МЕТОДЫ, МОДЕЛИ И ТЕХНОЛОГИИ
METHODS, MODELS AND TECHNOLOGIES

УДК 556.16:504.064.2.001.18 DOI: 10.34753/HS.2023.5.1.59

ВЫСОКОЧАСТОТНЫЙ МОНИТОРИНГ И МОДЕЛИ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ОЦЕНКИ СИНОПТИЧЕСКОЙ ИЗМЕНЧИВОСТИ СТОКА ВЗВЕШЕННЫХ НАНОСОВ МАЛОЙ ГОРОДСКОЙ РЕКИ

С. Р. Чалов[1,2], В. М. Морейдо[2], И. С. Денисова[1,2], И. А. Солоников[1]
[1]МГУ имени М.В. Ломоносова, Географический факультет, Москва, Россия
[2]Институт водных проблем РАН, Москва, Россия

ira.denisova@icloud.com

Аннотация. В статье приводятся результаты высокочастотного мониторинга стока наносов крупнейшего правого притока р. Москвы в пределах города – р. Сетуни, осуществлявшегося в 2020–2021 гг. автоматическими регистраторами оптической мутности и уровня воды. Задействованы 2 ряда записи с 16 ноября 2019 г. по 19 марта 2020 г. (частота записи 10 минут) и с 1 марта 2021 г. по апрель 2021 г. и с июля 2021 г. по октябрь 2021 г. (частота записи 1 минута). В условиях высокой антропогенной нагрузки средняя мутность воды р. Сетунь достигает 100 мг/л, что в 3–4 раза выше фоновых значений. Отмечается корреляция мутности с расходами воды, которая возрастает при увеличении лага по времени от 0 до 50 часов. На основе совмещения измерений мутности воды с характеристиками уровня и расхода воды, атмосферных осадков и температуры воздуха, а также производных от них, проведена настройка 5 моделей машинного обучения для прогнозирования стока наносов с разрешением 1 час и 1 сутки. Наилучшие результаты продемонстрировала рекуррентная нейросеть LSTM RNN для прогноза суточных значений мутности воды: при среднеквадратической ошибке 10,8 NTU были

HIGH-FREQUENCY MONITORING AND MACHINE LEARNING MODELS FOR ASSESSING SYNOPTIC VARIABILITY OF SUSPENDED SEDIMENT FLOW OF A SMALL URBAN RIVER

Sergey R. Chalov[1,2], Vsevolod M. Moreido[2], Irina S. Denisova[1,2], Ivan A. Solonikov[1]
[1]Lomonosov Moscow State University, Faculty of Geography, Moscow, Russian Federation
[2]Water Problems Institute of the Russian Academy of Sciences, Moscow, Russian Federation
ira.denisova@icloud.com

Abstract. The article presents the results of high-frequency monitoring of sediment flow of the largest right tributary of the Moscow River within the city – the Setun River, carried out in 2020–2021 by automatic recorders of optical turbidity and water level. Several recordings were involved from November 16, 2019 to March 19, 2020 (recording frequency 10 minutes) and from March 1, 2021 to April 2021 and from July 2021 to October 2021 (recording frequency 1 minute). Under conditions of high anthropogenic load, the average turbidity of the Setun River water reaches 100 mg/l, which is 3–4 times higher than background values. There is a correlation between turbidity and water flow rates, which increases with increasing time lag from 0 to 50 hours. Based on combining measurements of water turbidity with characteristics of water level and flow, precipitation, and air temperature, as well as derivatives from them, 5 machine learning models were configured to predict sediment runoff with a resolution of 1 hour and 1 day. The best results were demonstrated by the recurrent neural network LSTM RNN for forecasting daily values of water turbidity: with a root-mean-square error of 10.8 NTU, the time and magnitude of local

С. Р. Чалов, В. М. Морейдо, И. С. Денисова, И. А. Солоников. Высокочастотный мониторинг и модели машинного обучения для оценки синоптической изменчивости стока взвешенных наносов малой городской реки // Гидросфера. Опасные процессы и явления. 2023. Т. 5. Вып. 1. С. 59–74. DOI: 10.34753/HS.2023.5.1.59

верно определены время и величина локальных увеличений пика мутности.

Ключевые слова: городская гидрология; сток взвешенных наносов; мутность воды; модели машинного обучения.

increases in the turbidity peak were correctly determined.

Keywords: urban hydrology; suspended sediment flow; water turbidity; machine learning models.

Введение

Внутригодовое изменение водности рек обусловливает сезонную цикличность стока наносов. На малых реках, помимо режима выпадения осадков и водности, характеристики стока наносов в значительной степени определяются свойствами поверхности водосбора. Здесь существенное влияние оказывают физические характеристики почво-грунтов, режим их промерзания и увлажнения, состав и распространение растительности, режим залегания и стаивания снега на водосборе и т.п. На урбанизированных территориях в условиях сокращенного времени добегания дождевого стока исключительную выраженность приобретают суточные колебания, как правило, связанные с отдельными гидрологическими событиями (периодами повышения расходов воды в реке, вызванных погодными (синоптическими) факторами). В этих условиях характеристика режима стока наносов возможна исключительно на основе высокочастотных наблюдений, которые применительно к взвешенными наносам обеспечиваются автоматическими датчиками мутности. Подобные технологии позволяют получать принципиально новые данные о стоке наносов и, соответственно, применять к ним современные методы статистического анализа и математического моделирования. К последним относятся модели машинного обучения, которые в последние десятилетия стали активно внедряться в гидрологическую практику [Bhattacharya, Solomatine, 2006; Solomatine, Ostfeld, 2008; Moreido et al., 2021].

Применительно к стоку взвешенных наносов методы машинного обучения позволяли разработать прогнозные модели, основанные на учете метеорологических, гидрологических и геолого-геоморфологических факторов рек разного размера – от малых рек Пуэрто-Рико до р. Миссисипи [Sharafati et al., 2020; Zounemat-Kermani et al., 2020]. Ряд работ продемонстрировал перспективность применения методов машинного обучения для описания рядов стока наносов урбанизированных территорий [Behrouz, Yazdi, Sample, 2022]. По сравнению с известными эмпирическими моделями, описывающими сток наносов [Cohen, Kettner, Syvitski, 2014], методы машинного обучения значительно расширяют спектр учитываемых параметров и воспроизводимость мутности воды и расхода наносов. В России методы автоматизированного высокочастотного мониторинга и машинного обучения применялись только для изучения уровня воды. В литературе приводятся примеры успешного воспроизведения паводков октября 2018 г. на р. Мзымта и Пшиш Краснодарского края, где моделирование уровней воды с разной заблаговременностью (1–20 часов) было выполнено с использованием трёх моделей машинного обучения: ансамблевых моделей M5P и XGBoost на основе деревьев решений и нейросетевой модели MLP [Белякова и др., 2022]. Однако в целом внедрение подобных технологий в России до настоящего времени развития не получило в связи с недостатком рядов мониторинга.

Данная работа направлена на анализ результатов комплексного мониторинга стока воды и наносов, ведущегося на р. Сетунь в Москве с 2019 г. Здесь на основе автоматической регистрации уровней воды с помощью логгеров (самописцев) гидростатического уровня и оптической мутности воды собран уникальный ряд высокочастотных (частотой 1 измерение от 1 до 30 минут) значений уровня и мутности воды. На территории бассейна реки расположена метеорологическая обсерватория МГУ, проводящая полный комплекс наблюдений за осадками. В совокупности этот массив данных определил возможности комплексного

Chalov S.R., Vsevolod M. Moreido V.M., Denisova I.S., Solonikov I.A. High-frequency monitoring and machine learning models for assessing synoptic variability of suspended sediment flow of a small urban river. *Hydrosphere. Hazard processes and phenomena*, 2023, vol. 5, iss. 1, pp. 59–74. (In Russian; abstract in English). DOI: 10.34753/HS.2023.5.1.59.

исследования реакции стока наносов урбанизированного бассейна на ход метеорологических и гидрологических факторов, что и являлось целью данной статьи. Задачами работы являлось статистическое описание режима мутности воды на основе высокочастотного мониторинга, адаптация нескольких алгоритмов машинного обучения для анализа стока взвешенных наносов и проведение тестовых оценок их применимости.

Материалы и методы

Бассейн р. Сетунь

Сетунь – крупнейший правый приток р. Москвы в её верхнем и среднем течении. Общая длина – 38 км, площадь водосбора – 190 км2. Водосбор примерно в равных частях разделен Московской кольцевой автомобильной дорогой (МКАД), пересекающей реку в 18 км от истока. Еще в 50-е гг. река полностью протекала по неурбанизированной территории и пересекала только Аминьевское и Староволынское шоссе и Киевскую железную дорог, однако в последнее десятилетие изменение границ природоохранных территорий значительно усилило массовое возведение капитальных строений и жилого фонда. В этих условиях активизировалась трансформация характеристик речного стока, в том числе стока наносов. Его основными факторами стало усиление сбросов сточных вод (в том числе со снегоплавильных станций в зимний период), 2–3 кратное увеличение времени добегания дождевого стока с водосбора [Chalov et al., 2023]. Значительная часть стока зарегулирована прудами, которые снижают максимальные значения расходов весеннего половодья и дождевых паводков, занимая на водосборе реки около одного процента территории [Черных, Сабитов, Алтунин, 2015].

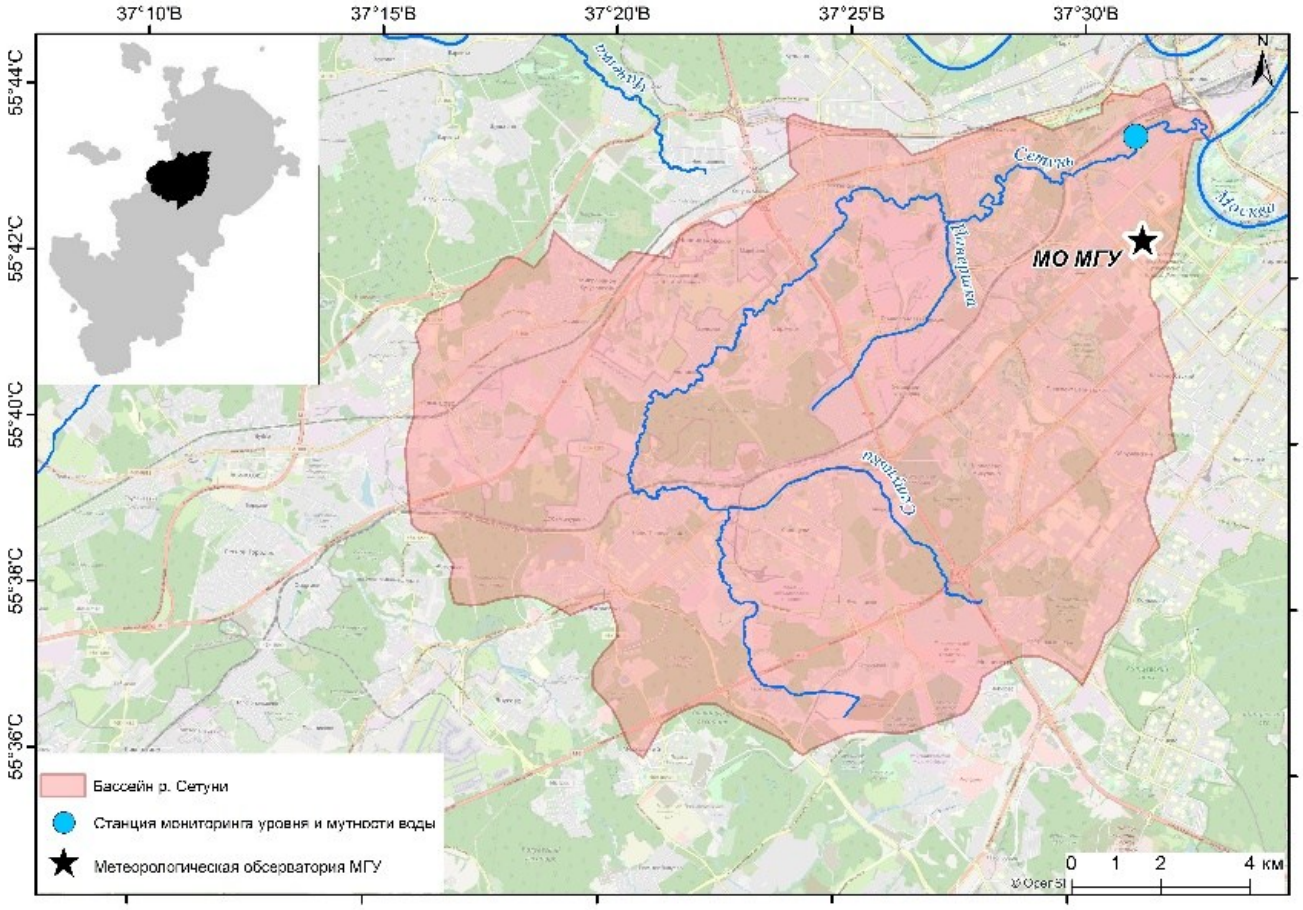


Рисунок 1. Станция мониторинга в бассейне р. Сетуни
Figure 1. Monitoring station in the Setun River basin

Мониторинговые данные

Данные, анализируемые в работе, получены по замыкающему створу, расположенному в нижнем течении р. Сетунь (рисунок 1). Здесь 14 ноября 2019 г. была организована станция мониторинга уровня и мутности воды, на которой с помощью барометрического логгера ведется запись значений уровня воды с интервалом 30 минут. Используется логгер HOBO Water Level (30 ft) Data Logger, который фиксирует суммарное гидростатическое давление атмосферы и толщи воды над гидростатическим датчиком давления прибора; запись ведется на внутренний накопитель. Барометрическая компенсация проводится по данным атмосферного давления метеостанции МГУ.

Для перехода от значений уровня к значениям расходов воды на данном створе наблюдений были организованы регулярные измерения расходов воды, которые позволили получить зависимости Q=f(H) в виде 1:

$$Q = \begin{cases} 8.935 \times H^2 - 4{,}056 \times H, \text{если } H < 1{,}07 \\ 15 \times H - 10{,}15, \text{если } H > 1{,}07 \end{cases} \quad (1)$$

где Q – расход воды (м3/с), H – уровень воды (м).

Уровни воды записывались автоматически с частотой 30 минут, таким образом полученный ряд расходов воды также имеет получасовое временное разрешение.

Данные о мутности воды записывались автоматическими логгерами, основанными на нефелометрическом принципе, (таблица 1) на протяжении двух периодов:

а) логгер Analite NEP-495, период с 16 ноября 2019 г. по 19 марта 2020 г., частота записи 10 минут (далее ряд 1);

б) логгер In-Situ Aqua Troll 600, период с 1 марта 2021 г. по апрель 2021 г. и с июля 2021 г. по октябрь 2021 г. частота записи 1 минута (далее ряд 2).

Нефелометрический метод измерения мутности речных вод основан на способности взвешенных частиц рассеивать световое излучение, проходящее через единицу объема воды: чем сильнее затухание светового потока с известными характеристиками, которое регистрирует один или несколько датчиков излучения, тем больше содержание в воде взвешенных частиц. Данный принцип используется большинством современных датчиков in situ. Приёмные датчики погружают в водный объект, их детекторы расположены под различными углами к источнику излучения, что позволяет определить степень рассеяния исходного пучка света. Регистрируемое на датчиках излучения изменение светового потока калибруется относительно стандартных растворов, что позволяет получить универсальное уравнение преобразования для достаточно широкого диапазона изменения мутности (от 0 до 4000 NTU [Droujko, Molnar, 2022]). Преимуществами такого метода являются: быстрое получение величины мутности, мобильность, возможность проведения непрерывного мониторинга мутности с высокой частотой измерений [Промахова, Чалов, 2015].

Таблица 1. Технические характеристики моделей регистраторов мутности.
Table 1. Technical characteristics of turbidity logger models.

Обозначение	Ряд 1	Ряд 2
Название модели	ANALITE NEP-495	In-Situ Aqua TROLL 600
Измеряемые параметры	Оптическая мутность, температура воды	Оптическая мутность, температура воды
Диапазон измеряемой мутности, NTU[1]	0 – 5000	0 – 4000
Погрешность определения	±1%	±2%
Частота записи использованная (возможная)	10 минут (1 сек – 18 ч)	1 минута (1 мин – 99 ч)
Период записи	16.11.2019 –19.03.2020	01.03.2021 – 29.10.2021

[1]NTU - нефелометрические единицы измерения оптической мутности потока (nephelometric turbidity units)

Подготовка данных производилась в среде RStudio при помощи средств языка программирования R. Первичное сглаживание данных производилось с помощью фильтра

Савитцки-Голея [Press, Teukolsky, 1990]. Затем из рядов были удалены бракованные участки записей, в которых значения мутности резко аномально повышались. Основная причина появления таких участков – заиление датчиков, при котором на чувствительных элементах прибора накапливались глинисто-илистые частицы, несмотря на проводившуюся периодическую очистку. Таким образом были получены основные характеристики высокочастотных рядов наблюдений, поделенные на отрезки с удаленными отбракованными участками (таблица 2).

Таблица 2. Основные характеристики высокочастотных рядов наблюдений.
Table 2. Main characteristics of high-frequency observation series.

Период	**2019–2020 часть 1**	**2019–2020 часть 2**	**2019–2020 полный**	**2021 часть 1**	**2021 часть 2**	**2021 полный**
Начало периода	05.12.19 9:00	31.12.19 7:00	05.12.19 9:00	09.03.21 11:45	22.07.21 14:41	01.03.21 10:00
Конец периода	24.12.19 19:00	29.01.20 6:50	29.01.20 6:50	03.05.21 22:30	28.10.21 7:53	28.10.21 7:53
Продолжительность, сут.	19 (каждые 10 минут)	29 (каждые 10 минут)	48 (каждые 10 минут)	51 (каждую 1 минуту)	93 (каждую 1 минуту)	144 (каждую 1 минуту)
Средняя мутность, NTU	54,09	26,37	37,49	80,99	31,61	47,19
СКО[2] мутности, NTU	14,15	14,83	19,92	108,28	60,24	81,81
Минимальная мутность, NTU	36,11	2,33	2,33	0,00	0,01	0,00
Максимальная мутность, NTU	90,37	81,03	90,37	1692,54	1359,99	1692,54
Средний расход воды, $м^3/с$	2,51	2,39	2,44	3,20	2,03	2,51
СКО расхода воды, $м^3/с$	0,32	0,30	0,31	1,77	1,30	1,93
Минимальный расход воды, $м^3/с$	2,07	2,05	2,05	1,44	0,93	0,93
Максимальный расход воды, $м^3/с$	5,10	4,62	5,10	11,74	16,82	30,76

[2]СКО - среднеквадратическое отклонение

Бракованные участки рядов характеризовались крайне высокой дисперсией значений, отчётливо выраженной на графиках хода мутности, что позволяло отличить бракованный участок с аномально высокой мутностью от пиков мутности, связанных с увеличением расходов воды (например, во время паводков).

Использованные в работе данные метеорологической обсерватории МГУ представляли собой срочные наблюдения за температурой воздуха и количеством осадков с января 2019 г. по февраль 2023 г. Температура воздуха измерялась термометром ТМ-4 («психрометром Августа») в метеорологической будке каждые 3 часа, количество осадков – осадкомером Третьякова О-1 при наличии осадков, 4 раза в сутки: в 6, 9, 18 и 21 час. Обработка данных о количестве осадков производилась в связи с характером данных следующим образом:

1) Значения в ряде данных для дней, когда осадки не наблюдались, были заполнены нулями;

2) Ряды данных были внутри каждых суток поделены на группы (для дней с осадками). Каждая группа включала снятое по осадкомеру значение суммы осадков за наблюдённый период и пустые строки выше вплоть до начала суток или до предыдущего измерения внутри данных суток;

3) Накопленные суммы осадков внутри каждой группы данных равномерно распределялись по этой группе. Так как ряд имеет

временное разрешение 3 часа, для каждой строки ряда было получено приближённое значение суммы осадков за 3 часа;

4) Полученные значения сумм осадков были поделены на 3, в результате были получены средние интенсивности осадков в мм/час.

Поскольку временно́е разрешение метеорологических данных значительно ниже разрешения данных о мутности и расходах воды, было принято решение привести данные о температуре воздуха и осадках к 1-часовой временной сетке. Также были подготовлены суточные ряды метеорологических данных, содержащие среднюю температуру и суточные суммы осадков.

Данные о мутности, расходах воды, температуре и осадках были собраны в единые наборы данных отдельно для двух рядов записей мутности (с логгеров Analite и In-Situ, соответственно). Так как данные о мутности и расходах воды имеют более высокое временно́е разрешение, значения температуры воздуха были линейно интерполированы для приведения к новой временно́й сетке. Значения интенсивностей осадков же, напротив, не могли быть линейно интерполированы в силу характера выпадения осадков, поэтому для каждой строки ряда данных была принята соответствующая ей среднечасовая величина интенсивности осадков. Ряд осадков, таким образом, получился ступенчатым, и для компенсации резких изменений значений он был сглажен скользящим средним с окном в 15 соседних значений.

Единичные пропуски данных наблюдений в рядах были заполнены с помощью линейной интерполяции. Значительные временны́е разрывы в рядах данных не могли быть каким-либо обоснованным образом заполнены, поэтому последующий анализ рядов проводился как для полных рядов, содержащих временные разрывы, так и для отдельных их частей, где такие разрывы отсутствовали. Для получения рядов с суточным разрешением величины мутности и расходов воды пересчитывались на новую временну́ю сетку, их значения осреднялись, после чего объединялись с суточными рядами метеоданных. Пример графика совмещённого хода мутности, расхода воды и осадков, построенного по объединённому суточному ряду, представлен на рисунке 2.

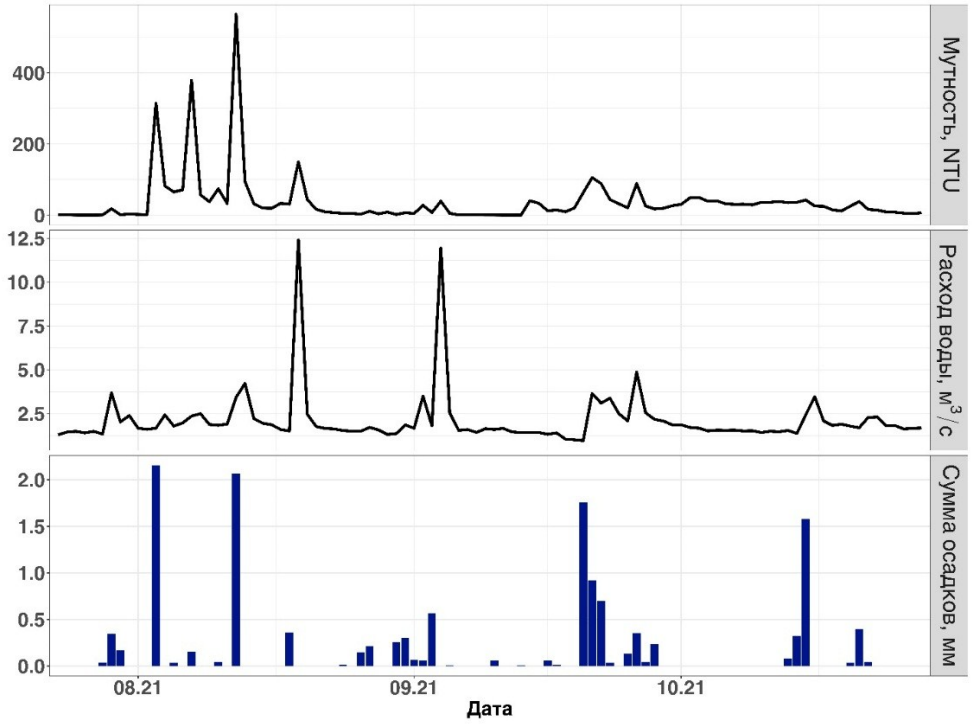


Рисунок 2. Ход мутности, расхода воды и осадков для р. Сетуни по суточным данным за лето–осень 2021 г.

Figure 2. Progress in turbidity, water flow and precipitation for the Setun River according to daily data for summer–autumn period, 2021.

Для общей характеристики стока наносов определялась средняя мутность по формуле 2:

$$\overline{x} = \frac{1}{n} \sum_i^n x_i \quad (2)$$

где $\overline{x}$ – среднее значение мутности, n – число измерений, x_i – i-е измерение мутности.

Стандартное отклонение определялось по формуле 3:

$$S_x = \sqrt{\frac{\sum_i^n (x_i - \overline{x})^2}{n-1}} \quad (3)$$

где $\bar{x}$ – среднее значение мутности, n – число измерений, x_i – i-е измерение мутности.

Разброс величины мутности был рассчитан как разность максимального и минимального значений всего ряда по формуле 4:

$$Sp_x = \max(x_i) - \min(x_i) \quad (4)$$

Коэффициент вариации мутности определялся по формуле 5:

$$c_v = \frac{S_x}{x} \times 100\% \quad (5)$$

Дополнительно по методике [Чалов, Цыпленков, 2020] рассчитывался индекс интенсивности турбулентных колебаний как среднее значение разности между соседними измерениями мутности воды по всему ряду наблюдений по формуле 6:

$$T2 - T1 = \frac{1}{n-1} \sum_{i=1}^{n-1} (|x_{i+1} - x_i|) \quad (6)$$

где x_{i+1} и x_i – соседние измерения мутности.

Индекс *T2-T1/*Sp_x рассчитывался как отношение средней разности между соседними измерениями к разности максимального и минимального значения мутности по всему ряду наблюдений по формуле 7:

$$(T2 - T1)/Sp_x = \frac{T2 - T1}{\max(x_i) - \min(x_i)} \quad (7)$$

Данный индекс характеризует выраженность турбулентных колебаний мутности на фоне её общих флуктуаций и является мерой соотношения турбулентных и синоптических колебаний.

Описание использованных моделей

Для выявления закономерностей формирования стока взвешенных наносов и определения возможностей его количественного описания были использованы 5 моделей машинного обучения: ARIMA, Prophet, Random Forest Regressor, XGBRegressor, LSTM. Модель ARIMA [Box, Jenkins, 1970] является авторегрессионной моделью скользящего среднего для нестационарных временных рядов. Prophet [Taylor, Letham, 2018] – модель, реализованная в одноименной библиотеке, разработанной компанией Facebook, для прогнозирования временных рядов с открытым кодом на основе оптимизации аддитивных регрессионных моделей. Модель Random Forest Regressor [Pedregosa et al., 2011] из библиотеки Scikit-learn языка Python является ансамблевым алгоритмом на основе деревьев решений. Модель XGBRegressor [Chen, Guestrin, 2016] из библиотеки XGBoost языка Python также является ансамблевым алгоритмом на основе решающих деревьев, однако компенсация ошибок в процессе обучения модели в данном случае происходит не за счёт усреднения, а через обучение (модель градиентного бустинга). LSTM [Hochreiter, Schmidhuber, 1997; Kratzert et al., 2018] – модель, реализованная в библиотеке Keras, являющаяся разновидностью рекуррентных нейронных сетей и приспособленная к задачам прогнозирования временных рядов в случаях, когда важные события разделены временными лагами с неопределённой продолжительностью и границами.

Модели были использованы как для расчета ряда измеренных величин мутности, так и для прогнозирования её будущих значений на период до 3 недель. Для обучения и проверки моделей применялись данные о мутности воды с соответствующими предикторами за несколько периодов: с декабря 2019 г. по январь 2020 г., когда запись величин мутности велась логгером Analite; с марта 2021 г. по апрель 2021 г. и с июля 2021 г. по октябрь 2021 г., когда запись велась логгером

In-Situ; последний рассматривавшийся период представлял собой объединение обоих периодов наблюдений 2021 г. Поскольку ряды мутности и расходов имели различное временное разрешение (ряды мутности – 1 минута и 10 минут, ряды расходов – 30 минут), они были приведены к 1-часовому шагу путём осреднения. Таким образом, на основе исходных данных были созданы 2 типа рядов: с разрешением 1 час и с суточным разрешением (осредненные значения за 1 сутки).

В качестве предикторов модели использовались характеристики уровня и расхода воды, атмосферных осадков и температуры воздуха, а также их производных параметров (таблица 3). Список отобранных предикторов отличался между моделями, при этом основными предикторами выступали расход воды и суммы осадков за день и за неделю. Также в качестве предикторов обычно использовались температура воздуха и сдвинутые с определённым лагом (шагом во времени) величины расходов и сумм осадков. Измеренные температуры воздуха и воды имели высокую парную корреляцию, поэтому ради сокращения размерности данных и во избежание переобучения было принято решение объединить температуры воды и воздуха в один вектор-признак с помощью метода главных компонент (PCA).

Таблица 3. Перечень использованных для моделирования характеристик.
Table 3. List of the modelling dataset variables.

Обозначение	Характеристика
(1) NTU	мутность воды (в NTU);
(2) Q	расход воды (в $м^3/с$);
(3) RainIntensity	сглаженные значения интенсивности осадков (в мм/час);
(4) Hour	время суток (в часах);
(5) DayOfWeek	номер дня недели;
(6) DayOfYear	день года, нормализованный по шкале от 0 до 1;
(7) Month	номер месяца;
(8) QLag10	значение расхода воды, взятого с лагом в 10 наблюдений ($м^3/с$);
(9) QLag20	значение расхода воды, взятого с лагом в 20 наблюдений ($м^3/с$);
(10) QLag50	значение расхода воды, взятого с лагом в 50 наблюдений ($м^3/с$);
(11) QLag100	значение расхода воды, взятого с лагом в 100 наблюдений ($м^3/с$);
(12) RainIntensityLag10	сглаженные значения интенсивности осадков, взятые с лагом в 10 наблюдений (в мм/час);
(13) RainIntensityLag20	сглаженные значения интенсивности осадков, взятые с лагом в 20 наблюдений (в мм/час);
(14) RainIntensityLag50	сглаженные значения интенсивности осадков, взятые с лагом в 50 наблюдений (в мм/час);
(15) RainIntensityLag100	сглаженные значения интенсивности осадков, взятые с лагом в 100 наблюдений (в мм/час);
(16) AirTempSign	знак температуры воздуха;
(17) Temperature	объединённый показатель температуры воздуха и воды, полученный методом главных компонент (PCA);
(18) NTUdiff	разность между соседними измерениями мутности воды (в NTU);

Для подбора оптимальных параметров моделей использовались, при возможности, функции GridSearchCV и RandomizedSearchCV библиотеки Scikit-learn языка Python. Возможность количественной оценки мутности определялась как для её часовых значений, так и для суточных величин. Модели проверялись на способность оценивать мутность с использованием данных о её величинах за предшествующий период и без него. Метрикой эффективности модели служила среднеквадратическая ошибка (RMSE), вычисляемая по формуле 8:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2} \quad (8)$$

где n – длина ряда наблюдений, Y_i – фактические значения показателя, $\hat{Y}_i$ – оценки значений показателя, i – номер наблюдения.

Результаты и обсуждение

Средняя мутность воды р. Сетунь в мониторинговом створе за период наблюдений изменяется от 37,49 до 47,19 NTU (таблица 4), что с учетом региональной формулы 9 пересчета оптической мутности воды в весовую [Sokolov et al., 2020], полученной по результатам одновременного определения обеих характеристик в пробах воды, соответствует значениям весовой мутности от 31,39 до 39,15 мг/л:

$$SSC = 1{,}4 + 0{,}8 \times T \quad (9)$$

где SSC – весовая мутность (мг/л), T – оптическая мутность (NTU).

Максимальная мутность воды достигала 73,7 и 1355,4 мг/л для Ряда 1 и Ряда 2 соответственно (таблица 4). Это по крайней мере в 3–4 раза выше фоновых значений мутности воды, характерных для малых рек верхней части Волжского бассейна [Караушев, 1977]. Полученные в ходе настоящей работы ряды мутности р. Сетунь свидетельствуют о наличии 20-минутных макротурбулентных изменений мутности воды. Потоку характерна максимальная интенсивность пульсаций мутности по сравнению с реками других природных условий [Чалов, Цыпленков, 2020], проявляющаяся в усилении вклада макротурбулентных изменений мутности в синоптические колебания. Отмечена высокая автокорреляция (cor≥0,75) значений мутности воды в пределах 100 минут. Существование изменений мутности воды в таком масштабе времени хорошо согласуется с когерентными структурами в речном потоке [Carling, Orr, 2000], которые связаны с неровностями руслового рельефа. Размеры вихрей столь низкой частоты охватывают всё сечение реки, в результате чего пульсирующим оказывается и сам расход воды. В условиях урбанизированного бассейна на нестационарный режим переноса взвесей оказывает роль и наличие случайных коротких сбросов из дренажно-коллекторной сети, фиксируемых по всему бассейну. Наличие этого фактора играет важную роль в обосновании дискретного режима измерений с использованием автоматических оптических датчиков.

Таблица 4. Средние характеристики рядов мутности воды р. Сетунь.
Table 4. Average characteristics of water turbidity datasets of the Setun River.

№	Сред. мутность, NTU	Макс. мутность, NTU	Мин. мутность, NTU	Станд. отклонение	Разброс	Коэф-т вариации *cv*, %	*T2-T1*, NTU	*T2-T1/Sp*$_x$, NTU
Ряд 1	37,49	90,37	2,33	19,92	88,04	0,53	0,4	0,00448
Ряд 2	47,19	1692,54	0,004	81,81	1692,54	1,73	0,48	0,00028

Анализ матриц парной корреляции (рисунок 3) свидетельствует о преимущественном воздействии на сток взвешенных наносов атмосферных осадков и водности реки ($cor \geq 0{,}5$). При этом с увеличением лага времени по осадкам (от нулевого лага до лага в 100 наблюдений, что соответствует 100 часам) скоррелированность между мутностью воды и интенсивностью осадков увеличивается, а при аналогичном увеличении лага расходов воды – уменьшается. В частности, корреляция ряда мутности с одновременным расходом воды составляет 0,5, с расходом воды, взятым с лагом 10 измерений (10 часов) – 0,48, а с лагом 100 измерений (100 часов) – 0,38. Корреляция с одновременными осадками составляет 0,53, с лагом 50 часов – 0,56, 100 часов – 0,58. Это, вероятно, связано с характерным для р. Сетуни временем склонового и руслового добегания – около 6–8 часов для осадков часовой интенсивности и около 36 часов для их накопленных сумм, что было показано в работе [Чалов и др., 2023].

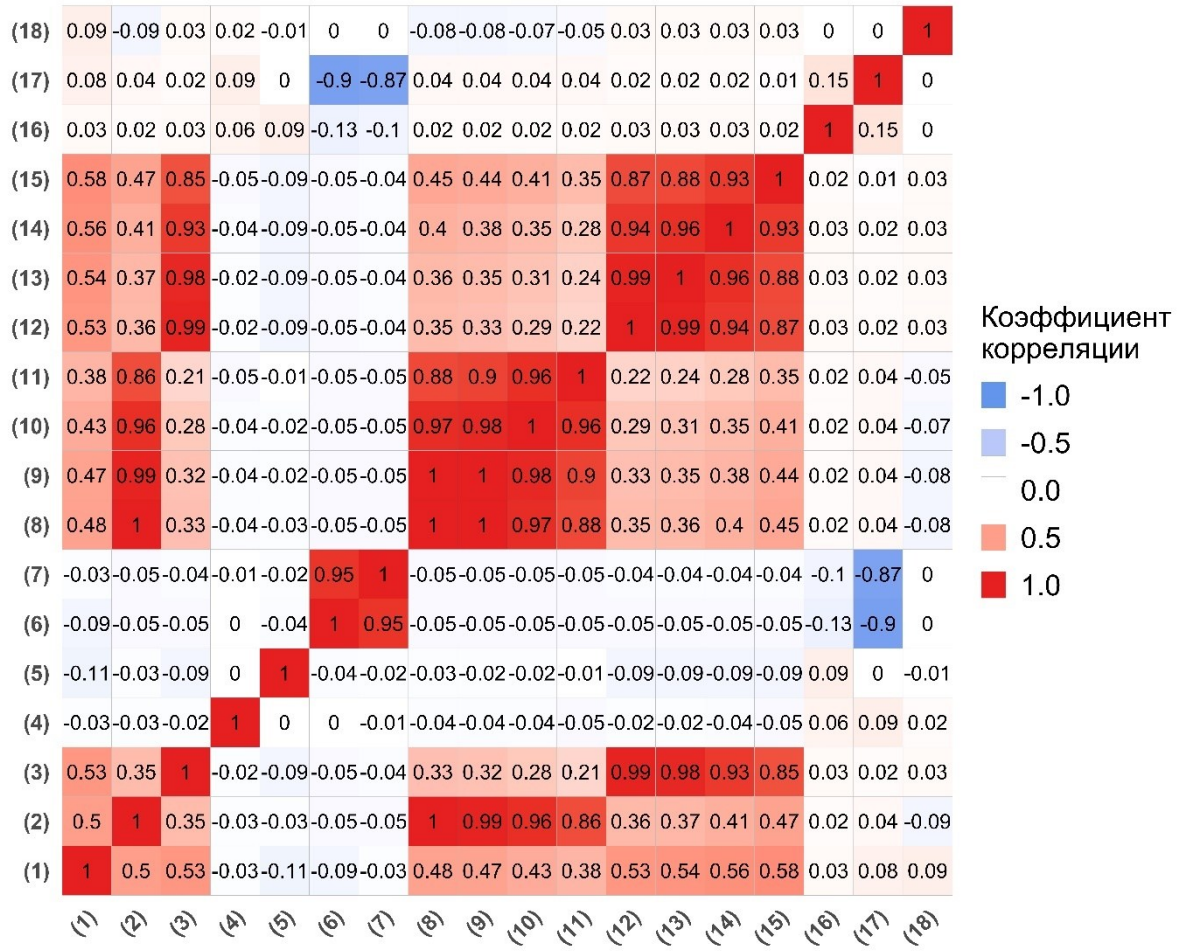


Рисунок 3. Корреляционная матрица предикторов по ряду 2 (обозначения в соответствии с табл. 3).
Figure 3. Correlation matrix of predictors for dataset 2 (see Table 3 for description).

Прогнозы по часовым рядам данных на моделях Prophet и ARIMA дали неудовлетворительные результаты. Так, например, прогноз по часовым данным на тестовой выборке моделью Prophet, обученной по данным за 2021 г., дал среднеквадратическую ошибку (RMSE) 46,3 NTU, а значение эффективности Нэша-Сатклиффа составило –11,7. Модель градиентного бустинга XGBRegressor показала лучшие результаты по часовым данным. Обученная только на данных о расходах воды, модель в общем сумела правильно определить пики мутности и тренды для недельного прогноза, однако среднеквадратическая ошибка прогноза (RMSE) составила 28,8 NTU.

Применение модели XGBRegressor для суточных величин по данным за вторую половину 2021 г. позволил спрогнозировать общий ход мутности воды на 4 недели вперёд. Один пик предсказанных значений мутности оказался сильно завышен, что дало вклад в величину среднеквадратической ошибки (RMSE), составившей 32,6 NTU. При использовании в XGBRegressor всех данных с 2019 по 2021 гг. при тестовой выборке 45 дней среднеквадратическая ошибка прогноза составила 37,2 NTU, однако при этом в целом верно был предсказан ход мутности и время наступления пиков (таблица 5).

Таблица 5. Рассмотренные методы количественной оценки мутности.
Table 5. Turbidity quantification methods reviewed.

Модель	Ряд наблюдений для обучения	Временное разрешение ряда	RMSE (в NTU)	Комментарий
Prophet	Полный ряд наблюдений 2021 г.	1час	46,3	Модель неверно оценила тренды, цикличность и пики. Верно оценено время наступления отдельных пиков.
ARIMA	Данные наблюдений за лето–осень 2021 г.	1 час	14,7	Несмотря на небольшую величину ошибки, модель неверно оценила тренды, цикличность и пики.
	Данные наблюдений за весну 2021 г.	1 час	24,2	Несмотря на небольшую величину ошибки, модель неверно оценила тренды, цикличность и пики.
XGBRegressor	Полный ряд наблюдений 2021 г.	1 час	28,8	Прогноз на 1 неделю. Правильно оценены общие тренды и время наступления пиков.
	Данные наблюдений за лето–осень 2021 г.	1 сутки	32,6	Прогноз на 4 недели. Правильно оценены общие треды, сильно завышен один из пиков.
	Все имеющиеся наблюдения	1 сутки	37,2	Прогноз на 45 дней. Правильно оценён общий ход мутности; сильно завышены пиковые значения.
RandomForestRegressor	Все имеющиеся наблюдения	1 сутки	27,5	Прогноз на 45 дней. Правильно оценён общий ход мутности, время наступления пиков, пиковые значения оказались завышены.
	Полный ряд наблюдений 2021 г.	1 сутки	29,5	Прогноз на 35 дней. Правильно оценён общий ход мутности, время наступления пиков, пиковые значения оказались завышены.
LSTM RNN	Полный ряд наблюдений 2021 г.	1 сутки	10,8	Прогноз на 3 недели. В первую неделю мутность была предсказана с высокой точностью, правильно смоделированы время наступления пика и его абсолютная величина. К концу периода прогноз разошёлся с данными наблюдений.
	Полный ряд наблюдений 2021 г.	1 час	14,4	Прогноз на 3 недели. Модель смогла правильно оценить общий ход мутности, время наступления пиков и их абсолютную величину. Также моделью была выявлена цикличность колебаний.

Более высокую эффективность при обучении на всех имеющихся данных за 2019–2021 гг. и прогнозе на 45 дней продемонстрировала модель RandomForestRegressor. Среднеквадратическая ошибка прогноза составила 27,5 NTU. Моделью был верно спрогнозирован общий ход мутности и время наступления пиков, которые по своей величине также оказались завышенными, но менее выраженно, чем при прогнозе моделью

XGBRegressor. При использовании в качестве обучающей выборки только данных 2021 г. прогноз на 35 дней среднеквадратическая ошибка прогноза составила 29,5 NTU. Общий тренд хода мутности был предсказан верно, однако значительное завышение пиковых величин мутности внесло вклад в большую величину погрешности. При этом моделью было правильно определено время наступления пиков.

Наиболее успешным является описание среднесуточных значений мутности при помощи нейросетевой модели LSTM. Среднеквадратическая ошибка составила 10,8 NTU. Величины мутности за первую неделю прогноза были предсказаны с высокой точностью, были верно определены время и величина локального пика мутности, однако далее расхождение фактических и прогнозных значений стало более явным (рисунок 4). Прогноз значения мутности на срок в 3 недели при использовании для обучения часовых наблюдений за мутностью, расходами, осадками и температурой по данным 2021 г. показали среднеквадратическую ошибку прогноза 14,4 NTU. Модель смогла предсказать общий ход мутности и время наступления пиков, однако несколько пиков моделью выявлены не были.

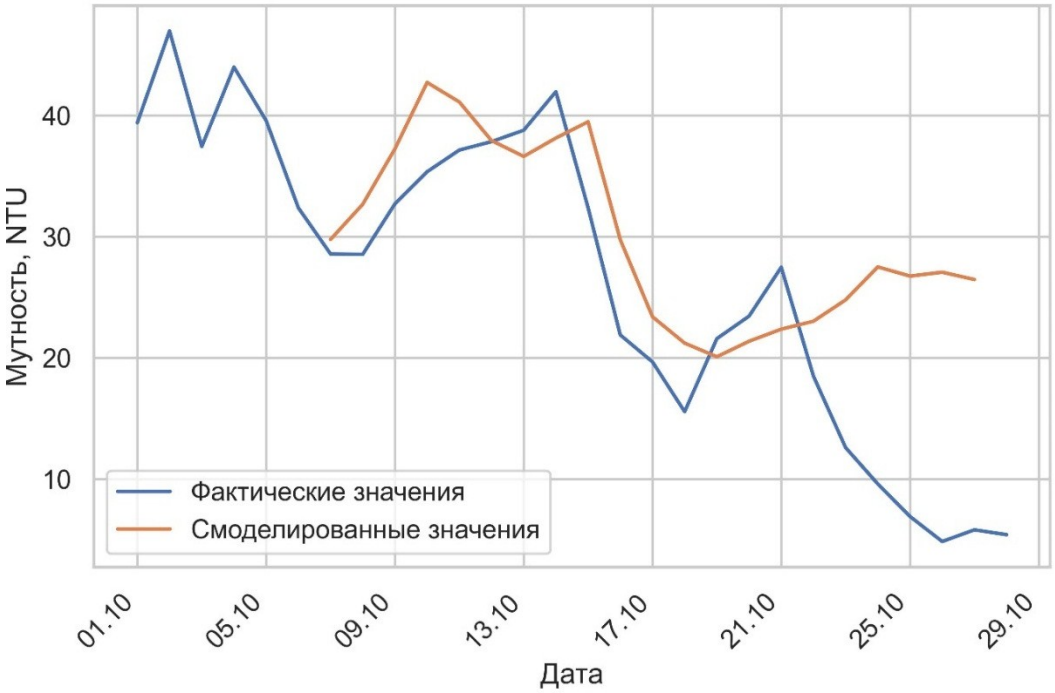


Рисунок 4. Фактические (синие) и смоделированные (оранжевый) величины оптической мутности воды (NTU) с помощью LSTM RNN по суточным данным в 2021 г.
Figure 4. Actual (blue) and simulated (orange) optical turbidity (NTU) values using LSTM RNN for daily data in 2021.

Выводы

На основе обработки данных высокочастотного автоматизированного мониторинга уровней и мутности воды дана оценка режима взвешенных наносов р. Сетунь. Выявлено увеличение значений мутности воды в 3–4 раза по сравнению с фоновыми величинами; сделан вывод о нестационарном режиме переноса взвесей; выявлена роль случайных коротких сбросов из дренажно-коллекторной сети, фиксируемых по всему бассейну. Отмечено наличие статистически достоверных корреляций мутности воды и интенсивности осадков и расходов воды, при этом корреляция с расходами снижается при увеличении лага по времени от 0 до 100 часов.

Дана оценка возможности прогнозирования величин мутности воды на р. Сетунь по данным об атмосферных осадках и расходах воды. Прогностические модели регрессии ARIMA и

Prophet дали неудовлетворительные результаты. Модели на основе деревьев решений XGBRegressor и RandomForestRegressor справились с задачей прогнозирования времени наступления пиков и общего тренда колебаний, однако среднеквадратическая ошибка прогнозов этих моделей зачастую сопоставима со среднеквадратическими отклонениями мутности за соответствующие периоды, что приводит к неудовлетворительным результатам. Рекуррентная нейросеть LSTM RNN, показавшая лучшие результаты с наименьшими погрешностями, является оптимальным из рассмотренных методов прогнозирования мутности воды.

Благодарности

Развитие мониторинговой сети и натурные исследования в бассейне р. Сетунь выполнено в рамках проекта РНФ 19-77-30004. Концепция исследования подготовлена в рамках программы развития Междисциплинарной научно-образовательной школы МГУ имени М.В. Ломоносова «Будущее планеты и глобальные изменения окружающей среды».

Литература

Караушев А.В. Сток наносов, его изучение и географическое распределение. Л.: Гидрометеоиздат, 1977. 240 с.

Прогнозирование уровней воды на реках Краснодарского края с использованием методов машинного обучения / *Белякова П.А., Морейдо В.М., Цыпленков А.С., Амербаев А.Н., Гречишникова Д.А., Курочкина Л.С., Филиппов В.А., Макеев М.С.* // Водные ресурсы. 2022. Т. 49. № 1. С. 13–26. https://doi.org/10.31857/S0321059622010047.

Реакция водного стока малой городской реки на экстремальные дождевые осадки на территории Москвы в 2020 и 2021 гг. / *Чалов С.Р., Платонов В.С., Морейдо В.М., Самохин М.А., Ярынич Ю.И., Коршунова Н.Н., Болгов М.В., Касимов Н.С.* // Метеорология и гидрология. 2023. Т. 2. С. 69–79. https://doi.org/10.52002/0130-2906-2023-2-69-79.

Промахова Е.В., Чалов С.Р. Современные технологии определения мутности воды // Маккавеевские чтения-2014. 2015. С. 82–95.

Чалов С.Р., Цыпленков А.С. Роль крупномасштабной турбулентности в изменении мутности речных вод // Вестник Московского университета. Серия 5. География. 2020. №. 3. С. 34–46.

Черных О.Н., Сабитов М.А., Алтунин В.И. Типизированные приемы экологического восстановления малых рек Москвы (на примере р.

References

Behrouz M.S., Yazdi M.N., Sample D.J. Using Random Forest, a machine learning approach to predict nitrogen, phosphorus, and sediment event mean concentrations in urban runoff. *Journal of Environmental Management*, 2022, vol. 317, pp. 115412. https://doi.org/10.1016/j.jenvman.2022.115412.

Belyakova P.A., Moreido V.M., Tsyplenkov A.S., Amerbaev A.N., Grechishnikova D.A., Kurochkina L.S., Filippov V.A., Makeev M.S. Prognozirovanie urovnei vody na rekakh Krasnodarskogo kraya s ispol'zovaniem metodov mashinnogo obucheniya [Forecasting water levels on the rivers of the Krasnodar Territory using machine learning methods]. *Vodnye resursy [Water resources],* 2022, vol. 49, no. 1, pp. 13–26. https://doi.org/10.31857/S0321059622010047. (In Russian).

Bhattacharya B., Solomatine D.P. Machine learning in sedimentation modelling. *Neural Networks,* 2006, vol. 19, no. 2, pp. 208–214. https://doi.org/10.1016/j.neunet.2006.01.007.

Box G.E.P., Jenkins G.M. *Time Series Analysis: Forecasting and Control Holden-Day series in time series analysis and digital processing.* San Francisco, CA: Holden-Day, 1970, 575 p.

Carling P.A., Orr H.G. Morphology of riffle–pool sequences in the River Severn, England. *Earth Surface Processes and Landforms: The Journal of the British Geomorphological Research Group*, 2000, vol. 25, no. 4, pp. 369–384.

Сетунь) // Природообустройство. 2015. №. 3. С. 57–64.

Behrouz M.S., Yazdi M.N., Sample D.J. Using Random Forest, a machine learning approach to predict nitrogen, phosphorus, and sediment event mean concentrations in urban runoff // Journal of Environmental Management. 2022. V. 317. P. 115412. https://doi.org/10.1016/j.jenvman.2022.115412.

Bhattacharya B., Solomatine D.P. Machine learning in sedimentation modelling // Neural Networks. 2006. V. 19. № 2. P. 208–214. https://doi.org/10.1016/j.neunet.2006.01.007.

Box G.E.P., Jenkins G.M. Time Series Analysis: Forecasting and Control Holden-Day series in time series analysis and digital processing. / San Francisco, CA: Holden-Day, 1970, 575 p.

Carling P.A., Orr H.G. Morphology of riffle–pool sequences in the River Severn, England // Earth Surface Processes and Landforms: The Journal of the British Geomorphological Research Group. 2000. V. 25. № 4. P. 369–384. https://doi.org/10.1002/(SICI)1096-9837(200004)25:4<369::AID-ESP60>3.0.CO;2-M.

Chalov S.R., Platonov V.S., Moreido V.M., Samokhin M.A. Small Urban River Runoff Response to 2020 and 2021 Extreme Rainfalls on the Territory of Moscow // Russian meteorology and hydrology. 2023. V. 48. № 2. P. 138–146. https://doi.org/10.3103/S1068373923020061.

Chen T., Guestrin C. XGBoost: A Scalable Tree Boosting System // Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining : KDD '16. – New York, NY, USA: ACM, 2016. P. 785–794. https://doi.org/10.1145/2939672.2939785.

Cohen S., Kettner A.J., Syvitski J.P.M. Global suspended sediment and water discharge dynamics between 1960 and 2010: Continental trends and intra-basin sensitivity // Global and planetary change. 2014. V. 115. P. 44–58. https://doi.org/10.1016/j.gloplacha.2014.01.011.

Droujko J., Molnar P. Open-source, low-cost, in-situ turbidity sensor for river network monitoring // https://doi.org/10.1002/(SICI)1096-9837(200004)25:4<369::AID-ESP60>3.0.CO;2-M.

Chalov S.R., Platonov V.S., Moreido V.M., Samokhin M.A. Small Urban River Runoff Response to 2020 and 2021 Extreme Rainfalls on the Territory of Moscow. *Russian meteorology and hydrology*, 2023, vol. 48, no. 2, pp. 138–146. https://doi.org/10.3103/S1068373923020061.

Chalov S.R., Platonov V.S., Moreido V.M., Samokhin M.A., Yarynich Yu.I., Korshunova N.N., Bolgov M.V., Kasimov N.S. Reaktsiya vodnogo stoka maloi gorodskoi reki na ekstremal'nye dozhdevye osadki na territorii Moskvy v 2020 i 2021 gg. [Reaction of water flow of a small urban river to extreme rainfall in Moscow in 2020 and 2021]. *Meteorologiya i gidrologiya [Meteorology and hydrology]*. 2023, vol. 2, pp. 69-79. https://doi.org/10.52002/0130-2906-2023-2-69-79. (In Russian)

Chalov S.R., Tsyplenkov A.S. Rol' krupnomasshtabnoi turbulentnosti v izmenenii mutnosti rechnykh vod [Influence of the macroturbulence on the dynamics of river water turbidity]. *Vestnik Moskovskogo universiteta. Seriya 5. Geografiya,* 2020, no. 3, pp. 34–46. (In Russian; abstract in English).

Chen T., Guestrin C. XGBoost: A Scalable Tree Boosting System. *Proceedings of the twenty second ACM SIGKDD International Conference on Knowledge Discovery and Data Mining : KDD '16*, New York, NY, USA: ACM, 2016, pp. 785–794. https://doi.org/10.1145/2939672.2939785.

Chernykh O.N., Sabitov M.A., Altunin V.I. Tipizirovannye priemy ekologicheskogo vosstanovleniya malykh rek Moskvy (na primere r. Setun') [Typical methods of ecological restoration of small rivers of Moscow (using the example of the Setun River)]. *Prirodoobustroistvo,* 2015, no. 3, pp. 57–64. (In Russian; abstract in English).

Cohen S., Kettner A.J., Syvitski J.P.M. Global suspended sediment and water discharge dynamics between 1960 and 2010: Continental trends and intra-basin sensitivity. *Global and planetary change*, 2014, vol. 115, pp. 44–58. https://doi.org/10.1016/j.gloplacha.2014.01.011.

Scientific Reports. 2022. V. 12. №1. P. 10341. https://doi.org/10.1038/s41598-022-14228-4.

Hochreiter S., Schmidhuber J. Long Short-Term Memory // Neural Computation. 1997. V. 9. № 8. P. 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735.

Moreido V., Gartsman B., Solomatine D.P., Suchilina Z. How Well Can Machine Learning Models Perform without Hydrologists? Application of Rational Feature Selection to Improve Hydrological Forecasting // Water. 2021. V. 13. № 12. P. 1696. https://doi.org/10.3390/w13121696.

On the complexities of sediment load modeling using integrative machine learning: Application of the great river of Loíza in Puerto Rico / *Zounemat-Kermani M., Mahdavi-Meymand A., Alizamir M., Adarsh S., Yaseen Z.M.* // Journal of Hydrology. 2020. V. 585. P. 124759. https://doi.org/10.1016/j.jhydrol.2020.124759.

Press W.H., Teukolsky S.A. Savitzky-Golay Smoothing Filters // Computers in Physics. 1990. V. 4. № 6. P. 669–672. https://doi.org/10.1063/1.4822961.

Rainfall-runoff modelling using Long Short-Term Memory (LSTM) networks / *Kratzert F., Klotz D., Brenner C., Schulz K., Herrnegger M.* // Hydrology and Earth System Sciences. 2018. V. 22. № 11. P. 6005-6022. https://doi.org/10.5194/hess-22-6005-2018.

Scikit-learn: Machine Learning in Python. / *Pedregosa F. et al.* // the Journal of machine Learning research. 2011. V. 12. P. 2825–2830.

Sharafati A., Haji Seyed Asadollah S.B., Motta D., Yaseen Z.M. Application of newly developed ensemble machine learning models for daily suspended sediment load prediction and related uncertainty analysis // Hydrological Sciences Journal. 2020. V. 65. № 12. P. 2022–2042. https://doi.org/10.1080/02626667.2020.1786571.

Sokolov D.I., Erina O.N., Tereshina M.A., Puklakov V. V. Impact Of Mozhaysk Dam On The Moscow River Sediment Transport // GEOGRAPHY, ENVIRONMENT, SUSTAINABILITY. 2020. V. 13. №. 4. P. 24–31. https://doi.org/10.24057/2071-9388-2019-150.

Droujko J., Molnar P. Open-source, low-cost, in-situ turbidity sensor for river network monitoring. *Scientific Reports*, 2022, vol. 12, no. 1, pp. 10341. https://doi.org/10.1038/s41598-022-14228-4.

Hochreiter S., Schmidhuber J. Long Short-Term Memory. *Neural Computation,* 1997, vol. 9, no. 8, pp. 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735.

Karaushev A.V. *Stok nanosov, ego izuchenie i geograficheskoe raspredelenie [Sediment runoff, its study and geographical distribution]*. Leningrad: Gidrometeoizdat, 1977, 240 p. (In Russian)

Kratzert F., Klotz D., Brenner C., Schulz K., Herrnegger M. Rainfall-runoff modelling using Long Short-Term Memory (LSTM) networks. *Hydrology and Earth System Sciences,* 2018, vol. 22, no. 11, pp. 6005–6022. https://doi.org/10.5194/hess-22-6005-2018.

Moreido V., Gartsman B., Solomatine D.P., Suchilina Z. How Well Can Machine Learning Models Perform without Hydrologists? Application of Rational Feature Selection to Improve Hydrological Forecasting. *Water,* 2021, vol. 13, no. 12, pp. 1696. https://doi.org/10.3390/w13121696.

Pedregosa F. et al. Scikit-learn: Machine Learning in Python. *the Journal of machine Learning research,* 2011, vol. 12, pp. 2825–2830.

Press W.H., Teukolsky S.A. Savitzky-Golay Smoothing Filters. *Computers in Physics,* 1990, vol. 4, no. 6, pp. 669–672. https://doi.org/10.1063/1.4822961.

Promakhova E.V., Chalov S.R. Sovremennye tekhnologii opredeleniya mutnosti vody. *Makkaveevskie chteniya-2014*, 2015, pp. 82–95.

Sharafati A., Haji Seyed Asadollah S.B., Motta D., Yaseen Z.M. Application of newly developed ensemble machine learning models for daily suspended sediment load prediction and related uncertainty analysis. *Hydrological Sciences Journal*, 2020, vol. 65, no. 12, pp. 2022–2042. https://doi.org/10.1080/02626667.2020.1786571.

Sokolov D.I., Erina O.N., Tereshina M.A., Puklakov V. V. Impact Of Mozhaysk Dam On The Moscow River Sediment Transport. *GEOGRAPHY, ENVIRONMENT, SUSTAINABILITY,* 2020, vol. 13,

Solomatine D.P., Ostfeld A. Data-driven modelling: some past experiences and new approaches // Journal of Hydroinformatics. 2008. V. 10. № 1. P. 3. https://doi.org/10.2166/hydro.2008.015.

Taylor S.J., Letham B. Forecasting at Scale // The American Statistician. 2018. V. 72. № 1. P. 37–45. https://doi.org/10.1080/00031305.2017.1380080.

no. 4, pp. 24–31. https://doi.org/10.24057/2071-9388-2019-150.

Solomatine D.P., Ostfeld A. Data-driven modelling: some past experiences and new approaches. *Journal of Hydroinformatics*, 2008, vol. 10, no. 1, p. 3. https://doi.org/10.2166/hydro.2008.015.

Taylor S.J., Letham B. Forecasting at Scale. *The American Statistician*, 2018, vol. 72, no. 1, pp. 37–45. https://doi.org/10.1080/00031305.2017.1380080.

Zounemat-Kermani M., Mahdavi-Meymand A., Alizamir M., Adarsh S., Yaseen Z.M. On the complexities of sediment load modeling using integrative machine learning: Application of the great river of Loíza in Puerto Rico. *Journal of Hydrology,* 2020, vol. 585, pp. 124759. https://doi.org/10.1016/j.jhydrol.2020.124759.