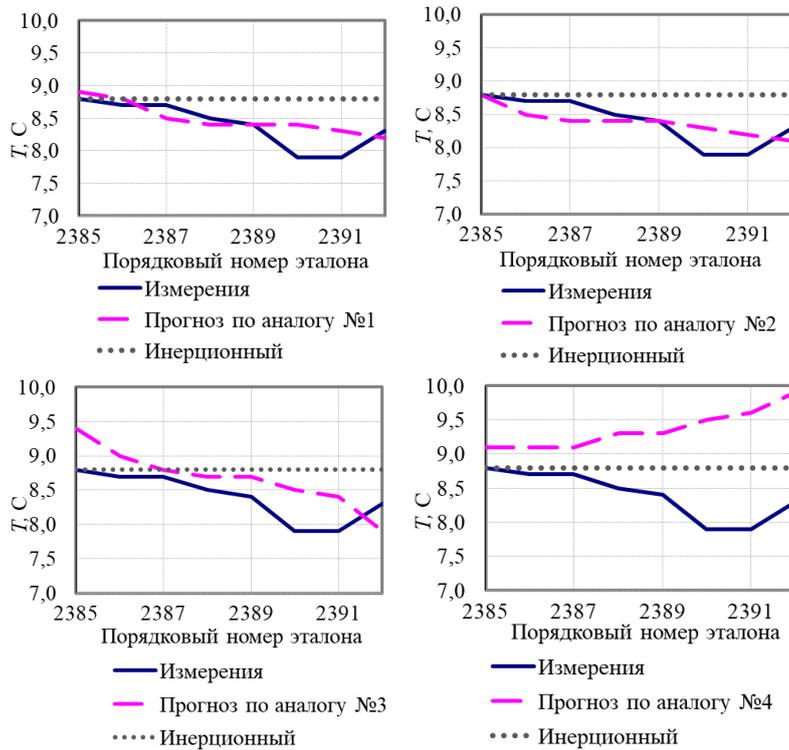


Рис. 8: Сопоставление фактических значений и четырех прогнозов температуры воздуха с заблаговременностью два часа. Прогноз выполнен методом АКМ по четырем аналогам, найденным Манхеттенской метрикой

0.03  $^{\circ}\text{C}$ , тогда как СКО, наоборот, увеличивается с 0.27  $^{\circ}\text{C}$  до 0.29  $^{\circ}\text{C}$ .

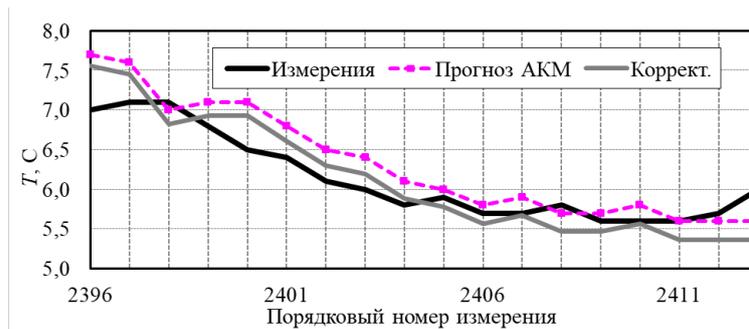


Рис. 9: Прогноз методом АКМ и методом экстраполяции с линейной коррекцией прогностического отрезка (Манхеттенская метрика)

## Заключение

Тестирование построенных моделей на независимой выборке температуры воздуха по данным автоматической метеорологической станции позволило выделить оптимальную метрику — Манхеттенскую. При заблаговременности прогноза до 4 часов используемая метрика дает ошибку прогноза не более 0,4 °С.

Полученные результаты позволяют рекомендовать использование метода аналогов для текущего прогноза нестационарных временных рядов температуры воздуха.

## Список литературы

- [1] Восканян К.Л., Дивинский Л.И., Кузнецов А.Д., Сероухова О.С., Симакина Т.Е. Применение сплайн-экстраполяции для текущего прогнозирования приземной температуры воздуха // Труды Военно-космической академии им. А.Ф. Можайского. 2015. № 647. С. 88–92.
- [2] Алдошкина Е.С., Кузнецов А.Д., Пугачев А.А., Сероухова О.С., Симакина Т.Е., Чукин В.В. Опыт использования аппарата нейронных сетей для анализа и прогноза временного ряда температуры воздуха // Ученые записки Российского государственного гидрометеорологического университета. 2009. № 11. С. 91–100.
- [3] Васечкина Е.Ф., Ярин В.Д. Прогноз временных рядов с применением метода аналогов // Морской гидрофизический журнал. 2007. № 4. С. 70–80. <https://doi.org/10.1007/s11110-007-0019-3>
- [4] Восканян К.Л., Кузнецов А.Д., Сероухова О.С., Симакина Т.Е. Текущее прогнозирование экологических измерений на основе поиска аналогов // Сборник тезисов XI научно-прикладной международной конференции «Естественные и антропогенные аэрозоли» (СПб., 16-18 октября 2018 г.).
- [5] Чучуева И. Модель прогнозирования временных рядов по выборке максимального подобия: пояснение и пример [Электронный ресурс] // Хабр. URL: <https://habr.com/ru/post/267035/> (дата обращения: 12.12.2018).
- [6] Aldenderfer M.S., Blashfield R.K. Cluster Analysis. Series: Quantitative Applications in the Social Sciences. Vol. 44. SAGE Publications, Inc, 1984. 88 p.
- [7] Сокэл Р.Р. Кластер-анализ и классификация: предпосылки и основные направления // Классификация и кластер. Под ред. Дж. Вэн Райзин. М.: Мир, 1980. С. 7–19.
- [8] Hamer R., Cunningham J. Cluster analyzing profile data confounded with interrater differences: A comparison of profile association measures // Applied Psychological Measurement. 1981. Vol. 5. Pp. 63–72. <https://doi.org/10.1177/014662168100500109>
- [9] Синельник С.А., Коврижных О.Е. Определение меры сходства объектов в кластерном анализе [Электронный ресурс] // Бизнес концепт. URL: <http://www.konspekt.biz/index.php?text=51064> (дата обращения: 01.11.2018).

- [10] Магнус Я.Р., Катышев П.К., Пересецкий А.А. Метод наименьших квадратов (МНК) // Эконометрика. Начальный курс: учебное пособие. М.: Издательство «Дело», 2004. С. 34–37.
- [11] Восканян К.Л., Кузнецов А.Д., Сероухова О.С. Автоматические метеорологические станции. Часть 1. Тактико-технические характеристики: учебное пособие. СПб.: РГГМУ, 2016. 195 с.
- [12] Восканян К.Л., Кузнецов А.Д., Сероухова О.С. Автоматические метеорологические станции. Часть 2. Цифровая обработка данных автоматических метеорологических станций: практикум. СПб.: РГГМУ, 2016. 99 с.

### Образец цитирования

Восканян К.Л., Кузнецов А.Д., Сероухова О.С., Симакина Т.Е. Оценка возможностей метода аналогов для текущего прогноза температуры воздуха // Вестник ТвГУ. Серия: Прикладная математика. 2019. № 1. С. 101–116. <https://doi.org/10.26456/vtpmk528>

### Сведения об авторах

- 1. Восканян Карина Левановна**  
доцент кафедры экспериментальной физики атмосферы Российского государственного гидрометеорологического университета.  
*Россия, 195196, г. Санкт-Петербург, Малоохтинский пр., д. 98, РГГМУ.*  
*E-mail: [kvosia@mail.ru](mailto:kvosia@mail.ru)*
- 2. Кузнецов Анатолий Дмитриевич**  
заведующий кафедрой экспериментальной физики атмосферы Российского государственного гидрометеорологического университета.  
*Россия, 195196, г. Санкт-Петербург, Малоохтинский пр., д. 98, РГГМУ.*  
*E-mail: [kuznetsov1946@inbox.ru](mailto:kuznetsov1946@inbox.ru)*
- 3. Сероухова Ольга Станиславовна**  
доцент кафедры экспериментальной физики атмосферы Российского государственного гидрометеорологического университета.  
*Россия, 195196, г. Санкт-Петербург, Малоохтинский пр., д. 98, РГГМУ.*  
*E-mail: [serouhova@inbox.ru](mailto:serouhova@inbox.ru)*
- 4. Симакина Татьяна Евгеньевна**  
доцент кафедры экспериментальной физики атмосферы Российского государственного гидрометеорологического университета.  
*Россия, 195196, г. Санкт-Петербург, Малоохтинский пр., д. 98, РГГМУ.*  
*E-mail: [tatiana.simakina@gmail.com](mailto:tatiana.simakina@gmail.com)*

**ASSESSMENT OF THE POSSIBILITIES OF THE METHOD  
OF ANALOGUES FOR THE CURRENT FORECAST  
OF AIR TEMPERATURE**

**Voskanan Karina Livanovna**

Associate Professor of the Department of Experimental Atmospheric Physics,  
Russian State Hydrometeorological University  
*Russia, 195196, St. Petersburg, 98 Malookhtinsky ave, RSHU.*  
*E-mail: [kvosia@mail.ru](mailto:kvosia@mail.ru)*

**Kuznetcov Anatolii Dmitriyevich**

Head of the Department of Experimental Atmospheric Physics,  
Russian State Hydrometeorological University  
*Russia, 195196, St. Petersburg, 98 Malookhtinsky ave, RSHU.*  
*E-mail: [kuznetsov1946@inbox.ru](mailto:kuznetsov1946@inbox.ru)*

**Seroukhova Olga Stanislavovna**

Associate Professor at the Department of Experimental Atmospheric Physics,  
Russian State Hydrometeorological University  
*Russia, 195196, St. Petersburg, 98 Malookhtinsky ave, RSHU.*  
*E-mail: [seroukhova@inbox.ru](mailto:seroukhova@inbox.ru)*

**Simakina Tatiana Evgenyevna**

Associate Professor at the Department of Experimental Atmospheric Physics,  
Russian State Hydrometeorological University  
*Russia, 195196, St. Petersburg, 98 Malookhtinsky ave, RSHU.*  
*E-mail: [tatiana.simakina@gmail.com](mailto:tatiana.simakina@gmail.com)*

---

*Received 05.03.2019, revised 23.03.2019.*

---

An algorithm for forecasting non-stationary time series based on the analog method is considered. The proposed method is tested at air temperature values according to an automatic meteorological station. Six forecast models using various measures of similarity were constructed, the forecast error was determined on an independent sample. Comparison with the estimate of the inertial prediction led to the conclusion that the method of analogs gives satisfactory results of the current temperature prediction.

**Keywords:** analogue method, current temperature forecast, proximity measures.

#### **Citation**

Voskanan K.L., Kuznetcov A.D., Seroukhova O.S., Simakina T.Ye., "Assessment of the possibilities of the method of analogues for the current forecast of air temperature", *Vestnik TvGU. Seriya: Prikladnaya Matematika [Herald of Tver State University. Series: Applied Mathematics]*, 2019, № 1, 101–116 (in Russian). <https://doi.org/10.26456/vtpmk528>

### References

- [1] Voskanyan K.L., Divinskij L.I., Kuznetsov A.D., Seroukhova O.S., Simakina T.E., “The use of spline extrapolation for the current prediction of surface air temperature”, *Proceedings of the Mozhaisky Military Space Academy*, 2015, № 647, 88–92 (in Russian).
- [2] Aldoshkina E.S., Kuznetsov A.D., Pugachev A.A., Seroukhova O.S., Simakina T.E., Chukin V.V., “Experience of using the neural network apparatus for analyzing and forecasting the air temperature time series”, *Scientific notes of the Russian State Hydrometeorological University*, 2009, № 11, 91–100 (in Russian).
- [3] Vasechkina E.F., Yarin V.D., “Prediction of time series by the method of analogs”, *Physical Oceanography*, **17**:4 (2007), 242–251, <https://doi.org/10.1007/s11110-007-0019-3>.
- [4] Voskanyan K.L., Kuznetsov A.D., Seroukhova O.S., Simakina T.E., “Current forecasting of environmental measurements based on the search for analogues”, *Abstracts of the XI Scientific and Applied International Conference “Natural and anthropogenic aerosols”* (St. Petersburg, October 16–18, 2018) (in Russian).
- [5] Chuchueva I., *The time series forecasting model for the maximum similarity sample: an explanation and an example*, Habr (in Russian), <https://habr.com/ru/post/267035/> (accessed at 12.12.2018).
- [6] Aldenderfer M.S., Blashfield R.K., *Cluster Analysis*. V. 44, Quantitative Applications in the Social Sciences, SAGE Publications, Inc, 1984, 88 pp.
- [7] Sokel R.R., “Cluster analysis and classification: prerequisites and main directions”, *Klassifikatsiya i klaster [Classification and Cluster]*, ed. Dzh. Ven Rajzin, Mir Publ., Moscow, 1980, 7–19 (in Russian).
- [8] Hamer R., Cunningham J., “Cluster analyzing profile data confounded with interrater differences: A comparison of profile association measures”, *Applied Psychological Measurement*, **5** (1981), 63–72, <https://doi.org/10.1177/014662168100500109>.
- [9] Sinelnik S.A., Kovrizhnykh O.E., *Determining the measure of similarity of objects in cluster analysis*, konspekt.biz (in Russian), <http://www.konspekt.biz/index.php?text=51064> (accessed at 01.11.2018).
- [10] Magnus Ya.R., Katyshev P.K., Peresetskij A.A., “Least Squares Method”, *Ekonometrika. Nachalnyj kurs: uchebnoe posobie [Econometrics. Initial course: tutorial]*, Delo Publ., M., 2004, 34–37 (in Russian), 576 pp.
- [11] Voskanyan K.L., Kuznetsov A.D., Seroukhova O.S., *Avtomaticheskie meteorologicheskie stantsii. Chast 1. Taktiko-tekhnicheskie kharakteristiki: uchebnoe posobie [Automatic weather stations. Part 1. Performance characteristics: tutorial]*, RSHU, SPb., 2016 (in Russian), 195 pp.
- [12] Voskanyan K.L., Kuznetsov A.D., Seroukhova O.S., *Automatic weather stations. Part 2. Digital data processing of automatic meteorological stations: workshop*, RSHU, SPb., 2016 (in Russian), 99 pp.