

ПРОСТРАНСТВЕННО-ВРЕМЕННОЙ АНАЛИЗ ЭВОЛЮЦИИ ПЛОЩАДИ ТЕРМОКАРСТОВЫХ ОЗЕР С ИСПОЛЬЗОВАНИЕМ КОСМИЧЕСКОГО ЗОНДИРОВАНИЯ ЗЕМНОЙ ПОВЕРХНОСТИ И ПРОЦЕДУР РАНДОМИЗИРОВАННОГО МАШИННОГО ОБУЧЕНИЯ И ПРОГНОЗИРОВАНИЯ

Ю. А. Дубнов^{*1}, А. Ю. Попков¹, Ю. С. Попков^{1,2}, М. А. Куприянов³,
В. Ю. Полищук⁴, А. В. Мельников³ и Ю. М. Полищук³

¹Федеральный исследовательский центр «Информатика и управление» РАН, г. Москва, Россия

²Институт проблем управления РАН, г. Москва, Россия

³Югорский научно-исследовательский институт информационных технологий, г. Ханты-Мансийск, Россия

⁴Институт мониторинга климатических и экологических систем СО РАН, г. Томск, Россия

* **Контакт:** Алексей Юрьевич Попков, apopkov@isa.ru, alexeypopkov@yandex.ru

Работа посвящена проблеме прогнозирования эволюции площади термокарстовых озер в зоне вечной мерзлоты Арктики на примере анализа тестовых участков из нескольких географических регионов. Предлагаемый в работе подход основан на методе рандомизированного машинного обучения (РМО) для построения математических моделей динамики площади озер в условиях климатических изменений, ее обучения на реальных данных и дальнейшего прогнозирования. Приводятся и сравниваются результаты моделирования динамики площадей озер с помощью линейных статических и динамических моделей. Показано, что использование динамической модели площади озер позволяет значительно уменьшить среднюю ошибку моделирования.

Ключевые слова: термокарстовые озера, метод информационной энтропии, рандомизированное машинное обучение, статические и динамические модели, пропущенные данные, рандомизированное прогнозирование, климатические изменения.

Цитирование: Дубнов, Ю. А., А. Ю. Попков, Ю. С. Попков, М. А. Куприянов, В. Ю. Полищук, А. В. Мельников и Ю. М. Полищук Пространственно-временной анализ эволюции площади термокарстовых озер с использованием космического зондирования земной поверхности и процедур рандомизированного машинного обучения и прогнозирования // Russian Journal of Earth Sciences. — 2024. — Т. 24. — ES5011. — DOI: 10.2205/2024es000935 — EDN: TGQMBR

1. Введение

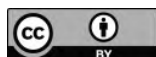
Исследования природных явлений в зонах вечной мерзлоты становятся особенно актуальными в последние годы в связи с наблюдаемым потеплением климата, вызванным парниковыми газами, основными из которых являются углекислый газ и метан. Многочисленные профильные исследования арктических регионов показывают, что существенный вклад в эмиссию метана вносят термокарстовые озера [Миронов и др., 2022; Степаненко и др., 2011; Фельдман, 1984; Karlsson et al., 2014; Kirpotin et al., 2009; Sudakov and Vakulenko, 2014; Zabelina et al., 2020]. Обнаружена связь между площадью термокарстовых озер и объемом выделяемого метана [Walter et al., 2007]. Поэтому прогнозирование эволюции площади озер является важным для оценивания климатических изменений. Источником данных о площади термокарстовых озер и климатических изменениях является технология дистанционного зондирования земной поверхности [Полищук и др., 2020; 2018; Verpoorter et al., 2014].

<https://elibrary.ru/tgqmbR>

Получено: 15 февраля 2024 г.

Принято: 30 августа 2024 г.

Опубликовано: 30 декабря 2024 г.



© 2024. Коллектив авторов.

Процессы образования и эволюции площади термокарстовых озер изучены недостаточно, следствием чего является несовершенство математических моделей, предназначенных для описания процессов их образования. Кроме того, данные о климатических показателях и площадях озер сопровождаются значительными погрешностями и пропусками. Все это свидетельствует о высоком уровне неопределенности при моделировании и обучении модели, и, следовательно, при прогнозировании эволюции площади [Дубнов и др., 2023].

Многие проблемы исследования эволюции площади термокарстовых озер удается эффективно решать, опираясь на методы рандомизированного машинного обучения (РМО) и рандомизированного прогнозирования (РП) [Попков и др., 2019], основанные на максимизации энтропийной меры неопределенности и использовании математических моделей со случайными (рандомизированными) параметрами. Вычислительное ядро метода строится на энтропийно-рандомизированном оценивании функций плотности распределения вероятностей (ПРВ) параметров [Попков и др., 2019; Popkov et al., 2022], базирующемся на принципе условной максимизации функционала информационной энтропии, провозглашенном и развитом в работах [Golan et al., 1996; Jaynes, 1957; Kapur, 1989; Kullback and Leibler, 1951; Levine and Tribus, 1978].

Полученные энтропийно-оптимальные функции ПРВ используются для генерирования ансамбля прогнозных траекторий пространственно-временной динамики термокарстовых озер. Методы РМО и РП применялись для решения ряда задач, в том числе прогнозирования мирового населения [Попков и Дубнов, 2016; Popkov et al., 2016], классификации [Дубнов, 2019], прогнозирования суточной электрической нагрузки в энергетических системах [Попков и др., 2020; Popkov et al., 2020], анализа развития пандемии COVID-19 [Попков, 2021; Попков и др., 2021].

Вклад данной работы в развитие методов анализа данных, моделирования и прогнозирования рассматривается в следующих трех аспектах. Первый аспект состоит в адаптации РМО к особенностям моделирования и обучения стандартными статистическими подходами, основанными на статических регрессионных моделях и динамических моделях временных рядов типа ARMA, тем самым создавая новый инструмент оценивания моделей в рамках общего теоретического инструментария. Второй аспект состоит в применении рандомизированного подхода для восстановления пропущенных данных, но с моделью, отличной от основной, что позволяет методологически корректно подойти к сравнению результатов моделирования с помощью предлагаемого подхода со стандартным. Третий аспект состоит в том, что предлагаемый подход позволил подойти к решению исследуемой задачи в условиях большого количества нерегулярных пропусков в данных, что делает невозможным применение моделей временных рядов и их восстановление стандартными методами.

В первом разделе приводится описание имеющихся реальных данных о площади озер и климатических характеристиках регионов, второй и третий разделы посвящены методологии использования РМО при моделировании и прогнозировании площади термокарстовых озер, в четвертом разделе приводятся результаты экспериментов и их обсуждение.

2. Климатические и спутниковые данные

Имеющиеся в нашем распоряжении данные для анализа динамики площади термокарстовых озер представляют собой таблицы для 51 тестового участка (ТУ), включающие значения средней площади озер (S , га), среднегодовой температуры воздуха (T) и годовой суммы осадков для данной территории (R) за период с 1985 по 2021 гг. Нумерация и расположение ТУ и названия регионов даны на рис. 1, где для краткости «Юг» обозначает южную часть Западно-Сибирской Арктики.

Одной из особенностей подготовки данных для анализа эволюции площадей термокарстовых озер является формирование временных рядов значений зависимой переменной и предикторов для построения математической модели. В настоящей работе используются данные о среднегодовой температуре воздуха и годовой сумме осадков,



Рис. 1. Фрагмент карты арктической зоны России с показанным на ней размещением тестовых участков по территории исследованных регионов.

полученные с помощью реанализа метеорологических полей на основе системы ERA-5 [Полищук и др., 2020].

Данные о площадях озер основаны на результатах измерения площадей озер по спутниковым снимкам системы Landsat, детальное описание которых приводится в работе [Полищук и др., 2020]. Ключевыми особенностями данных о средней площади озер являются наличие пропущенных значений для различных и неравномерных отчетов времени и присутствие возможных погрешностей измерений. Пропуски и погрешности в данных обусловлены в первую очередь отсутствием в некоторые годы безоблачных спутниковых снимков, а также с методикой расчета средней площади по спутниковым изображениям. На рис. 2 приведены доли пропущенных значений для исследуемых ТУ.

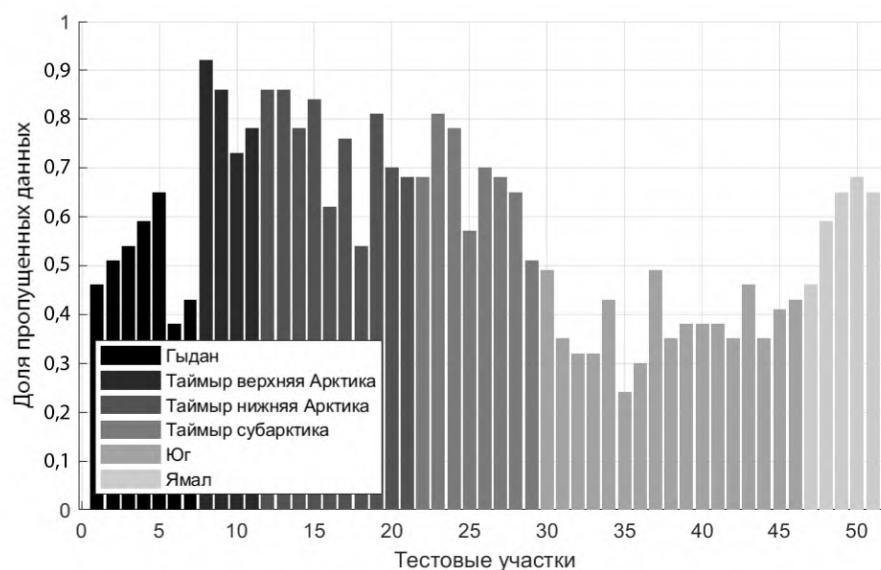


Рис. 2. Доля пропущенных данных о площадях озер для всех ТУ (%).

График на рис. 2 показывает, что доля пропусков в данных о площади озера существенная и в среднем варьируется от 30–40% (Юг) и 50–60% (Гыдан, Ямал) до 70–90% (Таймыр). Наибольшая доля пропущенных данных составляет 85% и более для ТУ № 95, 96, 100, 101, то есть для данных ТУ при построении модели площади имеется лишь 3–5 точек данных за весь период с 1985 по 2021 гг. Таким образом, в связи с малым объемом обучающей выборки и наличием неравномерных пропусков в данных, построение регрессионных моделей классическими статистическими методами, такими как метод наименьших квадратов (МНК) и ARMA/ARIMA методами, становится затруднительным, а проверка необходимых статистических гипотез – невозможной.

3. Рандомизированные модели площади термокарстовых озера

Основы концепции рандомизированного машинного обучения описаны в [Попков, 2023; Попков и др., 2019; Popkov et al., 2022] и сводятся к трем этапам. На первом, согласно принятой модели со случайными параметрами исследуемого объекта и имеющихся данных, определяются энтропийно-оптимальные функции ПРВ параметров модели и измерительных шумов. На втором этапе эти функции сэмпляются и с помощью метода Монте Карло генерируются соответствующие ансамбли траекторий выхода модели. На третьем этапе определяются числовые характеристики указанных ансамблей. На втором и третьем этапах этой процедуры реализуется алгоритм рандомизированного прогнозирования.

В настоящей работе рассматриваются две модели площади термокарстовых озера S :

1. Линейная статическая модель (ЛСМ) с двумя входами: среднегодовой температурой T и годовой суммой осадков R [Дубнов и др., 2021a];
2. Линейная динамическая модель (ЛДМ) с теми же входами, но учитывающая предыдущие значения самой площади, таким образом, входами такой модели являются: среднегодовая температура T , осадки R и p исторических (предыдущих) значений площади $S_{-p}, \dots, S_{-1}$ [Дубнов и Булычев, 2017].

3.1. Линейная статическая модель

Рассмотрим линейную статическую модель площади следующего вида:

$$\begin{aligned} S[t|a] &= a_0 + a_1 T[t] + a_2 R[t], \\ v[t|a] &= S[t|a] + \xi[t], \quad t \in \mathcal{T} = [0, T], \end{aligned} \quad (1)$$

где $v[t]$ – измеряемая площадь озера, $\xi[t]$ – случайный шум, имитирующий измерительные ошибки, сопровождающие трансформацию космических изображений в значения площади озера, $\mathcal{T}$ – временной интервал, на котором осуществляется обучение модели.

Точки, в которых происходят измерения, составляют счетное и конечное множество размера m , соответствующего количеству точек измерения, поэтому иногда удобно использовать индексную нотацию вместо дискретной:

$$\begin{aligned} \mathcal{T} = [0, T] &= [t_1, t_1, \dots, t_m], \\ S[t_j] &= S_j, \quad T[t_j] = T_j, \quad R[t_j] = R_j, \quad v[t_j] = v_j, \quad \xi[t_j] = \xi_j, \quad j = \overline{1, m}, \end{aligned}$$

таким образом из (1) будем иметь:

$$\begin{aligned} S_j(a) &= a_0 + a_1 T_j + a_2 R_j, \\ v_j(a) &= S_j(a) + \xi_j, \quad j = \overline{1, m}. \end{aligned} \quad (2)$$

С учетом (1) и (2) можно определить векторный функционал Φ , реализующий преобразование входа в выход в виде:

$$y = \Phi(\mathbf{x}, \mathbf{a}) = \Phi(T, R, \mathbf{a}) = a_0 + a_1 T + a_2 R, \quad \mathbf{x} = (T, R), \quad \mathbf{a} = (a_0, a_1, a_2), \quad (3)$$

где $y = S$ – выход модели, в данном случае скалярный, $\mathbf{x}$ – вектор входа, $\mathbf{a}$ – вектор параметров модели.

Важно отметить, что в общем случае измерения могут проводиться не регулярно, то есть расстояние между соседними моментами времени не равны между собой для всех точек. Реальные данные наблюдений имеют регулярный характер за исключением пропусков, вследствие этого далее предполагается регулярность измерений и в используемых моделях.

Параметры модели предполагаются случайными интервального типа

$$a_i \in \mathcal{A}_i = [a_i^-, a_i^+], \quad a \in \mathcal{A} = \bigcup_{i=0}^2 \mathcal{A}_i, \quad (4)$$

вероятностные свойства которых характеризуются непрерывно-дифференцируемой функцией ПРВ $P(a)$, определенной на $\mathcal{A}$.

Измерительный шум $\xi[t]$ также предполагается случайным, независимым и интервального типа:

$$\xi \in \Xi_j = [\xi_j^-, \xi_j^+], \quad \xi = \{\xi_1, \dots, \xi_m\} \in \Xi = \bigcup_{j=1}^m \Xi_j, \quad (5)$$

вероятностные свойства которого характеризуются совместной функцией ПРВ $Q(\xi)$, которая с учетом независимости измерений определяется как

$$Q(\xi) = \prod_{j=1}^m q_j(\xi), \quad (6)$$

где $q_j(\xi)$ – функции ПРВ шума в j -й точке измерения.

3.2. Линейная динамическая модель

Рассмотрим теперь динамическую модель площади

$$\begin{aligned} S[t] &= a_0 + \sum_{k=1}^p a_k S[t-k] + b_1 T[t] + b_2 R[t], \\ v[t] &= S[t] + \xi[t], \quad t = [0, T], \end{aligned} \quad (7)$$

где p – порядок модели, $a_0, \dots, a_p, b_1, b_2$ – параметры модели, индекс $-k$ означает k -е предыдущее значение времени по отношению к исходному.

С учетом индексной нотации будем иметь:

$$\begin{aligned} S_j &= a_0 + \sum_{k=1}^p a_k S_{j-k} + b_1 T_j + b_2 R_j, \\ v_j &= S_j + \xi_j. \end{aligned} \quad (8)$$

С учетом (7) и (8) определим векторный функционал Ψ , реализующий преобразование входа в выход в виде:

$$\begin{aligned} y &= \Psi(\mathbf{x}, \mathbf{a}) = \Phi(T, R, S_{-1}, \dots, S_{-p}, \mathbf{a}) = a_0 + \sum_{k=1}^p a_k S_{-k} + b_1 T + b_2 R, \\ \mathbf{x} &= (T, R, S_{-1}, \dots, S_{-p}), \quad \mathbf{a} = (a_0, a_1, \dots, a_p, b_1, b_2), \end{aligned} \quad (9)$$

где $y = S$ – скалярный выход модели, $\mathbf{x}$ – вектор входа, $\mathbf{a}$ – вектор параметров модели.

Как и в случае статической модели, ее параметры и независимые шумы предполагаются случайными интервального типа

$$a_i \in \mathcal{A}_i = [a_i^-, a_i^+], \quad b_\ell \in \mathcal{B}_\ell = [b_\ell^-, b_\ell^+], \quad a \in \mathcal{A} = \bigcup_{i=0}^p \mathcal{A}_i \bigcup_{\ell=1}^2 \mathcal{B}_\ell, \quad (10)$$

$$\xi \in \Xi_j = [\xi_j^-, \xi_j^+], \quad \xi = \{\xi_1, \dots, \xi_m\} \in \Xi = \bigcup_{j=1}^m \Xi_j, \quad (11)$$

вероятностные свойства которых характеризуются непрерывно-дифференцируемой функцией ПРВ параметров $P(a)$ определенной на $\mathcal{A}$, и функцией ПРВ шумов $Q(\xi)$ определяемой (6).

Динамическая модель в отличие от статической позволяет учесть влияние предыдущих значений площади на ее значение в текущий момент времени. Наличие такого влияния исключает пропуски в данных, и требует предварительного восстановления пропущенных значений. Для восстановления пропусков в данных предлагается использовать рассмотренную ранее модель [Дубнов и др., 2021b]. Примеры восстановления данных приведен на рис. 3 для ТУ № 1 и ТУ № 95.

4. Обучение рандомизированных моделей

Задача энтропийного оценивания (первый этап РМО) формулируется следующим образом:

$$H(P, Q) = - \int_{\mathcal{A}} P(a) \ln P(a) da - \int_{\Xi} Q(\xi) \ln Q(\xi) d\xi \rightarrow \max_{P, Q}, \quad (12)$$

при условиях:

- нормировки

$$\int_{\mathcal{A}} P(a) da = 1, \quad \int_{\Xi} Q(\xi) d\xi = 1; \quad (13)$$

- соблюдения эмпирических балансов

$$\int_{\mathcal{A}} \Omega(\mathbf{x}_j, \mathbf{a}) P(a) da + \int_{\Xi_j} \xi q_j(\xi) d\xi = y_j, \quad j = \overline{1, m}, \quad (14)$$

где Ω – функционал модели, реализующий преобразование входа в выход, $\mathbf{x}_j$ – вход в j -й точке наблюдения, определяемые (3) или (9), y_j – наблюдаемое (реальное) значение площади на интервале обучения T .

Описанная оптимизационная задача представляет собой функциональную энтропийно-линейную задачу с интегральными ограничениями, которая имеет аналитическое решение в предположении, что функции $P(a)$ и $Q(\xi)$ являются непрерывно-дифференцируемыми функциями [Попков и др., 2019]:

$$P^*(a, \lambda) = \frac{\exp(-\sum_{j=1}^m \lambda_j \Omega(x_j, a))}{\mathcal{P}(\lambda)}, \quad \mathcal{P}(\lambda) = \int_{\mathcal{A}} \exp\left(-\sum_{j=1}^m \lambda_j \Omega(\mathbf{x}_j, a)\right) da, \quad (15)$$

$$Q^*(\xi, \lambda) = \frac{\exp(-\sum_{j=1}^m \lambda_j \xi_j)}{\mathcal{Q}(\lambda)}, \quad \mathcal{Q}(\lambda) = \int_{\Xi} \exp\left(-\sum_{j=1}^m \lambda_j \xi_j\right) d\xi, \quad (16)$$

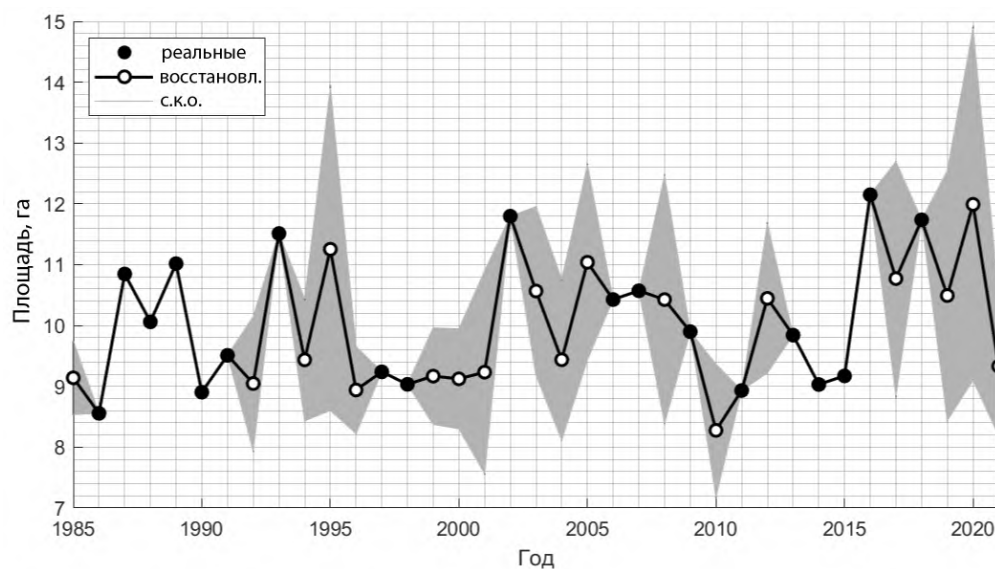
где $\lambda = (\lambda_1, \dots, \lambda_m)$ – вектор множителей Лагранжа, которые определяются решением балансовых уравнений:

$$R_j(\lambda) = \frac{1}{P(\lambda)} \int_{\mathcal{A}} \Omega(\mathbf{x}_j, a) \exp \left(- \sum_{j=1}^m \lambda_j \Omega(\mathbf{x}_j, a) \right) da + \frac{1}{Q(\lambda)} \int_{\Xi} \xi \exp \left(- \sum_{j=1}^m \lambda_j \xi_j \right) d\xi = y_j. \quad (17)$$

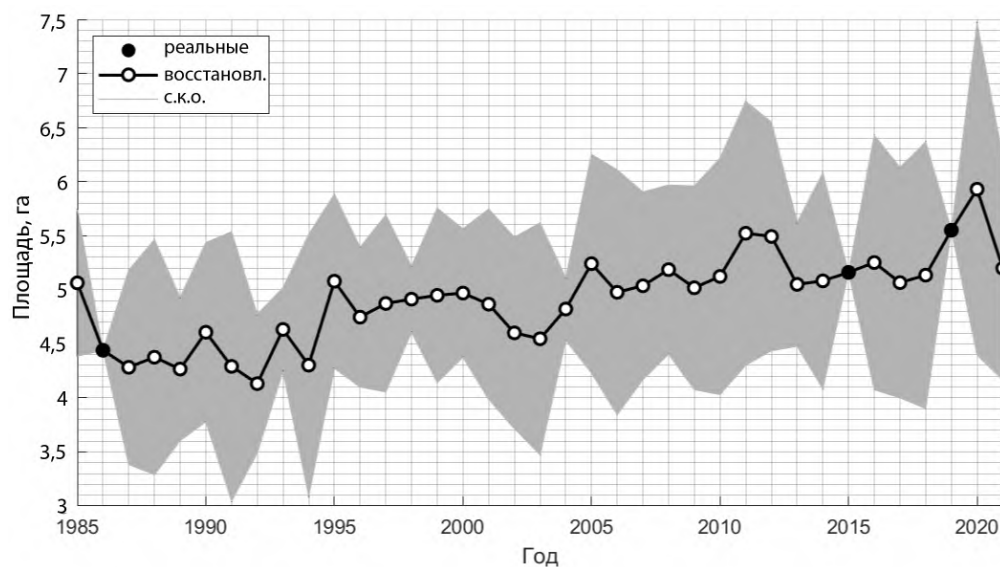
Решение указанных уравнений определяется с использованием алгоритмов минимизации квадратичной невязки [Golan et al., 1996]

$$J(\lambda) = \|R(\lambda) - Y\|^2 \Rightarrow \min_{\lambda}, \quad (18)$$

где $\|\cdot\|$ – евклидова норма, $R(\lambda)$ и Y – векторы с компонентами R_j и y_j соответственно.



(а) ТУ № 1.



(б) ТУ № 95.

Рис. 3. Восстановление значений площади.

Таким образом, в результате применения энтропийно-рандомизированного оценивания восстанавливаются функция ПРВ для параметров модели $P^*(a)$ и измерительных шумов $Q^*(\xi)$, которые в дальнейшем используются для сэмпирования и генерации соответствующих ансамблей и для вычисления точечных оценок среднего $\bar{a}$ и дисперсии σ_a^2 параметров:

$$\bar{a} = \mathbb{E}[a] = \int_{\mathcal{A}} aP(a)da, \quad \sigma_a^2 = D[a] = \int_{\mathcal{A}} (a - \mathbb{E}[a])^2 P(a)da. \quad (19)$$

Аналогичным образом оцениваются средние значения и дисперсии измерительных шумов.

5. Прогнозирование с использованием рандомизированных моделей

Концепция рандомизированного прогнозирования, предложенная в работах [Попков, 2023; Попков и др., 2019; Popkov et al., 2022], может применяться для любых параметрических моделей со случайными параметрами и известных функций ПРВ. Согласно данной методике для прогнозирования будущих траекторий процесса используются не точечные оценки параметров, а ансамбль возможных значений случайных величин, полученный посредством сэмпирования соответствующих функций ПРВ. В данной работе предлагается модифицировать данный подход, используя технологию bootstrap для размножения выборки. Он нацелен на уменьшение дисперсии прогнозных значений благодаря многократному сэмпированию одной и той же функции ПРВ с вычислением эмпирических средних значений и последующим усреднением результатов.

Алгоритм рандомизированного прогнозирования площади термокарстовых озер, ориентированный на использование линейных моделей площади представляет собой следующую процедуру:

1. Полученные функции ПРВ параметров и измерительных шумов сэмперируются и соответствующие последовательности случайных величин используются для построения ансамбля траекторий, характеризующих временную эволюцию площади термокарстовых озер. В случае линейных моделей площади, вычисляются выборочные средние значения и с их помощью строится прогнозная траектория;
2. Предыдущий этап повторяется заданное количество раз (например, 1000), в результате чего ансамбль прогнозных траекторий усредняется повторно и вычисляется итоговая средняя траектория динамики площади озер.

В данной работе, для сэмпирования используется метод исключений (Acceptance-Rejection method) [Neumann, 1951], который иногда приводит к увеличению общего времени сэмпирования, особенно для функций ПРВ экспоненциального типа, когда сэмплируемые значения должны быть сосредоточены в малой части области определения функции.

Область определения функций ПРВ $P(a)$ и $Q(\xi)$ задается на этапе оценивания, исходя из ограничений на значения параметров модели и шумов измерений. В данной работе отрезки значений для параметров модели формируются с помощью предварительных оценок по методу МНК по правилу «трех сигм» [Феллер, 1967].

6. Результаты исследования

Реализация всех вычислительных экспериментов проводилась в среде MATLAB версии 9.7 (2019b) на платформе Windows x64. Для решения системы уравнений (18) использовался алгоритм Trust-Region-Dogleg [Nocedal and Wright, 2006], реализованный в модуле Optimization Toolbox, для сэмпирования энтропийно-оптимальных распределений использовался алгоритм AR собственной разработки.

6.1. Моделирование климатических показателей

Прежде чем приступить к прогнозированию площади озер на основе среднегодовых значений температуры и количества осадков, необходимо построить вспомогательные модели для прогнозирования данных значений. В [Дубнов и др., 2023] предлагалась линейная динамическая модель со случайной последовательностью на входе. В процедуру обучения включалась совместная функция ПРВ параметров и входного шума, которая восстанавливалась с помощью решения соответствующей задачи энтропийно-рандомизированного оценивания.

Здесь мы развиваем иной подход, предложенный в работах [Полищук и др., 2020; 2018], где температура и осадки описывались линейными моделями от времени. Такая модель позволяет оценить общий тренд, к примеру, на рост средней температуры со временем, но, с другой стороны, никак не учитывает динамику изменения температуры за весь период наблюдений. В данном исследовании предлагается дополнить линейную модель для температуры и осадков посредством введения циклических составляющих:

$$T[t] = \mu + \alpha t + \beta_1 \sin\left(\frac{2\pi t}{4}\right) + \beta_2 \sin\left(\frac{2\pi t}{8}\right) + \beta_3 \sin\left(\frac{2\pi t}{12}\right) + \varepsilon, \quad (20)$$

$$t \in T.$$

Аналогичная модель использовалась для описания временной эволюции годовой суммы осадков $R[t]$.

Частоты с периодом 4, 8 и 12 выбраны так потому, что для большинства ТУ на данных значениях наблюдаются пики функций автокорреляции значений температуры и осадков. Параметры моделей температуры и осадков определялись МНК. Пример применения данной модели представлен на рис. 4.

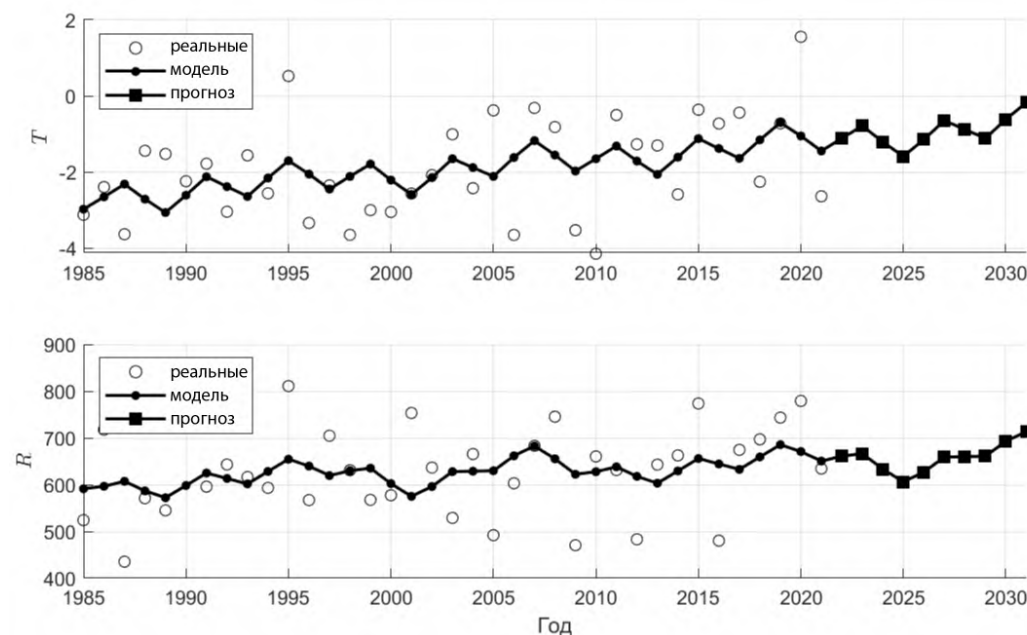


Рис. 4. Прогнозирование температуры и осадков для ТУ № 1.

6.2. Моделирование площади по исходным данным с пропусками

Как было отмечено выше, применение динамических моделей сопряжено с необходимостью иметь данные предыдущих состояний исследуемого процесса в количестве, определяемым порядком модели. В условиях наличия пропусков в данных, это обстоятельство приводит к необходимости их восстановления.

Исследуемая в работе задача характеризуется большим количеством пропущенных данных для значительного количества ТУ. По этой причине, в целях их восстановления, для моделирования используется линейная модель от значений климатических показателей (1) и метод энтропийно-рандомизированного оценивания для вычисления оценок функций ПРВ параметров данной модели.

На примере одного из ТУ рассмотрим алгоритм моделирования площади и результаты его работы. Для моделирования используется модель вида (1). Вектора входов для каждой точки наблюдения образуют строки матрицы входов X , первый столбец которой соответствует единичному вектору, наблюдения площади образуют матрицу наблюдений Y . Эта матрица определяет набор данных (датасет) для обучения:

$$Y_{m \times 1} = S, \quad X_{m \times 3} = [\mathbf{1}, \mathbf{T}, \mathbf{R}], \mathbf{a}_{3 \times 1} = (a_0, a_1, a_2)^T.$$

С учетом требований статистической теории, данные необходимо стандартизировать для удовлетворения требованиям нормальности:

$$\begin{aligned} X_{\min} &= \min(X^k), & X_{\max}^k &= \max(X^k), \\ Y_{\min} &= \min(Y), & Y_{\max} &= \max(Y), \\ X^k &= \frac{X^k - X_{\min}^k}{X_{\max}^k - X_{\min}^k}, & Y &= \frac{Y - Y_{\min}}{Y_{\max} - Y_{\min}}. \end{aligned}$$

где операции минимума и максимума берутся по столбцам.

Оценка модели методом наименьших квадратов приводит к следующим значениям среднего и дисперсии:

$$\begin{aligned} \mathbf{a}_{ols} &= (0,3250, 0,8059, -0,3759)^T, \quad \sigma^2 = 0,0459, \\ \sigma_{\mathbf{a}} &= (0,0886, 0,1653, 0,1478)^T. \end{aligned}$$

На основе этих оценок и правила «трех сигм» определяются интервалы случайных параметров (4) и шума (5), причем интервалы шумов одинаковы для всех точек наблюдения:

$$\begin{aligned} \mathcal{A} &= [\mathbf{a}_{ols} - 3\sigma_{\mathbf{a}}, \mathbf{a}_{ols} + 3\sigma_{\mathbf{a}}] = \begin{bmatrix} 0,0591 & 0,5910 \\ 0,3099 & 1,3019 \\ -0,8192 & 0,0675 \end{bmatrix}, \\ \Xi &= [-3\sigma, 3\sigma] = [-0,6424, 0,6424]. \end{aligned} \quad (21)$$

Решение уравнения (18) приводит к определению оптимального значения λ , которые определяют энтропийно-оптимальные распределения параметров в соответствии с (15)–(16) (см. табл. 1 и рис. 5).

Таблица 1. Оптимальные значения множителей Лагранжа λ

j	1	2	3	4	5	6	7	8	9	10
λ_j	-1,3274	-1,9910	2,0579	-0,8858	0,8545	-0,9223	-1,4929	-0,7539	2,6871	-0,8233
j	11	12	13	14	15	16	17	18	19	20
λ_j	1,0640	-1,1957	-0,6579	-0,2567	-1,0819	-1,0518	-0,1353	5,8824	-0,6201	-0,4658
j	21	22	23	24	25	26	27	28	29	30
λ_j	-0,5856	2,5382	-0,2708	0,3559	0,2002	-1,0294	-6,2172	-0,9045	-0,9670	-1,1241
j	31	32	33	34	35	36	37			
λ_j	-2,6671	1,9400	0,1695	9,3582	0,3763	1,2592	-0,6660			

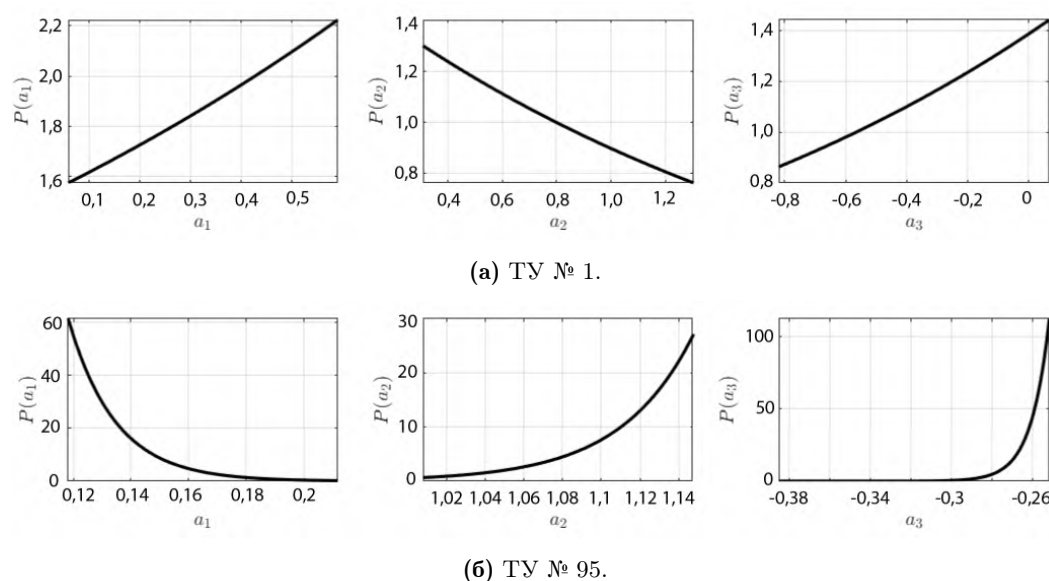


Рис. 5. Энтропийно-оптимальные распределения параметров модели.

Имея энтропийно-оптимальные распределения, можно вычислить средние (по распределению) значения параметров:

$$\mathbf{a}_{mee} = \mathbb{E}[\mathbf{a}] = \int_{\mathcal{A}} \mathbf{a}P(\mathbf{a})d\mathbf{a} = (0,3403, 0,7622, -0,3382)^T.$$

Результат моделирования и прогнозирования на 10 лет для ТУ № 1 и № 95 показан на [рис. 6](#).

Указанную последовательность действий необходимо повторить для каждого ТУ. Качество моделирования оценивается с помощью среднеквадратической ошибки RMSE и средней абсолютной процентной ошибки MAPE между прогнозной (модельной) $\widehat{S}$ и реальной траекториями S :

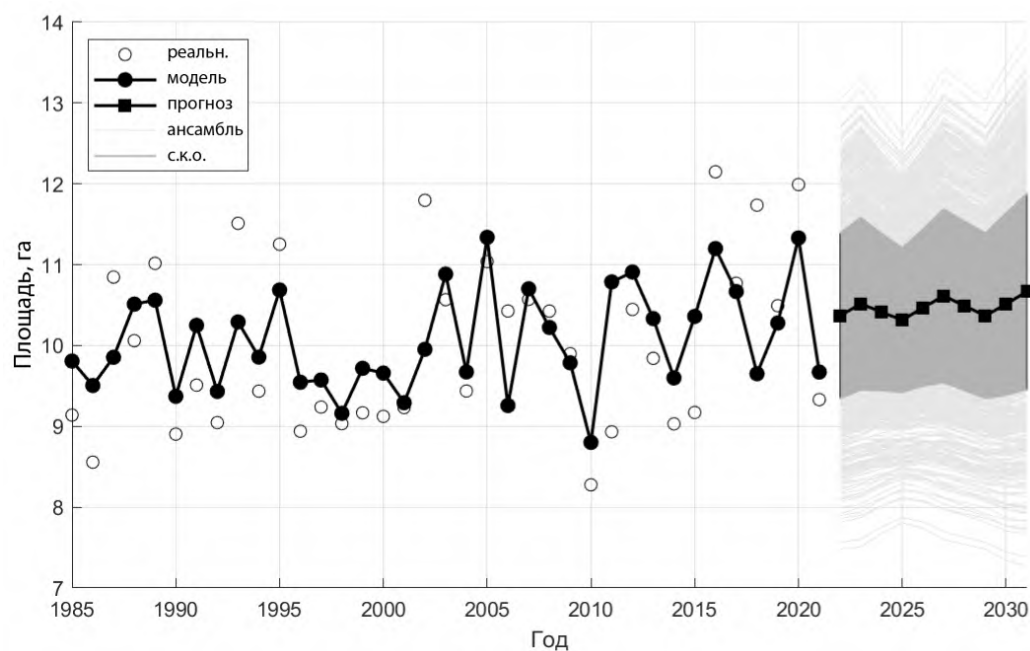
$$\text{RMSE} = \left(\frac{1}{n} \sum_{i=1}^n (S_i - \widehat{S}_i)^2 \right)^{1/2}, \quad \text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{S_i - \widehat{S}_i}{S_i} \right|. \quad (22)$$

Результаты экспериментов представлены в [рис. 2](#). Для каждого региона приведены усредненная по региону доля пропущенных значений ρ и ошибок RMSE и MAPE, а на [рис. 7](#) приведены значения ошибки MAPE для каждого тестового участка.

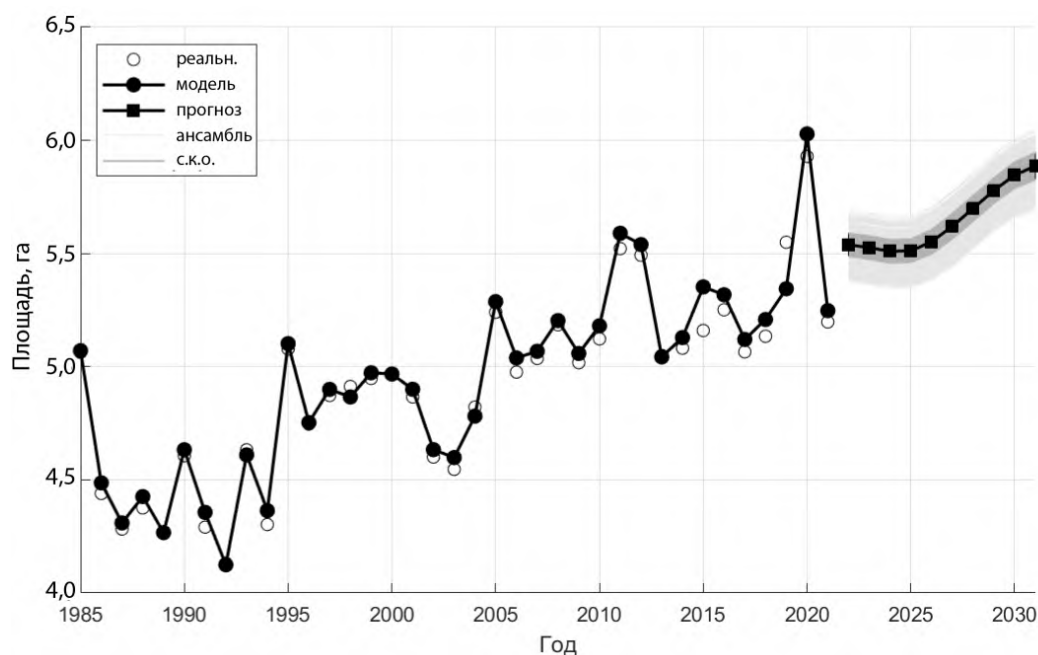
Таблица 2. Результаты моделирования по данным с пропусками

Регион	ρ	RMSE	MAPE
Юг	0,3782	0,9168	5,7762
Гыдан	0,5086	0,9274	7,9677
Ямал	0,6060	0,7126	10,0550
Таймыр верхняя Арктика	0,8225	0,4134	4,6721
Таймыр нижняя Арктика	0,7450	0,3536	3,1989
Таймыр субарктика	0,6725	0,3708	2,8365

Согласно результатам, приведенным в [рис. 2](#) и на [рис. 7](#), ошибка моделирования площади MAPE для большинства регионов не превышает 10%, однако для некоторых тестовых участков превосходит данное значение. Это может быть обусловлено как недостаточным качеством линейной модели, так и высокой долей пропущенных значений, в результате чего оценивание проводилось по малому набору точек.



(а) ТУ № 1.



(б) ТУ № 95.

Рис. 6. Результаты моделирования площади.

6.3. Моделирование площади по восстановленным данным

В данном разделе приведены результаты моделирования площади озёр с помощью линейной динамической модели (7) по восстановленным данным. Порядок p модели выбирается индивидуально для каждого тестового участка с целью минимизации среднеквадратической ошибки моделирования. Результаты экспериментов приведены в рис. 3 по регионам.

Полученные результаты показывают, что ошибка моделирования при использовании линейной динамической модели уменьшилась более чем в 2 раза в среднем по регионам по сравнению с использованием линейной статической модели.

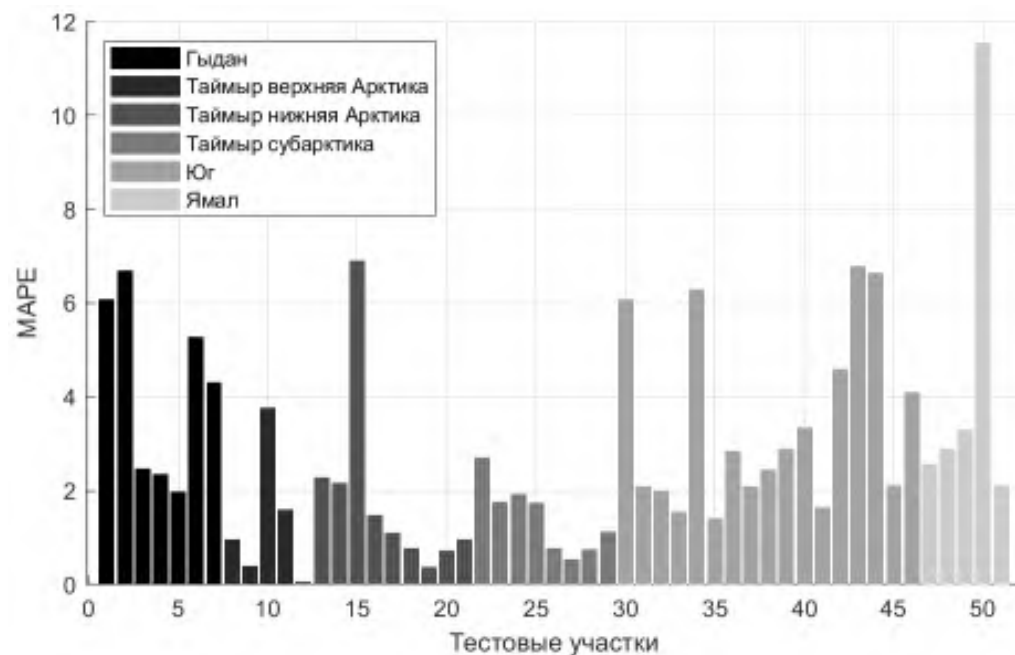


Рис. 7. Ошибки модели (1).

Таблица 3. Результаты моделирования по данным с пропусками

Регион	ρ	RMSE	MAPE
Юг	0,3782	0,5443	3,4984
Гыдан	0,5086	0,4844	4,0794
Ямал	0,6060	0,7908	4,5316
Таймыр верхняя Арктика	0,8225	0,1887	1,8872
Таймыр нижняя Арктика	0,7450	0,1691	1,5066
Таймыр субарктика	0,6725	0,1938	1,3951

6.4. Результаты прогнозирования площади

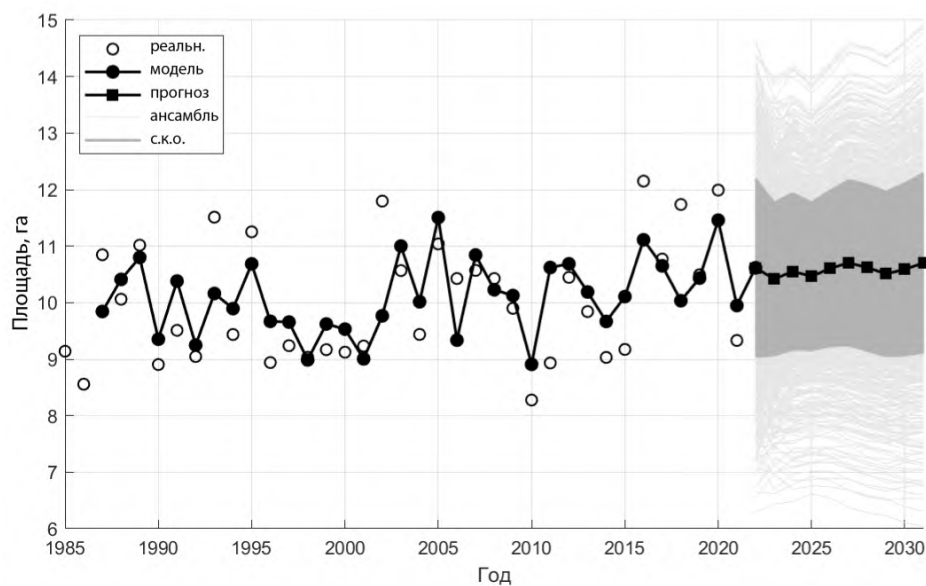
Прогнозирование площади термокарстовых озер проводилось отдельно для каждого тестового участка с помощью описанного ранее алгоритма рандомизированного прогнозирования и с использованием линейной динамической модели порядка p . Прогноз строился на ближайшие 10 лет отдельно для каждого тестового участка. Для примера приведены графики прогнозных траекторий для ТУ № 1 и ТУ № 95 (рис. 8).

6.5. Агрегация результатов

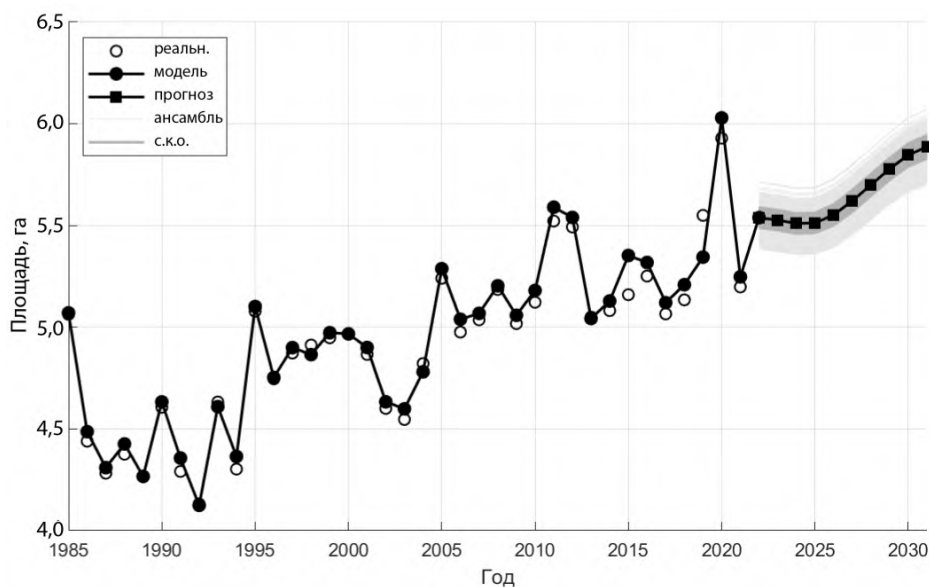
Учитывая большое количество ТУ трудно проанализировать результаты моделирования для всего региона. Для достижения целей обобщения полученных результатов на весь регион и его представление необходимо провести их агрегацию по всем ТУ, входящим в регион.

Моделирование площади термокарстовых озер проводилось с помощью обучения рандомизированной динамической модели (?). Прогноз строится на ближайшие 10 лет методом рандомизированного прогнозирования отдельно для каждого тестового участка.

На рис. 9 приведен график распределения порядка модели для всех тестовых участков с разбивкой по регионам. Так, например, для региона «Гыдан» 4 из 7 тестовых участков моделировались с помощью модели порядка 4, а остальные 3 – моделями порядка 1, 2 и 3. Преобладание в данном регионе моделей более высокого порядка может свидетельствовать о большей зависимости площади озер от значений предыдущих лет. С другой стороны, для региона «Юг» наблюдается обратная ситуация, когда



(а) ТУ № 1.



(б) ТУ № 95.

Рис. 8. Прогнозирование площади.

большинство ТУ моделировались моделями 0, 1 или 2 порядка и лишь несколько участков более высокими.

На рис. 10 приведен графики среднего прогнозируемого изменения площади озёр за ближайшие 10 лет по регионам. Изменение площади вычисляется относительно среднего значения за последние 3 года существующих данных, т.е. 2019, 2020 и 2021, и показывает средний относительный прирост площади

$$\bar{y} = \frac{1}{3}(y_{2019} + y_{2020} + y_{2021}),$$

$$\Delta S_n = \frac{y_n - \bar{y}}{\bar{y}}, \quad n = 2022, 2023, \dots, 2031. \quad (23)$$

Согласно графикам, приведенным на рис. 10, можно выделить группу регионов «Таймыр верхняя Арктика», «Таймыр нижняя Арктика» и «Таймыр субарктика», для

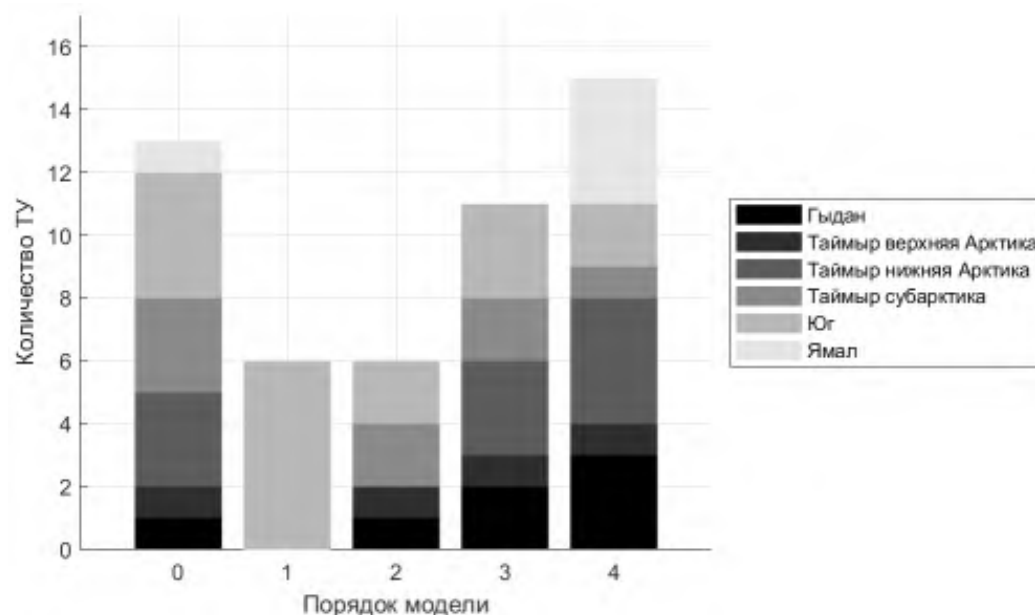


Рис. 9. Распределение по порядку модели для всех ТУ.

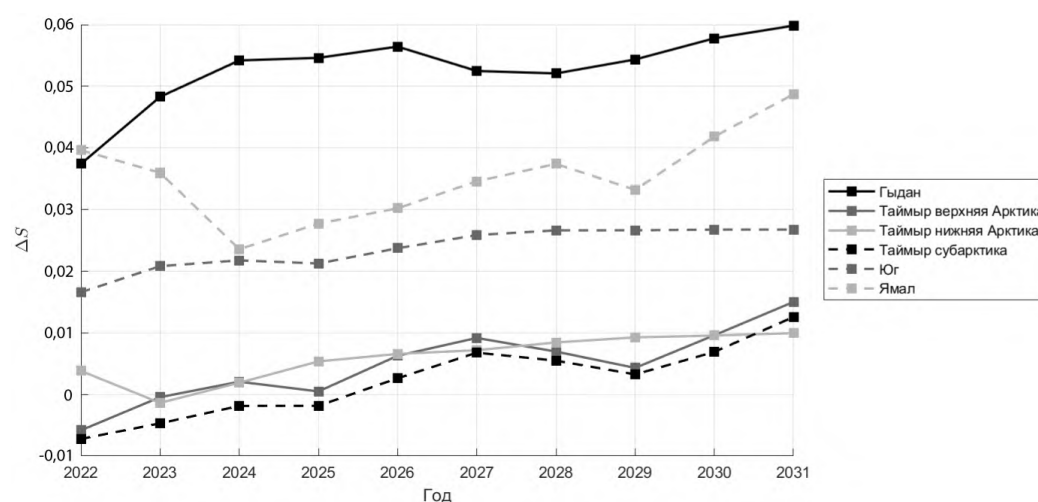


Рис. 10. Прогноз среднего изменения площади озёр по регионам.

которых графики ведут себя похожим образом, а среднее прогнозируемое изменение площади к 2031 году составляет 1–2%. Примерно такое же итоговое среднее значение прироста площади в 2% прогнозируется для региона «Юг», однако график для данного региона выглядит более «ровным» и «стабильным» благодаря большому количеству тестовых участков в нём. Наибольшие прирост средней площади около 4% и 5% прогнозируется для регионов «Ямал» и «Гыдан» соответственно.

На рис. 11 приведен график изменения площади тестовых участков. Точки соответствуют координатам тестовых участков, размер маркера обозначает степень изменения, а цвет соответствует региону.

В рис. 4 приведены некоторые показатели, полученные в рамках экспериментов. В целом можно отметить следующее:

- Гыдан – наибольший прирост 13.9% –ТУ № 16;
- Таймыр верхняя Арктика – наибольший прирост 9% –ТУ № 95;
- Таймыр нижняя Арктика – наибольший прирост 5.5% –ТУ № 102;
- Таймыр субарктика – наибольший прирост 4% –ТУ № 111;
- Юг – наибольший прирост 16.9% –ТУ № 14;
- Ямал – наибольший прирост 6.6% –ТУ № 28.

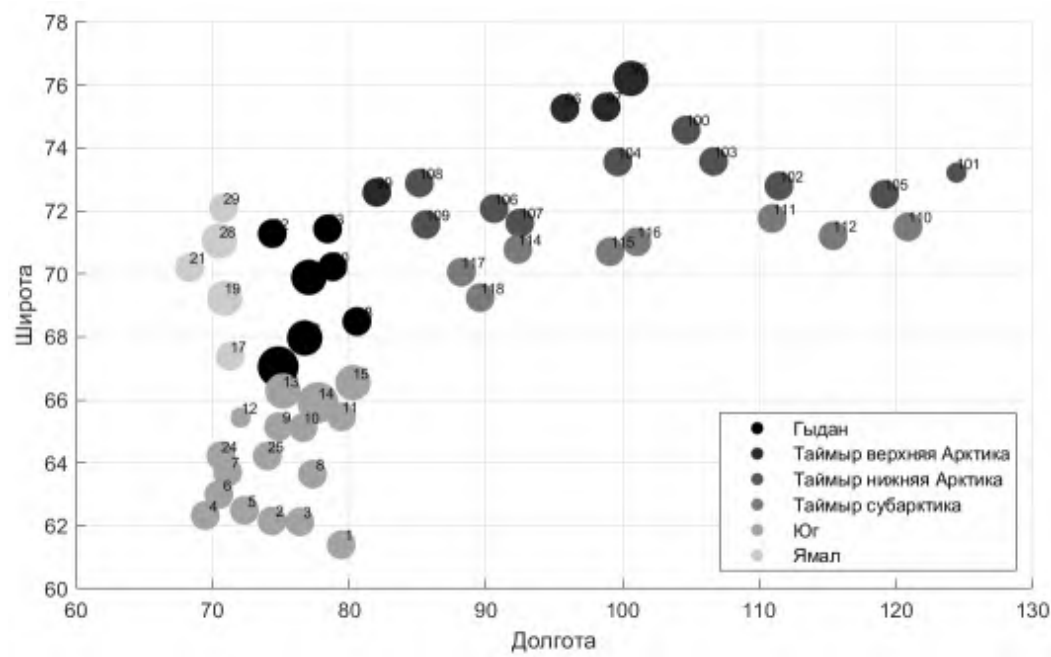


Рис. 11. Карта изменения площади ТУ.

Таблица 4. Итоговые показатели

Название региона	Количество ТУ	Среднее значение	ТУ с наименьшим значением	Наименьшее значение	ТУ с наибольшим значением	Наибольшее значение
Гыдан	7	0,0514	23	−0,0075	16	0,1391
Таймыр верхняя Арктика	4	0,0183	97	−0,0202	95	0,0899
Таймыр нижняя Арктика	10	0,0122	101	−0,0075	102	0,0548
Таймыр субарктика	8	0,0136	118	−0,0061	111	0,0401
Юг	17	0,0225	12	−0,0310	14	0,1688
Ямал	5	0,0422	29	0,0331	28	0,0656
ВСЕГО	51	0,0247	12	−0,0310	14	0,1688

7. Заключение

В работе рассматривается задача моделирования и прогнозирования площади термокарстовых озер с помощью метода рандомизированного машинного обучения, позволяющего строить статистические модели в условиях малого объема данных и неструктурированных шумов измерений. В ходе экспериментов было рассмотрено две модели площади озер: линейная статическая модель от значения климатических показателей температуры и количества осадков и линейная динамическая модель, позволяющая учитывать динамику исторических значений площади. В результате моделирования было показано, что переход к динамической модели площади и восстановление пропущенных данных позволяют значительно уменьшить среднюю ошибку моделирования.

Благодарности. Работа выполнена при финансовой поддержке Российского научного фонда (проект № 22-11-20023).

Список литературы

Дубнов Ю. А. Энтропийное оценивание в задачах классификации // Автоматика и телемеханика. — 2019. — № 3. — С. 138—151. — DOI: [10.1134/S0005231019030097](https://doi.org/10.1134/S0005231019030097).
Дубнов Ю. А. и Булычев А. В. Байесовская идентификация смешанных гауссовских моделей // Информационные технологии и вычислительные системы. — 2017. — № 1. — С. 101—114.

- Дубнов Ю. А., Полищук В. Ю., Попков А. Ю. *и др.* Энтропийно-рандомизированное прогнозирование эволюции площади термокарстовых озёр // Челябинский физико-математический журнал. — 2021а. — Т. 6, № 3. — С. 384–396. — DOI: [10.47475/2500-0101-2021-16312](https://doi.org/10.47475/2500-0101-2021-16312).
- Дубнов Ю. А., Полищук В. Ю., Попков Ю. С. *и др.* Метод энтропийно-рандомизированного восстановления пропущенных данных // Автоматика и телемеханика. — 2021b. — С. 140–160. — DOI: [10.31857/S0005231021040061](https://doi.org/10.31857/S0005231021040061).
- Дубнов Ю. А., Попков А. Ю., Полищук В. Ю. *и др.* Алгоритмы рандомизированного машинного обучения для прогнозирования эволюции площади термокарстовых озёр в зонах вечной мерзлоты // Автоматика и телемеханика. — 2023. — № 1. — DOI: [10.31857/S0005231023010051](https://doi.org/10.31857/S0005231023010051).
- Миронов М. С., Шорникова А. В., Сидорина И. Е. *и др.* Изучение эмиссии метана в термокарстовых озерах полуострова Ямал с помощью методов дистанционного зондирования Земли и наземных исследований // Материалы 20-й Международной конференции «Современные проблемы дистанционного зондирования Земли из космоса». — Москва : Институт космических исследований РАН, 2022. — DOI: [10.21046/20DZZconf-2022a](https://doi.org/10.21046/20DZZconf-2022a).
- Полищук В. Ю., Муратов И. Н., Куприянов М. А. *и др.* Моделирование полей термокарстовых озёр в зоне вечной мерзлоты на основе геоимитационного подхода и спутниковых снимков // Математические заметки СВФУ. — 2020. — Т. 27, № 1. — С. 101–114. — DOI: [10.25587/SVFU.2020.75.78.007](https://doi.org/10.25587/SVFU.2020.75.78.007).
- Полищук В. Ю., Муратов И. Н. *и* Полищук Ю. М. Проблемы моделирования пространственной структуры полей термокарстовых озёр в зоне вечной мерзлоты на основе спутниковых снимков // Вестник Югорского государственного университета. — 2018. — Т. 3, № 50. — С. 88–100. — DOI: [10.17816/byusu2018088-100](https://doi.org/10.17816/byusu2018088-100).
- Попков А. Ю. Рандомизированное машинное обучение нелинейных моделей с применением к прогнозированию развития эпидемического процесса // Автоматика и телемеханика. — 2021. — № 6. — С. 149–168. — DOI: [10.31857/S0005231021060064](https://doi.org/10.31857/S0005231021060064).
- Попков Ю. С. Рандомизация и энтропия в обработке данных, динамических системах, машинном обучении. — Москва : ЛЕНАНД, 2023.
- Попков Ю. С. *и* Дубнов Ю. А. Энтропийно-робастное рандомизированное прогнозирование при малых объемах ретроспективных данных // Автоматика и телемеханика. — 2016. — № 5. — С. 109–127.
- Попков Ю. С., Дубнов Ю. А. *и* Попков А. Ю. Прогнозирование развития эпидемии COVID-19 в странах Европейского союза с использованием энтропийно-рандомизированного подхода // Информатика и автоматизация. — 2021. — Т. 20, № 5. — С. 1010–1033. — DOI: [10.15622/20.5.1](https://doi.org/10.15622/20.5.1).
- Попков Ю. С., Попков А. Ю. *и* Дубнов Ю. А. Рандомизированное машинное обучение при ограниченных наборах данных: от эмпирической вероятности к энтропийной рандомизации. — Москва : ЛЕНАНД, 2019.
- Попков Ю. С., Попков А. Ю. *и* Дубнов Ю. А. Элементы рандомизированного прогнозирования и его применение для предсказания суточной электрической нагрузки энергетической системы // Автоматика и телемеханика. — 2020. — № 7. — С. 148–172. — DOI: [10.1134/S0005231019070107](https://doi.org/10.1134/S0005231019070107).
- Степаненко В. М., Мачульская Е. Е., Глаголев М. В. *и др.* Моделирование эмиссии метана из озёр зоны вечной мерзлоты // Известия РАН. Физика атмосферы и океана. — 2011. — Т. 47, № 2. — С. 275–288.
- Феллер В. Введение в теорию вероятностей и ее приложения. — Москва : Мир, 1967.
- Фельдман Г. М. Термокарст и вечная мерзлота. — Новосибирск : Наука, 1984.
- Golan A., Judge G. G. *and* Miller D. Maximum entropy econometrics. Robust estimation with limited data. — New York : Wiley, 1996. — 307 p.
- Jaynes E. T. Information theory and statistical mechanics // Physical review. — 1957. — Vol. 106, no. 4. — P. 620–630.
- Kapur J. N. Maximum-entropy models in science and engineering. — John Wiley & Sons, 1989.
- Karlsson J., Lyon S. *and* Destouni G. Temporal Behavior of Lake Size-Distribution in a Thawing Permafrost Landscape in Northwestern Siberia // Remote Sensing. — 2014. — Vol. 6, no. 1. — P. 621–636. — DOI: [10.3390/rs6010621](https://doi.org/10.3390/rs6010621).
- Kirpotin S. N., Polishchuk Y. *and* Bryksina N. Abrupt changes of thermokarst lakes in Western Siberia: impacts of climatic warming on permafrost melting // International Journal of Environmental Studies. — 2009. — Vol. 66, no. 4. — P. 423–431. — DOI: [10.1080/00207230902758287](https://doi.org/10.1080/00207230902758287).
- Kullback S. *and* Leibler R. A. On information and sufficiency // The annals of mathematical statistics. — 1951. — Vol. 22, no. 1. — P. 79–86.
- Levine R. D. *and* Tribus M. Maximum entropy formalism // Maximum Entropy Formalism Conference. — Massachusetts Institute of Technology, 1978.
- Neumann J. von. Various Techniques Used in Connection With Random Digits // Journal of Research of the National Bureau of Standards, Appl. Math. Series. — 1951. — Vol. 12. — P. 36–38.
- Nocedal J. *and* Wright S. Numerical optimization. — Springer Science & Business Media, 2006.

- Popkov Yu. S., Dubnov Yu. *and* Popkov A. New Method of Randomized Forecasting Using Entropy-Robust Estimation: Application to the World Population Prediction // Mathematics. — 2016. — Vol. 4, no. 1. — P. 16. — DOI: [10.3390/math4010016](https://doi.org/10.3390/math4010016).
- Popkov Yu. S., Popkov A. Yu., Dubnov Yu. A., *et al.* Entropy-Randomized Forecasting of Stochastic Dynamic Regression Models // Mathematics. — 2020. — Vol. 8, no. 7. — P. 1119. — DOI: [10.3390/math8071119](https://doi.org/10.3390/math8071119).
- Popkov Yu. S., Popkov A. Yu., Dubnov Yu. A., *et al.* Entropy Randomization in Machine Learning. — Chapman, Hall/CRC, 2022. — DOI: [10.1201/9781003306566](https://doi.org/10.1201/9781003306566).
- Sudakov I. *and* Vakulenko S. A. A mathematical model for a positive permafrost carbon-climate feedback // IMA Journal of Applied Mathematics. — 2014. — Vol. 80, no. 3. — P. 811–824. — DOI: [10.1093/imamat/hxu010](https://doi.org/10.1093/imamat/hxu010).
- Verpoorter Ch., Kutser T., Seekell D. A., *et al.* A global inventory of lakes based on high-resolution satellite imagery // Geophysical Research Letters. — 2014. — Vol. 41, no. 18. — P. 6396–6402. — DOI: [10.1002/2014GL060641](https://doi.org/10.1002/2014GL060641).
- Walter K. M., Smith L. C. *and* Stuart Chapin F. Methane bubbling from northern lakes: present and future contributions to the global methane budget // Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences. — 2007. — Vol. 365, no. 1856. — P. 1657–1676. — DOI: [10.1098/rsta.2007.2036](https://doi.org/10.1098/rsta.2007.2036).
- Zabelina S. A., Shirokova L. S., Klimov S. I., *et al.* Carbon emission from thermokarst lakes in NE European tundra // Limnology and Oceanography. — 2020. — Vol. 66, S1. — DOI: [10.1002/lno.11560](https://doi.org/10.1002/lno.11560).

SPATIOTEMPORAL ANALYSIS OF THE TERMOKARST LAKES EVOLUTION USING REMOTE SENSING AND RANDOMIZED MACHINE LEARNING AND FORECASTING

Yu. A. Dubnov^{*,1} , A. Yu. Popkov¹ , Yu. S. Popkov^{1,2} , M. A. Kupriyanov³ ,
V. Yu. Polishchuk⁴ , A. V. Melnikov³ , and Yu. M. Polishchuk³ 

¹Federal Research Center “Computer Science and Control” of RAS, Moscow, Russia

²Institute of Control Problems of RAS, Moscow, Russia

³Ugra Research Institute of Information Technologies, Khanty-Mansiysk, Russia

⁴Institute of Monitoring of Climatic and Ecological System of SB RAS, Tomsk, Russia

****Correspondence to:** Alexey Popkov, apopkov@isa.ru, alexeypopkov@yandex.ru

The work is devoted to the problem of forecasting the evolution of the area of thermokarst lakes in the Arctic permafrost zone using the analysis of test areas from several geographical regions. The approach proposed in the work is based on the Randomized Machine Learning method for constructing mathematical models of lake area from climate indicators, learning the model on real data and further forecasting. The results of modeling the dynamics of lake areas using linear static and dynamic models are presented and compared. It is shown that proposed dynamic model can significantly reduce the average modeling error.

Keywords: thermokarst lakes, information entropy, randomized machine learning, static and dynamic models, missing data, randomized forecasting, climate change.

Citation: Dubnov, Yu. A., A. Yu. Popkov, Yu. S. Popkov, M. A. Kupriyanov, V. Yu. Polishchuk, A. V. Melnikov, and Yu. M. Polishchuk (2024), Spatiotemporal Analysis of the Termokarst Lakes Evolution Using Remote Sensing and Randomized Machine Learning and Forecasting, *Russian Journal of Earth Sciences*, 24, ES5011, <https://doi.org/10.2205/2024ES000935>, EDN: TGQMBR

Received: 15 February 2024
Accepted: 30 August 2024
Published: 30 December 2024



© 2024. The Authors.

References

- Dubnov Yu. A. Entropy-Based Estimation in Classification Problems // *Avtomatika i Telemekhanika*. — 2019. — No. 3. — P. 138–151. — DOI: [10.1134/S0005231019030097](https://doi.org/10.1134/S0005231019030097).
- Dubnov Yu. A. and Boulytchev A. V. Bayesian Identification of a Gaussian Mixture Model // *Information technology and computing systems*. — 2017. — No. 1. — P. 101–114.
- Dubnov Yu. A., Polishchuk V. Yu., Popkov A. Yu., *et al.* Entropine-randomized forecasting of the evolution of the area of thermokarst lakes // *Chelyabinsk Physics and Mathematics Journal*. — 2021a. — Vol. 6, no. 3. — P. 384–396. — DOI: [10.47475/2500-0101-2021-16312](https://doi.org/10.47475/2500-0101-2021-16312).
- Dubnov Yu. A., Polishchuk V. Yu., Popkov Yu. S., *et al.* Entropy-Randomized Missing Date Imputation Method // *Avtomatika i Telemekhanika*. — 2021b. — P. 140–160. — DOI: [10.31857/S0005231021040061](https://doi.org/10.31857/S0005231021040061).
- Dubnov Yu. A., Popkov A. Yu., Polishchuk V. Yu., *et al.* Randomized machine learning algorithms to forecast the evolution of thermokarst lakes area in permafrost zones // *Automation and Telemekhanics*. — 2023. — No. 1. — DOI: [10.31857/S0005231023010051](https://doi.org/10.31857/S0005231023010051).
- Feldman G. M. Thermokarst and permafrost. — Novosibirsk : Nauka, 1984.
- Feller V. Introduction to Probability Theory and Its Applications. — Moscow : Mir, 1967.
- Golan A., Judge G. G. and Miller D. Maximum entropy econometrics. Robust estimation with limited data. — New York : Wiley, 1996. — 307 p.
- Jaynes E. T. Information theory and statistical mechanics // *Physical review*. — 1957. — Vol. 106, no. 4. — P. 620–630.
- Kapur J. N. Maximum-entropy models in science and engineering. — John Wiley & Sons, 1989.
- Karlsson J., Lyon S. and Destouni G. Temporal Behavior of Lake Size-Distribution in a Thawing Permafrost Landscape in Northwestern Siberia // *Remote Sensing*. — 2014. — Vol. 6, no. 1. — P. 621–636. — DOI: [10.3390/rs6010621](https://doi.org/10.3390/rs6010621).

- Kirpotin S. N., Polishchuk Y. and Bryksina N. Abrupt changes of thermokarst lakes in Western Siberia: impacts of climatic warming on permafrost melting // *International Journal of Environmental Studies*. — 2009. — Vol. 66, no. 4. — P. 423–431. — DOI: [10.1080/00207230902758287](https://doi.org/10.1080/00207230902758287).
- Kullback S. and Leibler R. A. On information and sufficiency // *The annals of mathematical statistics*. — 1951. — Vol. 22, no. 1. — P. 79–86.
- Levine R. D. and Tribus M. Maximum entropy formalism // *Maximum Entropy Formalism Conference*. — Massachusetts Institute of Technology, 1978.
- Mironov M. S., Shornikova A. V., Sidorina I. E., *et al.* Study of Methane Emissions in Thermokarst Lakes of the Yamal Peninsula Using Remote Sensing and Ground-Based Research // *Proceedings of the 20th International Conference «Modern Problems of Remote Sensing of the Earth from Space»*. — Moscow : Space Research Institute of the Russian Academy of Sciences, 2022. — DOI: [10.21046/20DZZconf-2022a](https://doi.org/10.21046/20DZZconf-2022a).
- Neumann J. von. Various Techniques Used in Connection With Random Digits // *Journal of Research of the National Bureau of Standards, Appl. Math. Series*. — 1951. — Vol. 12. — P. 36–38.
- Nocedal J. and Wright S. Numerical optimization. — Springer Science & Business Media, 2006.
- Polishchuk V. Yu., Muratov I. N., Kupriyanov M. A., *et al.* Modeling the fields of thermokarst lakes in the permafrost based on the geo-simulation approach and satellite images // *Mathematical notes of NEFU*. — 2020. — Vol. 27, no. 1. — P. 101–114. — DOI: [10.25587/SVFU.2020.75.78.007](https://doi.org/10.25587/SVFU.2020.75.78.007).
- Polishchuk V. Yu., Muratov I. N. and Polishchuk Yu. M. Problems of modeling the spatial structure of thermokarst lake fields in the permafrost based on satellite imagery // *Bulletin of the Yugra State University*. — 2018. — Vol. 3, no. 50. — P. 88–100. — DOI: [10.17816/byusu2018088-100](https://doi.org/10.17816/byusu2018088-100).
- Popkov A. Yu. Randomized machine learning of nonlinear models with application to forecasting the development of an epidemic process // *Avtomatika i Telemekhanika*. — 2021. — No. 6. — P. 149–168. — DOI: [10.31857/S0005231021060064](https://doi.org/10.31857/S0005231021060064).
- Popkov Yu. S. Randomization and Entropy in Data Processing, Dynamic Systems, Machine Learning. — Moscow : LENAND, 2023.
- Popkov Yu. S., Dubnov Yu. and Popkov A. New Method of Randomized Forecasting Using Entropy-Robust Estimation: Application to the World Population Prediction // *Mathematics*. — 2016. — Vol. 4, no. 1. — P. 16. — DOI: [10.3390/math4010016](https://doi.org/10.3390/math4010016).
- Popkov Yu. S., Dubnov Yu. and Popkov A. Forecasting Development of COVID-19 Epidemic in European Union Using Entropy-Randomized Approach // *Informatika i Avtomatizatsiya*. — 2021. — Vol. 20, no. 5. — P. 1010–1033. — DOI: [10.15622/20.5.1](https://doi.org/10.15622/20.5.1).
- Popkov Yu. S. and Dubnov Yu. A. Entropy-robust randomized forecasting under small sets of retrospective data // *Avtomatika i telemekhanika*. — 2016. — No. 5. — P. 109–127.
- Popkov Yu. S., Popkov A. Yu. and Dubnov Yu. A. Randomized Machine Learning with Limited Datasets: From Empirical Probability to Entropic Randomization. — Moscow : LENAND, 2019.
- Popkov Yu. S., Popkov A. Yu. and Dubnov Yu. A. Elements of randomized forecasting and its application to daily electrical load prediction in a regional power system // *Avtomatika i Telemekhanika*. — 2020a. — No. 7. — P. 148–172. — DOI: [10.1134/S0005231019070107](https://doi.org/10.1134/S0005231019070107).
- Popkov Yu. S., Popkov A. Yu., Dubnov Yu. A., *et al.* Entropy-Randomized Forecasting of Stochastic Dynamic Regression Models // *Mathematics*. — 2020b. — Vol. 8, no. 7. — P. 1119. — DOI: [10.3390/math8071119](https://doi.org/10.3390/math8071119).
- Popkov Yu. S., Popkov A. Yu., Dubnov Yu. A., *et al.* Entropy Randomization in Machine Learning. — Chapman, Hall/CRC, 2022. — DOI: [10.1201/9781003306566](https://doi.org/10.1201/9781003306566).
- Stepanenko V. M., Machul'skaya E. E., Glagolev M. V., *et al.* Numerical Modeling of Methane Emissions from Lakes in the Permafrost Zone // *Izvestiya RAS. Physics of the Atmosphere and Ocean*. — 2011. — Vol. 47, no. 2. — P. 275–288.
- Sudakov I. and Vakulenko S. A. A mathematical model for a positive permafrost carbon-climate feedback // *IMA Journal of Applied Mathematics*. — 2014. — Vol. 80, no. 3. — P. 811–824. — DOI: [10.1093/imat/hxu010](https://doi.org/10.1093/imat/hxu010).
- Verpoorter Ch., Kutser T., Seekell D. A., *et al.* A global inventory of lakes based on high-resolution satellite imagery // *Geophysical Research Letters*. — 2014. — Vol. 41, no. 18. — P. 6396–6402. — DOI: [10.1002/2014GL060641](https://doi.org/10.1002/2014GL060641).
- Walter K. M., Smith L. C. and Stuart Chapin F. Methane bubbling from northern lakes: present and future contributions to the global methane budget // *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*. — 2007. — Vol. 365, no. 1856. — P. 1657–1676. — DOI: [10.1098/rsta.2007.2036](https://doi.org/10.1098/rsta.2007.2036).
- Zabelina S. A., Shirokova L. S., Klimov S. I., *et al.* Carbon emission from thermokarst lakes in NE European tundra // *Limnology and Oceanography*. — 2020. — Vol. 66, S1. — DOI: [10.1002/lno.11560](https://doi.org/10.1002/lno.11560).