

НЕПАРАМЕТРИЧЕСКАЯ ОЦЕНКА ПРОГНОСТИЧЕСКОЙ ТОЧНОСТИ МОДЕЛЕЙ НА ПРИМЕРЕ ВРЕМЕННЫХ РЯДОВ КОНЦЕНТРАЦИЙ МЕТАНА В АТМОСФЕРНОМ ВОЗДУХЕ АРКТИЧЕСКОГО ОСТРОВА БЕЛЫЙ

Е. М. Баглаева, А. П. Сергеев, А. Г. Буевич, А. В. Шичкин, И. Е. Субботина

Институт промышленной экологии Уральского отделения РАН (Екатеринбург, Российская Федерация)

Для цитирования

Баглаева Е. М., Сергеев А. П., Буевич А. Г. и др. Непараметрическая оценка прогностической точности моделей на примере временных рядов концентраций метана в атмосферном воздухе арктического острова Белый // Арктика: экология и экономика. — 2024. — Т. 14, № 4. — С. 500—510. — DOI: 10.25283/2223-4594-2024-4-500-510.

Статья поступила в редакцию 17 июня 2024 г.

Анализ неопределенности прогнозов изменения выбросов парниковых газов приобретает серьезное значение в связи с катастрофическими последствиями изменения климата. Оценка точности моделей прогнозов содержания парниковых газов в атмосфере требует особого внимания и специальных подходов. В данной работе предложено использовать идеологию проверки статистических гипотез для оценки производительности предсказательной модели временного ряда. Данные для исследования получены при проведении мониторинга динамики приземной концентрации основных парниковых газов на арктическом острове Белый (Ямало-Ненецкий автономный округ, Россия). Три модели на основе авторегрессионных нейронных сетей с внешним входом NARX предсказывали изменение концентрации метана. Показано, что размеры эффекта не всегда определяют статистическую значимость различий.

Ключевые слова: парниковые газы, мониторинг, изменения климата, Арктическая зона Российской Федерации.

Введение

С начала индустриальной эпохи в 1750 г. наблюдаются изменения климата, связанные с потеплением, повышением температуры и учащением случаев экстремальной жары. В отчетах Межправительственной группы экспертов по изменению климата (МГЭИК) наиболее вероятной причиной потепления было названо увеличение содержания в атмосфере парниковых газов. Большинство парниковых газов, таких как водяной пар H_2O , углекислый газ CO_2 и метан CH_4 , имеют как природные, так и антропогенные источники.

В ближайшей перспективе (2021—2040 гг.) согласно отчету IPCC [2] глобальное потепление по всем сценариям и моделям продолжит усиливаться из-за роста совокупных выбросов парниковых газов. В ближайшей перспективе глобальная температура увеличится на $1,5^\circ C$ даже при сценарии

с очень низким уровнем выбросов парниковых газов, а при сценариях с более высокими выбросами ее рост превысит $1,5^\circ C$. Потепление в Арктике приводит к увеличению выбросов парниковых газов в атмосферу и, как следствие, дальнейшему повышению температуры, реальные масштабы которого до конца неизвестны [3; 4]. Для прогноза используются модели интерполяции на основе данных мониторинга парниковых газов [5; 6]. При таких катастрофических прогнозах приобретает первостепенное значение оценка эффективности используемых при расчетах моделей, их верификация и проверка точности расчетов. В связи с этим представляется важной разработка подходов к оценке точности моделей [7—9].

Географические (пространственные) модели интерполяции получили широкое распространение по мере развития вычислительной техники, и в настоящее время доступно большое разнообразие географических интерполяторов [10—12]. Числовые модели экологических систем выросли в количестве, разнообразии и сложности за последние несколько

десятилетий. Соответственно возрос интерес исследователей к сравнению и оценке прогностической точности моделей, в частности, чтобы определить, какие модели более точны для конкретной области исследования [7; 13—15].

Однако исследователи могут столкнуться с трудностями при оценке точности используемой модели. Ошибки могут быть связаны среди прочего с используемыми методологиями для такой оценки. Начальным шагом при проверке моделей природных явлений является вывод о том, похоже ли их поведение на наблюдаемое. График, показывающий, что некоторая закономерность наблюдаемой вариации достаточно хорошо воспроизводится моделью, представляется как свидетельство ее достоверности. Для моделей с множеством переменных и несколькими измерениями [16—18] визуальное сравнение смоделированных и наблюдаемых полей становится непрактичным, даже если учитывается только небольшая часть выходных данных модели. Исследователю необходимо либо сосредоточиться на каком-то ограниченном аспекте описываемой физической системы, либо использовать индексы для количественного определения общего несоответствия между моделируемым и наблюдаемым. Основными индексами для количественной оценки пространственных интерполяторов являются среднеквадратическая ошибка RMSE и средняя абсолютная ошибка MAE. Каждый из них является размерным в том смысле, что они выражают среднюю ошибку интерполятора в единицах измерения интересующей нас переменной. MAE — наиболее фундаментальный показатель общей производительности интерполятора, и использование несоответствующих ее показателей может привести к неправильной интерпретации эффективности модели [7; 19]. Вычисление MAE не представляет сложности. Взвешенные величины (абсолютные значения) интерполированных ошибок в узлах сетки суммируются для получения общей ошибки, которая затем делится на сумму весов ячеек сетки. Вычисление RMSE не так просто. RMSE получается сначала как сумма взвешенных индивидуальных квадратов ошибок узлов сетки, т. е. каждая ошибка узла сетки влияет на общую сумму пропорционально ее квадрату, а не величине. Как следствие, большие ошибки оказывают относительно большее влияние на общую квадратичную ошибку, чем меньшие ошибки. Затем общая квадратичная ошибка делится на сумму весов площадей ячеек сетки, что дает среднеквадратичную ошибку MSE. Последний шаг — получение RMSE как квадратного корня из MSE. Поскольку деление на сумму весов площадей ячеек сетки и квадратного корня масштабирует только общую квадратичную ошибку, очевидно, что MSE и RMSE будут увеличиваться вместе с общей квадратичной ошибкой по мере увеличения дисперсии, связанной с частотным распределением ошибок. Для более достоверной оценки точности модели было предложено несколько версий безразмерных индексов [19;

20]. Кроме того, существует еще целый ряд подобных индикаторов, которые используются для оценки моделей. Эти индексы представляют собой другой подход к оцениванию предсказательной точности моделей и достаточно гибки, что позволяет применять их к широкому диапазону приложений для работы с моделями.

В настоящей работе предлагается новый нетривиальный подход, позволяющий взглянуть на оценку эффективности модели с другой стороны. Он дает возможность оценить точность моделей не «количественно», а «качественно», на примере прогнозирования временного ряда изменения концентраций парниковых газов. Прогнозы были получены с помощью моделей, основанных на машинном обучении, которые показали высокую точность во многих исследованиях [5; 12].

Материалы и методы

Данные исследования

Данные для моделирования были получены в результате мониторинга динамики концентрации основных парниковых газов на территории арктического острова Белый, Ямало-Ненецкий автономный округ (рис. 1). Остров находится в 9 км от континентальной части Евразии в Карском море. Он представляет собой относительно ровную местность, покрытую тундровой растительностью, и находится далеко от основных промышленных объектов. Климат морской арктический с постоянными сильными ветрами преимущественно западных направлений [21].

Данные о приземной концентрации парниковых газов получены при помощи Picarro G2401 — анализатора углекислого газа, метана, угарного газа (монооксида углерода), паров воды. Измерения проводились в течение июля-августа 2017 г. каждые 5 с. Затем данные были усреднены и получены почасовые временные ряды. Измерения метеорологических параметров (температуры, влажности и давления) метеостанцией Vaisala AWS310 проводились каждый час в тот же временной период. Часовые временные ряды метеопараметров и содержания метана были синхронизированы. Для моделирования использовался временной интервал 1175 точек. Первые 1103 временных точки были использованы для обучения модели, для следующих 72 точек предсказывали значение концентрации метана. Для оценки точности модели использовали классические оценки точности, построенные на расчете разницы между прогнозируемыми и наблюдаемыми данными, и перестановочный подход, основанный на тестировании статистических гипотез.

Классические способы оценки точности модели

Традиционно для оценки предсказательной способности модели используют коэффициент корреляции $Corr$ между предсказанными и наблюдаемыми значениями и коэффициент детерминации R^2 . Коэффициент детерминации показывает процент объясненной моделью дисперсии.

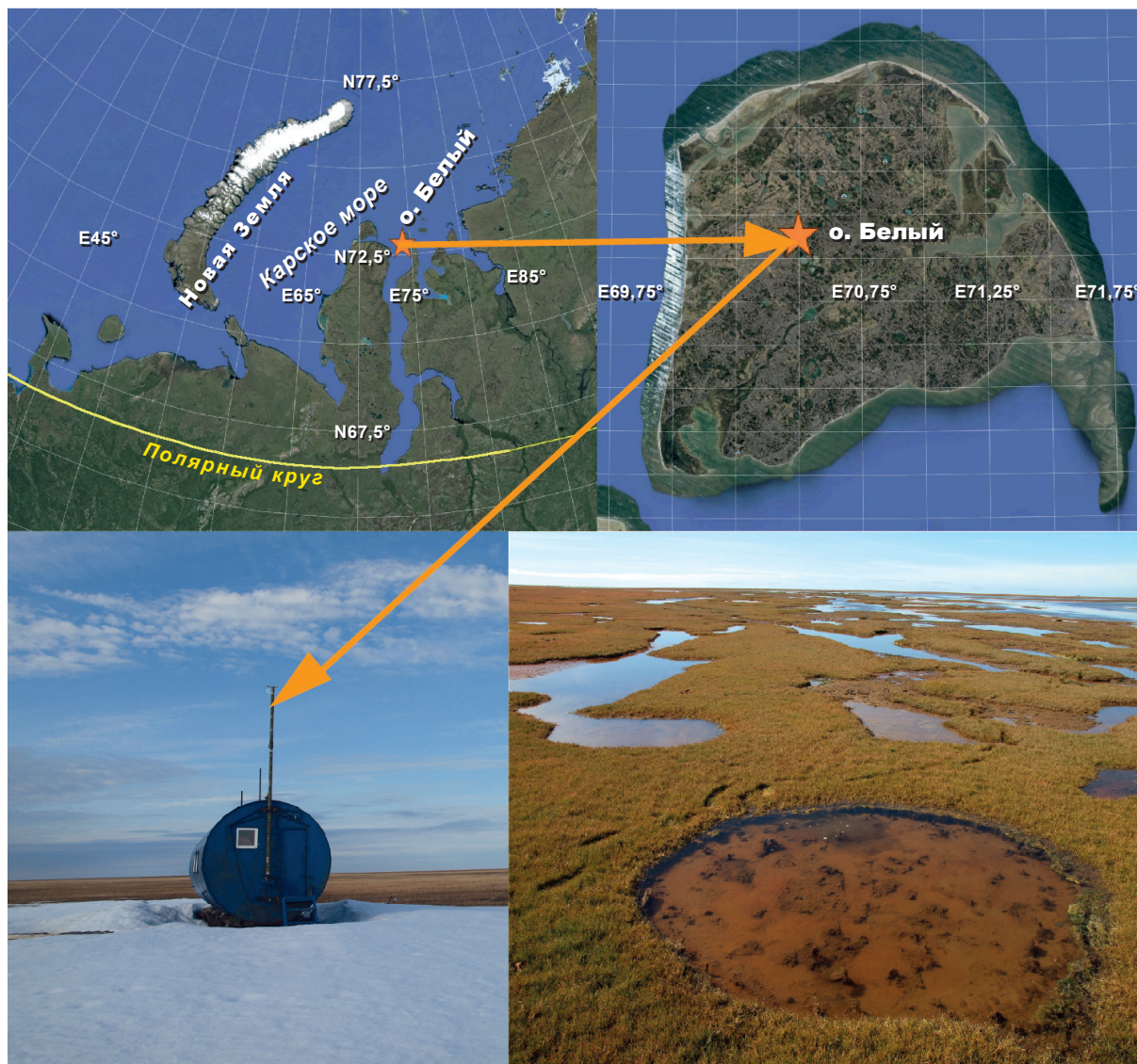


Рис. 1. Место отбора проб парниковых газов на острове Белый. Верхний ряд – космоснимки Google Earth. Нижний ряд – фотографии А. Г. Бувевича

Fig. 1. Greenhouse gas sampling site on Bely Island. Upper line – Google Earth satellite images. Lower line – photos by A. G. Buevich

Для оценки точности модели рассчитывают ошибки между предсказанными и наблюдаемыми значениями, или индексы согласия. Наиболее часто используются средняя абсолютная ошибка MAE, среднеквадратическая ошибка RMSE, относительная среднеквадратическая ошибка RRMSE. Формулы для их вычисления:

$$MAE = \frac{\sum_{i=1}^m |P(i) - O(i)|}{m}, \quad (1)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^m (P(i) - O(i))^2}{m}}, \quad (2)$$

$$RRMSE = \frac{\sqrt{\frac{1}{m} \sum_{i=1}^m (P(i) - O(i))^2}}{\frac{1}{m} \sum_{i=1}^m O(i)}, \quad (3)$$

где $O(i)$ — наблюдаемое значение временного ряда в i -й момент времени; $P(i)$ — предсказанное значение временного ряда в i -й момент времени; m — длина временного ряда.

Мы рассчитывали четыре индекса согласия: индекс Уилмотта IA, индекс остаточного отклонения прогноза RPD, индекс отношения эффективности к межквартильному расстоянию RPIQ, критерий эффективности Клинг — Гупта KGE.

$$IA = 1 - \frac{\sum_{i=1}^m (P(i) - O(i))^2}{\sum_{i=1}^m (|P(i) - \bar{O}| + |O(i) - \bar{O}|)^2}, \quad (4)$$

$$RPD = \frac{SD_{obs}}{RMSE}, \quad (5)$$

где SD_{obs} — стандартное отклонение наблюдаемого; $\bar{O} = \sum_{i=1}^m O(i) / m$ — среднее значение наблюдаемого временного ряда; $\bar{P} = \sum_{i=1}^m P(i) / m$ — среднее значение предсказанного временного ряда.

Если значение $RPD < 1,4$, то модель непригодна, при $1,4 < RPD < 2,0$ она хорошая, выше 2,0 — великолепная.

$$RPIQ = \frac{IQ}{RMSE}, \quad (6)$$

где $IQ = Q3 - Q1$ — межквартильное расстояние, соответствующее диапазону, на который приходится 50% всей выборки; $Q1$ — значение, ниже которого находится 25% точек; $Q3$ — значение, выше которого находится 25% точек.

$$KGE = 1 - \sqrt{\left(Corr - 1 \right)^2 + \left(\frac{SD_{pred}}{SD_{obs}} - 1 \right)^2 + \left(\frac{\bar{O}}{\bar{P}} - 1 \right)^2}. \quad (7)$$

Значение $KGE = 0$ считают порогом между хорошей и плохой моделями. Некоторые исследователи [22] дают более подробные границы моделей: $0,7 < KGE < 1,0$ — очень хорошая; $0,6 < KGE < 0,7$ — хорошая; $0,5 < KGE < 0,6$ — удовлетворительная; $0,4 < KGE < 0,5$ — допустимая; $KGE < 0,4$ — неудовлетворительная.

Модель NARX

Три модели на основе авторегрессионных нейронных сетей с внешним входом (NARX) были выбраны в предыдущих исследованиях для предсказания значений временных рядов [5]. Для построения моделей был использован пакет прикладных программ MATLAB 2021b. Архитектура сети NARX с функцией активации гиперболического тангенса представлена на рис. 2.

Модель NARX представлена уравнением

$$y(t) = f \left(\begin{matrix} y(t-1), y(t-2), \dots, y(t-n_y), \\ u(t-1), u(t-2), \dots, u(t-n_u) \end{matrix} \right) + \varepsilon, \quad (8)$$

где каждое последующее значение выходного вектора $y(t)$ зависит от предыдущих значений выходного и входного векторов; n_u — число входных задержек; n_y — число выходных задержек; ε — случайная ошибка.

Две другие из трех предложенных моделей использовали дискретное вейвлет-преобразование, которое разлагает временной ряд на основу, состоящую из семейства функций, образованных сдвигом и растяжением (сжатием) материнского вейвлета. Вейвлет-преобразование способно выявить различия в характеристиках временного ряда на разных масштабах, а путем сдвига может анализиро-

вать свойства временного ряда в разных точках на всем исследуемом интервале. Здесь использованы вейвлеты четвертого уровня db4 и sum4, принадлежащие семейству Добеши. Вейвлеты Добеши ортогональны, асимметричны и имеют компактный носитель. Sum4 — вейвлет Добеши с наименьшей асимметрией и компактным носителем. В пакете MATLAB эти вейвлеты задаются набором весовых коэффициентов — см. уравнение (9).

$$S(t) = A_J(t) + \sum_{m=1}^J D_m(t), \quad (9)$$

где $S(t)$ — исходный временной ряд; $A_J(t)$ и $D_m(t)$ ($m = 1, \dots, 4$) — аппроксимирующие и детальные компоненты.

Модели предсказывали изменение концентрации метана CH_4 . Для вейвлет-разложения временные ряды CH_4 масштабированы от 0 до 1. Первые 1103 точки времени были использованы для обучения модели, затем был получен прогноз для следующих 72 точек интервала (точки данных трехдневного прогнозирования). Метеорологические временные ряды: температура, влажность и давление были входными данными, а концентрации CH_4 были выходными данными NARX (см. рис. 2).

Логика перестановочного подхода

Классические способы оценки точности модели основаны на оценке размера эффекта: на расчете ошибок между прогнозируемыми и наблюдаемыми данными или на расчете коэффициентов согласия [7]. Часто полученные оценки являются противоречивыми, поскольку сложно сравнивать разные модели, их ошибки и коэффициенты согласия между собой. Индексы точности говорят о величине «эффекта» как о расстоянии между предсказанием и наблюдением. Чем выше ошибка (MAE, RMSE, RRMSE) и меньше индекс согласия (IA, KGE, RPD, RPIQ), тем хуже прогноз.

Математико-статистические методы оценки точности, основанные на тестировании статистических гипотез, позволяют измерить не только размер эффекта, но и его значимость посредством вычисления р-значения.

В настоящей работе мы предлагаем для оценки точности модели перестановочный подход, использующий расчет двух статистик: разности средних и корреляции.

Перестановочный подход предполагает перестановку всех значений наблюдаемого и прогнозируемого временных рядов. В предположении, что наблюдаемый и предсказанный ряды имеют одинаковую природу, наблюдаемый и предсказанный ряды неразличимы, получены по одной и той же системе, т. е. статистические характеристики наблюдаемого и предсказанного рядов дают одни и те же значения (нулевая гипотеза). Если всеми возможными способами поменять значения предсказанного и наблюдаемого ряда, для ряда из m значений

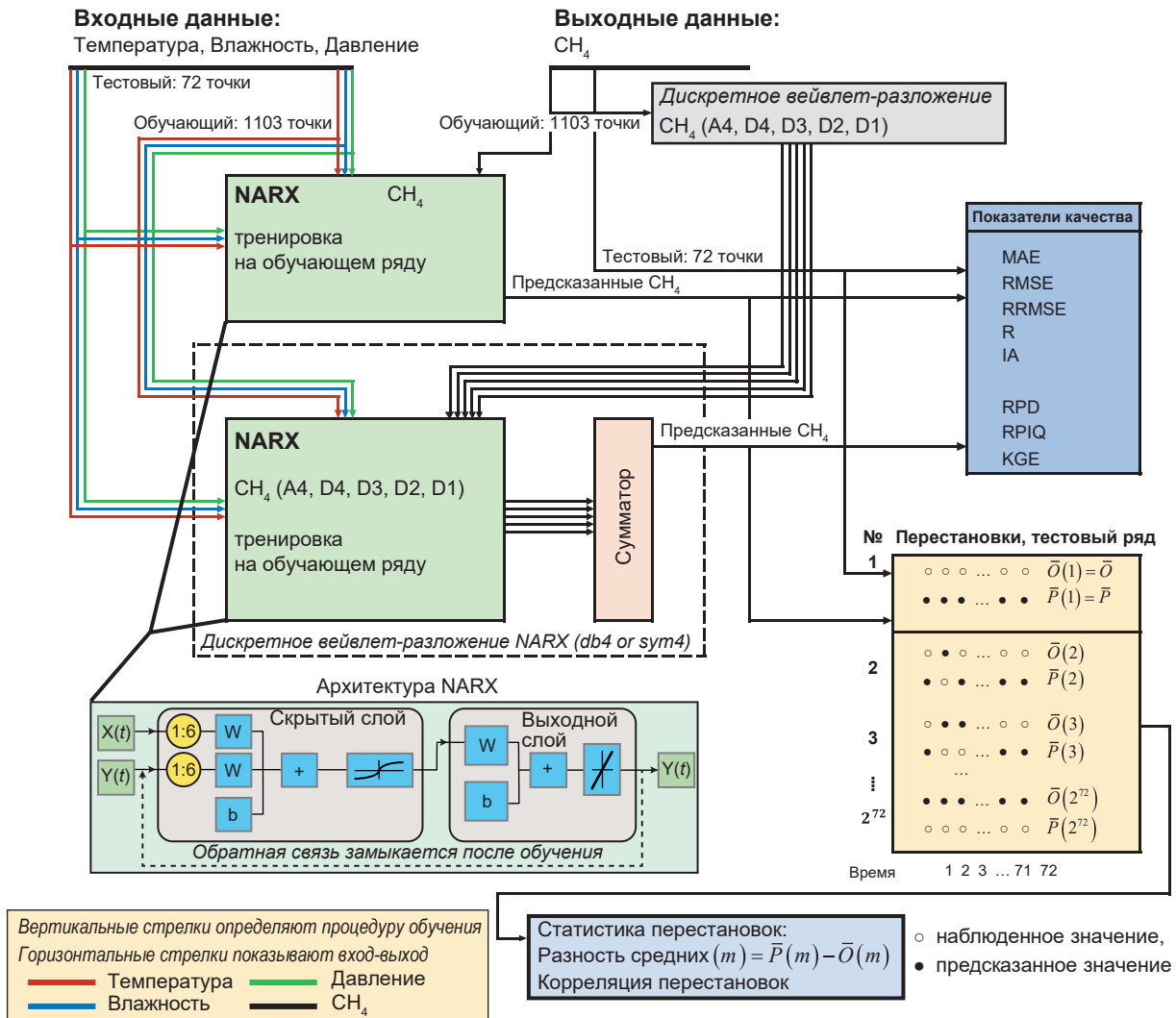


Рис. 2. Схема перестановочного подхода для трехсуточного прогноза метана
Fig. 2. Scheme of the permutation approach for three-day methane forecast

можно получить 2^m вариантов и построить точное распределение выбранных статистик. Наш прогнозируемый ряд содержит 72 значения, что дает 272 всевозможные перестановки. Для выполнения такого количества перестановок в настоящее время недостаточно ни вычислительных мощностей компьютеров, ни временных затрат. В настоящей работе было выбрано случайным образом 100 перестановочных временных рядов и для каждого из них вычислены две статистики: разность средних и корреляция.

Для оценки значимости различий выбранных статистик были рассчитаны р-значения. Под р-значением в настоящей работе понимали вероятность получить такое же или более экстремальное значение статистик при условии, что нулевая гипотеза верна. Перестановочное р-значение вычисляли для полученных перестановками распределений разности средних и корреляции. В случае с корреляцией измеряли правостороннее р-значение, а для

разности средних р-значение определяли как наименьшее из односторонних (слева и справа). Если вычисленное р-значение для распределения разности средних было высоким, нулевая гипотеза не отклонялась, делали вывод о неразличимости предсказанного и наблюдаемого наборов, что свидетельствовало о хорошем качестве модели.

В случае корреляции выдвигалась нулевая гипотеза об отсутствии корреляции между наблюдаемым и предсказанным наборами. Если было обнаружено низкое р-значение, то нулевая гипотеза отвергалась, делали вывод о наличии корреляции между наблюдаемыми и предсказанными значениями, что свидетельствует о высоком качестве модели.

Результаты и обсуждение

Временные ряды температуры, влажности, давления и содержания метана представлены на рис. 3. Статистические характеристики обучающего и те-

стового наборов наблюдаемых данных приведены в табл. 1. Значения температуры и содержания метана в тестовом и обучающем наборах отличаются слабой изменчивостью, о чем говорят низкие значения коэффициентов вариации. Исходные данные демонстрируют слабые отрицательные тренды для метана, температуры и давления и слабый положительный тренд для влажности.

Данные, используемые для моделирования, были измерены в летний сезон и не содержат компонент, связанных со сменой времен года. В нашей ранней работе при анализе временных рядов парниковых газов, в том числе метана, были обнаружены декадные, недельные, трехсуточные и суточные периодические компоненты в структуре временного ряда метана. На рис. 4 приведены оценки автокорреляционных функций исследуемых временных рядов и вейвлет-скалограмма метана.

Анализ оценок автокорреляционных функций показывает ярко выраженный 24-часовой период для метана, температуры и влажности и его отсутствие

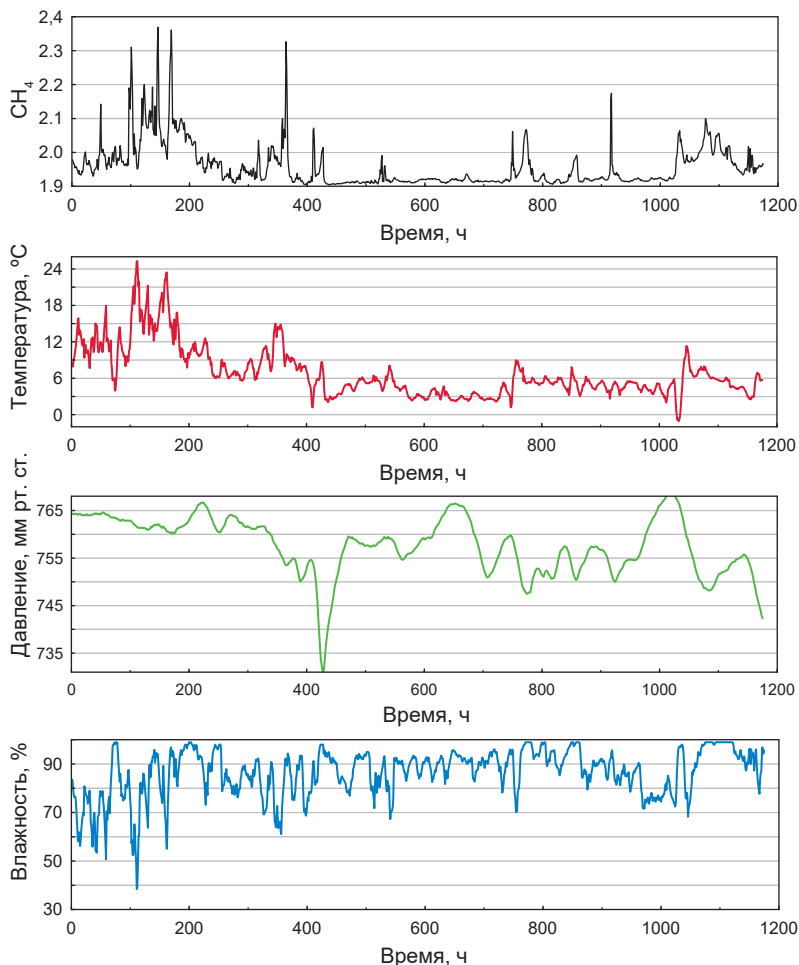


Рис. 3. Наблюдаемые временные ряды
Fig. 3. Observed time series

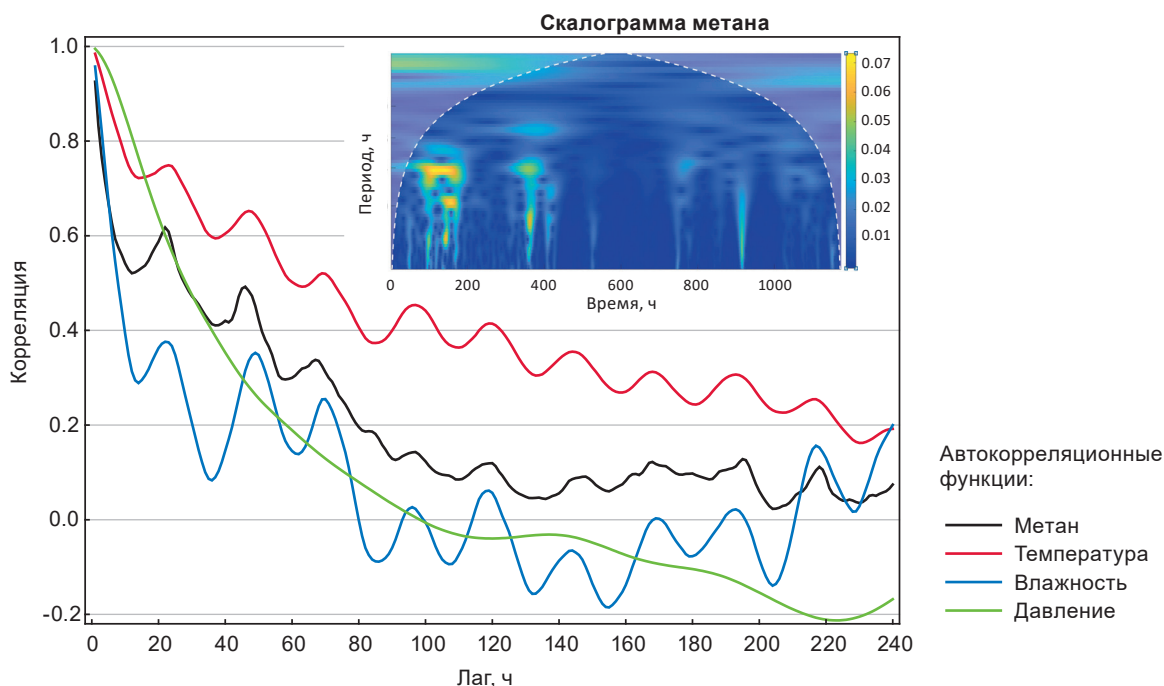


Рис. 4. Оценки автокорреляционных функций наблюдаемых временных рядов
Fig. 4. Estimates of autocorrelation functions of observed time series

Таблица 1. Статистические характеристики наблюдаемых временных рядов
Table 1. Statistical characteristics of the observed time series

Набор данных	Параметр	Минимум	Максимум	Среднее	Медиана	SD	Коэффициент вариации, %	Асимметрия	Экссесс
Обучающий (1103 точки)	Температура, °K	272,06	298,48	280,10	278,80	4,22	1,51	1,43	2,22
	Влажность, %	38	99	86	88	10	11	-1,09	1,54
	Давление, мм рт. ст.	731	769	758	759	6	0,79	-0,98	2,03
	CH ₄ , ppm	1,905	2,370	1,957	1,925	0,066	3,35	2,42	7,79
Тестовый (72 точки)	Температура, °K	275,08	275,17	275,12	275,11	0,02	0,01	0,70	-0,67
	Влажность, %	77	99	93	94	5	5	-1,04	1,07
	Давление, мм рт. ст.	742	756	752	753	5	5	-1,33	2,03
	CH ₄ , ppm	1,933	2,020	1,968	1,961	0,025	1,26	0,70	-0,76

Таблица 2. Оценки точности моделей предсказания
Table 2. Forecast model accuracy assessments

Модель	R ²	IA	MAE, ppm	RMSE	RRMSE	RPD	RPIQ	KGE
NARX	0,4633	0,7746	0,0119	0,0180	0,9169	0,8029	1,0577	0,4948
NARX (db4)	0,5709	0,8863	0,0124	0,0161	0,8199	1,5160	2,4313	0,7963
NARX (sym4)	0,6616	0,8812	0,0109	0,0143	0,7281	1,2718	1,7872	0,6866

Таблица 3. Статистические характеристики предсказанного ряда
Table 3. Statistical characteristics of the predicted series

Модель	Предсказанное среднее, ppm	SD, ppm	Разность средних, ppm	p-значение разности средних	Corr	p-значение корреляции
NARX	1,9636	0,0145	0,0042	0,0234	0,7133	0,0000
NARX (db4)	1,9719	0,0245	-0,0041	0,0145	0,7967	0,0038
NARX (sym4)	1,9715	0,0182	-0,0037	0,0133	0,8332	0,0000

для давления. Спад автокорреляционной функции временного ряда метана показывает влияние 72 значений ряда.

Вейвлет-скалограмма метана имеет ярко выраженный суточный период в диапазоне 90—180 ч.

Вейвлет-преобразование наблюдаемого обучающего набора содержания CH₄ на аппроксимирующие и детализирующие компоненты представлено в нашей ранней работе [5]. В табл. 2 показаны традиционные результаты оценки точности модели NARX и гибридных моделей NARX (db4) и NARX (sym4). Лучшие показатели выделены полужирным шрифтом. Как видно из табл. 2, по показателям ошибок лучшей оказывается модель NARX (sym4), но по индексам согласия лучшая модель — NARX (db4).

Традиционные оценки точности модели были дополнены методом перестановок. В табл. 3 представлены статистические характеристики предсказанного ряда. Для построения перестановочных распределений были выбраны статистика разности средних и корреляционная статистика. Метод перестановки возвращал p-значение — это вероятность получить наблюдаемое или более экстремальное значение используемой статистики (корреляции или разности средних), предполагая справедливость гипотезы о том, что как наблюдаемые, так и прогнозируемые временные ряды являются реализациями одного и того же случайного процесса.

Статистика разности средних отвечает на вопрос о средней выраженности признака в тестовой выборке без учета изменчивости прогнозируемых и на-

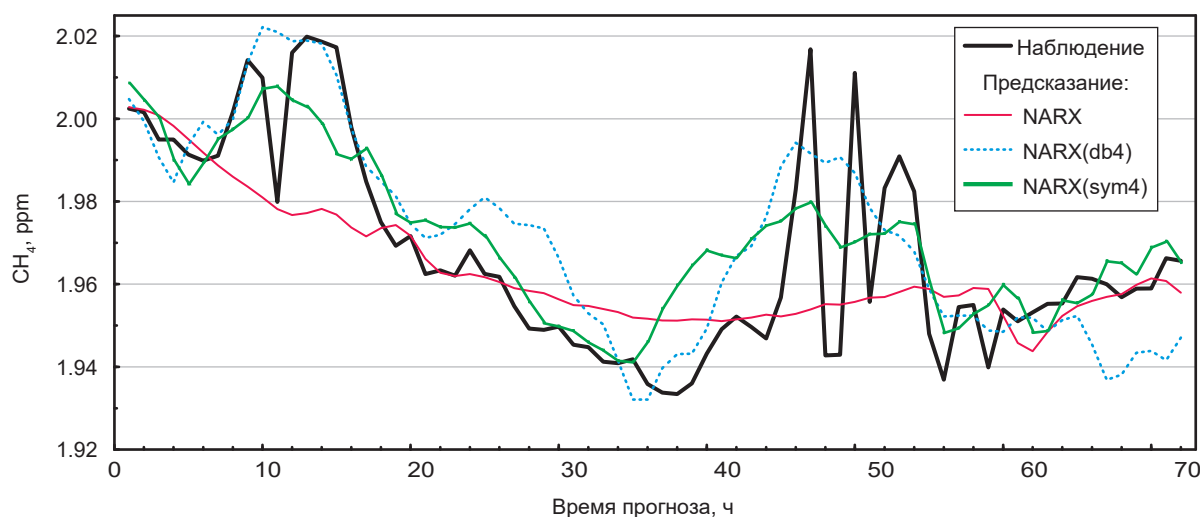


Рис. 5. Прогноз содержания CH_4 на 72 ч
Fig. 5. CH_4 forecast for 72 hours

блюдаемых значений временных рядов. Маленькая разность средних прогнозируемых и наблюдаемых данных показывает хорошую модель, а р-значение сравнивает изменчивость прогнозируемого и наблюдаемого наборов данных. Высокое р-значение для распределения разности средних означает хорошую модель.

Корреляционная статистика учитывала меру синхронности предсказанных и наблюдаемых временных рядов. Для корреляционной статистики модель лучше, когда р-значение меньше. Лучшие значения статистических показателей в табл. 3 выделены полужирным шрифтом.

Среднее значение обучающего набора составляет 1,9571 ppm, тестового набора — 1,9677 ppm, наиболее близко к ним предсказанное моделью NARX (sym4) среднее 1,9715 ppm. Стандартное отклонение обучающего набора ($SD_{obs} \text{ training} = 0,0655$) отличается от стандартного отклонения тестового набора ($SD_{obs} \text{ test} = 0,0248$), при этом наименьшее стандартное отклонение предсказанных значений — 0,0145 для модели NARX (см. табл. 3). Модель NARX (sym4) лучше всего предсказывает среднее значение, но разброс предсказанных значений вокруг среднего для этой модели больше, чем для NARX.

Предлагаемые гибридные модели NARX (sym4) имеют более низкие значения RMSE, RMSRE, MAE, а также более высокие коэффициенты корреляции Пирсона *Corr*, коэффициент детерминации R^2 .

Если сравнивать модели не только по размеру эффекта (ошибкам или индексам согласия), но и по значимости, по статистикам корреляции и разности средних, наилучшей оказывается модель NARX (sym4).

На рис. 5 показан прогноз содержания CH_4 в атмосферном воздухе в течение 72 ч с помощью модели NARX и гибридных моделей NARX (db4) и NARX (sym4). Прогнозирование гибридных моделей NARX (db4) и NARX (sym4) наиболее точно отражает изменение концентрации CH_4 за 72 ч. Форма прогно-

за модели NARX (sym4) повторяет наблюдаемый ряд CH_4 .

В целом оценка по р-значению и показателям точности и погрешности совпадают, но есть расхождение. Величина эффекта не всегда определяет статистическую значимость различий. Таким образом, перестановочный подход дополняет наше понимание «хорошего качества» модели. Он дает более точную картину качества моделей.

При прочих равных характеристиках для сравнения различных моделей можно использовать р-значения и можно предложить еще несколько статистик.

Заключение

Рост содержания парниковых газов, в частности метана, приводит к глобальному потеплению [1; 2]. Изменениям в большей степени подвержены чувствительные арктические экосистемы [23; 24]. Для построения политики адаптации к климатическим изменениям используют модельные расчеты возможных сценариев, оценка точности которых может быть критической [25; 26]. Традиционно используемые оценки ошибки модели отражают идею расстояния (различия) между наблюдаемыми и предсказанными значениями (MAE, RMSE и др.). Индексы согласия, с другой стороны, суть меры сходства между предсказанием и наблюдением. Перестановочный подход предлагает ответить на вопрос, насколько вероятно получить наблюдаемое или более экстремальное значение статистики в предположении, что предсказанный и наблюдаемый временные ряды имеют одинаковую природу.

В нашем исследовании было предложено использовать идеологию проверки статистических гипотез в оценке работы модели для прогнозирования временных рядов. Для демонстрации такого подхода использовались предсказательные модели из предыдущего исследования [5; 12]. Авторы полагают, что предложенный перестановочный подход позво-

ляет полнее оценить производительность используемой модели. В нашей работе, основанной на реальных данных мониторинга приземной концентрации парниковых газов на арктическом острове Белый, оценки производительности моделей на основе перестановочного подхода в целом совпали с оценками на основе традиционных индексов [5]. Однако возможны ситуации, при которых перестановочный подход даст другой взгляд на качество применяемой модели. В таком случае решение о применимости модели исследователь может принять, учитывая эту дополнительную информацию. Стоит отметить, что предложенный подход не заменяет применение упомянутых выше «традиционных» индексов для принятия решения о применимости той или иной модели для конкретного исследования, но дополняет их и делает суждения более объективными.

В дальнейших работах авторы планируют протестировать перестановочный метод на других данных для оценки производительности и экстраполяторов, и интерполяторов.

Для измерения концентрации парниковых газов на острове Белый была использована аппаратура Центра коллективного пользования арктических экологических исследований Института промышленной экологии Уральского отделения РАН.

Литература/References

1. Summary for Policymakers. Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. V. Masson-Delmotte, P. Zhai, A. Pirani, S. L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J. B. R. Matthews, T. K. Maycock, T. Waterfield, O. Yelekçi, R. Yu., B. Zhou (eds.). Cambridge, New York, Cambridge Univ. Press, 2021, pp. 3—32. Available at: <https://doi.org/10.1017/9781009157896.001>.
2. Summary for Policymakers. Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. H. Lee and J. Romero (eds). Geneva, 2023, pp. 1—34. DOI: 10.59327/IPCC/AR6-9789291691647.001.
3. Гурлев И. В., Макоско А. А., Малыгин И. Г. Экологические проблемы Арктической зоны России на примере Ямало-Ненецкого автономного округа // Арктика: экология и экономика. — 2024. — Т. 14, № 3. — С. 370—383. — DOI: 10.25283/2223-4594-2024-3-370-383.
4. Gurlev I. V., Makosko A. A., Malygin I. G. Ecological problems of the Russian Arctic Zone on the example of the Yamal-Nenets Autonomous Area. Arctic: Ecology and Economy, 2024, vol. 14, no. 3, pp. 370—383. DOI: 10.25283/2223-4594-2024-3-370-383. (In Russian).
5. Богоявленский В. И., Никонов Р. А., Богоявленский И. В. Новые данные об интенсивной дегазации Земли в Арктике на севере Западной Сибири: термокарстовые озера с кратерами выбросов газа и гра-

зевыми вулканами // Арктика: экология и экономика. — 2023. — Т. 13, № 3. — С. 353—368. — DOI: 10.25283/2223-4594-2023-3-353-368.

Bogoyavlensky V. I., Nikonov R. A., Bogoyavlensky I. V. New data on intensive Earth degassing in the Arctic in the north of Western Siberia: thermokarst lakes with gas blowout craters and mud volcanoes. Arctic: Ecology and Economy, 2023, vol. 13, no. 3, pp. 353—368. DOI: 10.25283/2223-4594-2023-3-353-368. (In Russian).

5. Rakhmatova A., Sergeev A., Shichkin A., Buevich A., Baglaeva E. Three-day forecasting of greenhouse gas CH₄ in the atmosphere of the Arctic Belyy Island using discrete wavelet transform and artificial neural networks. Neural Comput & Applic, 2021, vol. 33, pp. 10311—10322. Available at: <https://doi.org/10.1007/s00521-021-05792-3>.

6. Radersma R., Sheldon B. C. A new permutation technique to explore and control for spatial autocorrelation. Methods Ecol Evol, 2015, vol. 6, pp. 1026—1033. Available at: <https://doi.org/10.1111/2041-210X.12390>.

7. Ross S. M. Simulation, bootstrap statistical methods, and permutation tests. Introduction to Probability and Statistics for Engineers and Scientists. [S. l.], 2021, pp. 619—647. DOI: 10.1016/B978-0-12-824346-6.00024-7.

8. Berry K. J., Johnston J. E., Mielke P. W. A Primer of Permutation Statistical Methods. Cham, Switzerland, Springer-Verlag, 2019, 476 p. DOI: 10.1007/978-3-030-20933-9.

9. Berry K. J., Johnston J. E., Mielke P. W., Johnston L. A. Permutation methods. Part II. Wiley Interdisciplinary Reviews: Computational Statistics, 2018, vol. 10 (3), e1429. DOI: 10.1002/wics.1429.

10. Поддубный В. А., Наговицына Е. С., Маркелов Ю. И. и др. Оценка пространственного распределения концентрации метана в районе Баренцева и Карского морей в летний период 2016—2017 гг. // Метеорология и гидрология. — 2020. — № 3. — С. 77—86. — URL: <https://doi.org/10.15372/AOO20230204>.

Poddubny V. A., Nagovitsyna E. S., Markelov Y. I., Buevich A. G., Antonov K. L., Omel'kova E. V., Manzhurov I. L. Assessment of the spatial distribution of methane concentration in the Barents and Kara Seas in the summer 2016—2017. Russian Meteorology and Hydrology, 2020, vol. 45, no. 3, pp. 193—200. Available at: <https://doi.org/10.15372/AOO20230204>.

11. Субботина И. Е., Бувевич А. Г., Сергеев А. П. и др. Двухшаговый комбинированный алгоритм повышения точности прогнозирования концентрации метана в атмосферном воздухе на основе нейронной сети NARX и последующего прогнозирования невязок // Арктика: экология и экономика. — 2020. — № 2 (38). — С. 59—67. — URL: <https://doi.org/10.25283/2223-4594-2020-2-59-67>.

Subbotina I. E., Buevich A. G., Sergeev A. P., Shichkin A. V., Baglaeva E. M., Remezova M. S. Two-step combined algorithm for improving the accuracy of predicting methane concentration in atmospheric air based on the NARX neural network and subsequent

- prediction of residuals. *Arctic: Ecology and Economy*, 2020, no. 2 (38), pp. 59—67. Available at: <https://doi.org/10.25283/2223-4594-2020-2-59-67>. (In Russian).
12. *Buevich A., Sergeev A., Shichkin A., Baglaeva E.* A two-step combined algorithm based on NARX neural network and the subsequent prediction of the residues improves prediction accuracy of the greenhouse gases concentrations. *Neural Computing and Applications*, 2021, vol. 33 (1), pp. 1547—1557. DOI: 10.1007/s00521-020-04995-4.
13. *Cai L.* Multi-response permutation procedure as an alternative to the analysis of variance: An SPSS implementation. *Behavior Research Methods*, 2006, vol. 38 (1), pp. 51—59. Available at: <https://doi.org/10.3758/BF03192749>.
14. *Fişek M. H., Barlas Z.* Permutation tests for goodness-of-fit testing of mathematical models to experimental data. *Social Science Research*, 2013, vol. 42 (2), pp. 482—495. DOI: 10.1016/j.ssresearch.2012.09.010.
15. *Hart J. D. A., Weiss M. N., Brent L. J. N., Franks D. W.* Common permutation methods in animal social network analysis do not control for non-independence. *Behav Ecol Sociobiol*, 2022, vol. 76, p. 151. Available at: <https://doi.org/10.1007/s00265-022-03254-x>.
16. *Kończak G.* Applications of Permutation Methods in the Analysis of Associations. *Argumenta Oeconomica Cracoviensia*, 2020, vol. 1 (22), pp. 31—45. Available at: <https://doi.org/10.15678/AOC.2020.2203>.
17. *Mielke P. W., Berry K. J.* Permutation Methods. A Distance Function Approach. *Springer Series in Statistics*. New York, Springer, 2001, 446 p. DOI: 10.1007/978-1-4757-3449-2/.
18. *Taylor A. B., MacKinnon D. P.* Four applications of permutation methods to testing a single-mediator model. *Behavior Research Methods*, 2012, vol. 44 (3), pp. 806—844. DOI: 10.3758/s13428-011-0181-x/.
19. *Fekete A., Fuchs S., Garschagen M., Hutter G., Klepp S., Lüder C., Neise T., Sett D., von Elverfeldt K., Wannewitz M.* Adjustment or transformation? Disaster risk intervention examples from Austria, Indonesia, Kiribati and South Africa. *Land Use Policy*, 2022, vol. 120, p. 106230. Available at: <https://doi.org/10.1016/j.landusepol.2022.106230>.
20. *Willmott C. J., Ackleson S., Davis R., Feddema J., Klink K., Legates D., O'Donnell J., Rowe C.* Statistics for the Evaluation and Comparison of Models. *J. of Geophysical Research*, 1985, vol. 90, no. C5, pp. 8995—9005. Available at: <https://doi.org/10.1029/JC090iC05p08995>.
21. *Трофимова И. Е., Балыбина А. С.* Классификация климатов и климатическое районирование Западно-Сибирской равнины // *География и природ. ресурсы*. — 2014. — № 2. — С. 11—21. (In Russian).
- Trofimova I., Balybina A.* Classification of climates and climatic regionalization of the West-Siberian plain. *Geography and Natural Resources*, 2014, vol. 35, pp. 114—122. Available at: <https://doi.org/10.1134/S1875372814020024>.
22. *Daneshfaraz R., Norouzi R., Abbaszadeh H., Azamathulla H.* Theoretical and experimental analysis of applicability of sill with different widths on the gate discharge coefficients. *Water Science & Technology Water Supply*, 2022, vol. 22, pp. 7767—7781. DOI: 10.2166/ws.2022.354.
23. *Эдельгериев Р. С. Х., Романовская А. А.* Новые подходы к адаптации к изменению климата на примере Арктической зоны Российской Федерации // *Метеорология и гидрология*. — 2020. — № 5. — С. 12—28. (In Russian).
- Edel'geriev R. S. K., Romanovskaya A. A.* New Approaches to the Adaptation to Climate Change: The Arctic Zone of Russia. *Russ. Meteorol. Hydrol.*, 2020, vol. 45, pp. 305—316. Available at: <https://doi.org/10.3103/S1068373920050015/>.
24. *Богоявленский В. И., Богоявленский И. В., Каргина Т. Н., Никонов Р. А.* Цифровые технологии дистанционного выявления и мониторинга развития бугров пучения и кратеров катастрофических выбросов газа в Арктике // *Арктика: экология и экономика*. — 2020. — № 4 (40). — С. 90—105. — URL: <https://doi.org/10.25283/2223-4594-2020-4-90-105>.
- Bogoyavlensky V. I., Bogoyavlensky I. V., Kargina T. N., Nikonov R. A.* Digital technologies for remote detection and monitoring of the development of heaving mounds and craters of catastrophic gas blowouts in the Arctic. *Arctic: Ecology and Economy*, 2020, № 4 (40), pp. 90—105. Available at: <https://doi.org/10.25283/2223-4594-2020-4-90-105>. (In Russian).
25. *Tan Z., Zhuang Q.* Arctic lakes are continuous methane sources to the atmosphere under warming conditions. *Environmental Research Letters*, 2015, vol. 10, P. 054016.
26. *Zona D., Gioli B., Commane R. et. al.* Cold season emissions dominate the Arctic tundra methane budget. *Proceedings of the National Academy of Sciences USA*, 2016, vol. 113, pp. 40—45.

Информация об авторах

Баглаева Елена Михайловна, кандидат физико-математических наук, старший научный сотрудник, Институт промышленной экологии Уральского отделения РАН (620990, Россия, Екатеринбург, ул. Софьи Ковалевской, д. 20), e-mail: e.m.baglaeva@urfu.ru.

Сергеев Александр Петрович, кандидат физико-математических наук, заведующий лабораторией, Институт промышленной экологии Уральского отделения РАН (620990, Россия, Екатеринбург, ул. Софьи Ковалевской, д. 20).

Бувевич Александр Геннадьевич, научный сотрудник, Институт промышленной экологии Уральского отделения РАН (620990, Россия, Екатеринбург, ул. Софьи Ковалевской, д. 20).

Шичкин Андрей Васильевич, научный сотрудник, Институт промышленной экологии Уральского отделения РАН (620990, Россия, Екатеринбург, ул. Софии Ковалевской, д. 20).

Субботина Ирина Евгеньевна, кандидат физико-математических наук, научный сотрудник, Институт промышленной экологии Уральского отделения РАН (620990, Россия, Екатеринбург, ул. Софии Ковалевской, д. 20).

NONPARAMETRIC ASSESSMENT OF THE PREDICTIVE ACCURACY OF MODELS USING TIME SERIES OF METHANE CONCENTRATIONS IN THE ATMOSPHERIC AIR OF THE ARCTIC ISLAND OF BELY

Baglaeva, E. M., Sergeev, A. P., Buevich, A. G., Shichkin, A. V., Subbotina, I. E.

Institute of Industrial Ecology of the Ural Branch of the Russian Academy of Sciences (Ekaterinburg, Russian Federation)

The article was received on June 17, 2024

For citing

Baglaeva E. M., Sergeev A. P., Buevich A. G., Shichkin A. V., Subbotina I. E. Nonparametric assessment of the predictive accuracy of models using time series of methane concentrations in the atmospheric air of the Arctic Island of Bely. *Arctic: Ecology and Economy*, 2024, vol. 14, no. 4, pp. 500—510. DOI: 10.25283/2223-4594-2024-4-500-510. (In Russian).

Abstract

Over the past few decades, the number, variety, and complexity of time series forecasting models have grown. There has also been an increased interest among researchers in comparing and assessing the predictive accuracy and performance of models, and in determining which models are more accurate. The researchers propose using the hypothesis testing approach to assess the performance of a time series forecast model. They obtained data for the study while monitoring the dynamics of ground-level concentrations of the main greenhouse gases on the Arctic Island of Bely, Yamalo-Nenets Autonomous Area, Russia. A total of three models based on autoregressive neural networks with exogenous input (NARX) were considered to predict changes in methane concentration in the surface layer of atmospheric air. The performance of the models was assessed using 8 “traditional” indices (correlation and determination coefficients, Wilmott goodness-of-fit indices, mean absolute error, relative mean square error, etc.) and the proposed permutational approach using the hypothesis testing method. In general, based on the data under study, the estimates of the permutation approach and other indicators of accuracy and error coincide, but there are discrepancies. Effect sizes do not always determine the statistical significance of differences.

Keywords: *greenhouse gases, monitoring, climate change, Arctic zone of the Russian Federation.*

Information about the authors

Baglaeva, Elena Mikhailovna, PhD of Physical and Mathematical Sciences, Senior Researcher, Institute of Industrial Ecology of the Ural Branch of the Russian Academy of Sciences (20, Sofia Kovalevskaya St., Ekaterinburg, Russia, 620990), e-mail: e.m.baglaeva@urfu.ru.

Sergeev, Aleksandr Petrovich, PhD of Physical and Mathematical Sciences, Head of Laboratory, Institute of Industrial Ecology of the Ural Branch of the Russian Academy of Sciences (20, Sofia Kovalevskaya St., Ekaterinburg, Russia, 620990).

Buevich, Aleksandr Gennad'evich, Researcher, Institute of Industrial Ecology of the Ural Branch of the Russian Academy of Sciences (20, Sofia Kovalevskaya St., Ekaterinburg, Russia, 620990).

Shichkin, Andrey Vasil'evich, Researcher, Institute of Industrial Ecology of the Ural Branch of the Russian Academy of Sciences (20, Sofia Kovalevskaya St., Ekaterinburg, Russia, 620990).

Subbotina, Irina Evgen'evna, PhD of Physical and Mathematical Sciences, Researcher, Institute of Industrial Ecology of the Ural Branch of the Russian Academy of Sciences (20, Sofia Kovalevskaya St., Ekaterinburg, Russia, 620990).