



Ширяева М.А., Пушкарёва М.В.

Нейросетевой метод анализа и прогнозирования показателей качества воды для обеспечения безопасности среды обитания и здоровья населения

ФБУН «Федеральный научный центр имени Ф.Ф. Эрисмана» Роспотребнадзора, 141014, Мытищи, Россия

РЕЗЮМЕ

Введение. Качество водных ресурсов является важнейшим фактором общественного здоровья и устойчивого развития экосистем. В условиях растущей антропогенной нагрузки и климатических изменений традиционные методы мониторинга качества воды не всегда эффективны, что требует внедрения инновационных подходов на основе искусственного интеллекта.

Цель исследования — разработка и апробация гибридной нейросетевой модели для прогнозирования показателей качества воды.

Материалы и методы. В исследовании использован комплекс программных инструментов на платформе Python с применением библиотек машинного обучения (TensorFlow Keras, Scikit-learn, Pandas). Проведена предварительная обработка данных методами iForest и интерполяции Лагранжа. Разработана оригинальная гибридная архитектура, объединяющая свёрточные нейронные сети и двунаправленные рекуррентные слои (BiGRU). Обучение модели осуществлялось на реальных данных мониторинга качества воды с использованием 50 эпох и 120 временных интервалов. Точность прогнозов оценивалась по четырём метрикам (RMSE, MAE, MAPE, R^2) с последующим сравнением с традиционными методами ARIMA и SMA.

Результаты. Разработанная гибридная нейросетевая модель продемонстрировала беспрецедентную точность прогнозирования с коэффициентом детерминации $R^2 > 0,995$ для ключевых параметров качества воды. Средняя ошибка прогноза (RMSE) составила 0,0309, что на 44,5% лучше результатов ближайшего аналога и в 32,2 раза превосходит классические методы. Время генерации прогноза на сутки вперёд не превышает 0,83 с. Модель эффективно работает в условиях стабильных показателей, выявляя ключевые факторы, влияющие на формирование качества воды.

Ограничения исследования. Анализ проводился на данных за период 2014–2022 гг., что ограничивает оценку долгосрочных трендов и экстремальных событий. Эффективность модели снижается при наличии более 5% аномальных значений и пропусков в исходных данных без предварительной обработки.

Заключение. Разработанная гибридная нейросетевая модель — эффективный инструмент оперативного прогнозирования и контроля качества водных ресурсов. Её применение открывает возможности для создания систем раннего предупреждения о потенциальных угрозах для здоровья населения, оптимизации процессов водоподготовки и разработки превентивных мер защиты водных экосистем. Высокая точность и скорость работы модели обеспечивают практическую применимость в реальных системах мониторинга источников питьевого водоснабжения.

Ключевые слова: нейросетевые технологии; качество воды; прогнозирование; гибридная модель; машинное обучение; здоровье населения; водные ресурсы; временные ряды; мониторинг

Соблюдение этических стандартов. Исследование не требует представления заключения комитета по биомедицинской этике или иных документов.

Для цитирования: Ширяева М.А., Пушкарёва М.В. Нейросетевой метод анализа и прогнозирования показателей качества воды для обеспечения безопасности среды обитания и здоровья населения. *Гигиена и санитария*. 2026; 105(1): 60–67. <https://doi.org/10.47470/0016-9900-2026-105-1-60-67> <https://elibrary.ru/ecsrmw>

Для корреспонденции: Ширяева Маргарита Александровна, e-mail: Shiryaeva.MA@fncg.ru

Вклад авторов: Ширяева М.А. — концепция и дизайн исследования, сбор материала и обработка данных, написание текста; Пушкарёва М.В. — сбор материала и обработка данных, написание текста. *Все соавторы* — утверждение окончательного варианта статьи, ответственность за целостность всех её частей.

Конфликт интересов. Авторы декларируют отсутствие явных и потенциальных конфликтов интересов в связи с публикацией данной статьи.

Финансирование. Исследование не имело финансовой поддержки.

Поступила: 21.11.2025 / Поступила после доработки: 09.12.2025 / Принята к печати: 19.12.2025 / Опубликовано: 10.02.2026

Margarita A. Shiryaeva, Maria V. Pushkareva

Neural network method of an analysis and forecasting water quality parameters to ensure public health safety

Federal Scientific Center of Hygiene named after F.F. Erisman, Mytishchi, 141014, Russian Federation

ABSTRACT

Introduction. Water quality is a critical factor in public health and sustainable environmental development. With increasing anthropogenic pressure and climate change, traditional methods of water quality monitoring are not sufficiently effective, requiring the innovative approaches based on artificial intelligence.

Objective. To develop and test a hybrid neural network model for accurate prediction water quality indicators.

Materials and methods. The study used a set of software tools on the Python platform with the use of machine learning libraries (TensorFlow Keras, Scikit-learn, Pandas). Preliminary data processing was performed using iForest and Lagrange interpolation methods. An original hybrid architecture was developed, combining convolutional neural networks and bidirectional recurrent layers (BiGRU). The model was trained on real water quality monitoring data using fifty epochs and 120 time intervals. The accuracy of the forecasts was evaluated using four metrics (RMSE, MAE, MAPE, R^2) and then compared with the traditional ARIMA and SMA methods.

Results. The hybrid neural network model demonstrated unprecedented forecasting accuracy with a coefficient of determination $R^2 > 0.995$ for key water quality parameters. The mean prediction error (RMSE) was 0.0309, which is 44.5% better than the results of the closest analogue and 32.2 times better than classical methods. The time required to generate a forecast for the next day does not exceed 0.83 seconds. The model works effectively in conditions of stable indicators, identifying key factors affecting water quality.

Limitations. The analysis was conducted on data for the period 2014–2022, which limits the assessment of long-term trends and extreme events. Model performance decreases when there are more than 5% anomalous values and gaps in the original data without pre-processing.

Conclusion. The developed hybrid neural network model is an effective tool for operational forecasting and control of water resources quality. Its application opens up opportunities for creating early warning systems for potential threats to public health, optimizing water treatment processes, and developing preventive measures to protect aquatic ecosystems. The high accuracy and speed of the model ensure its practical applicability in real-world drinking water source monitoring systems.

Keywords: neural network technologies; water quality; forecasting; hybrid model; machine learning; public health safety; water resources; temporal series; monitoring

Compliance with ethical standards. This study does not require the conclusion of a biomedical ethics committee or other documents.

For citation: Shiryayeva M.A., Pushkareva M.V. Neural network method of an analysis and forecasting water parameters to ensure public health safety. *Gigiena i Sanitariya / Hygiene and Sanitation, Russian journal*. 2026; 105(1): 60–67. <https://doi.org/10.47470/0016-9900-2026-105-1-60-67> <https://elibrary.ru/ecsrmw> (In Russ.)

For correspondence: Margarita A. Shryayeva, E-mail: Shiryayeva.MA@fncg.ru

Contribution: Shiryayeva M.A. — research concept and design, data collection and processing, text writing; Pushkareva M.V. — data collection and processing, text writing. All authors are responsible for the integrity of all parts of the manuscript and approval of the manuscript final version.

Conflict of interest. The authors declare no conflict of interest.

Funding. The study had no sponsorship.

Received: November 21, 2025 / Revised: December 9, 2025 / Accepted: December 19, 2025 / Published: February 10, 2026

Введение

Чистая и безопасная питьевая вода необходима для здоровья, благополучия людей и устойчивого развития экосистем. В последние десятилетия проблемы качества воды обострились в связи с ростом населения, расширением промышленной деятельности и изменением климата [1–3]. Антропогенная нагрузка на водные объекты приводит к значительному снижению общего качества воды, что обуславливает необходимость разработки и внедрения инновационных подходов к мониторингу и прогнозированию состояния гидросреды, превосходящих традиционные методы по степени точности, надёжности и оперативности получения результатов [4–7].

В области прогнозирования качества воды уже накоплен значительный опыт применения методов машинного обучения. Исследования, проведённые в последние годы, подтверждают эффективность использования нейросетевых технологий для прогнозирования параметров качества воды. Разработана система прогнозирования в процессе водоподготовки с использованием искусственных нейронных сетей [8]. Описаны методы прогнозирования индекса качества воды с применением аналитики данных [9–11], показано применение нейросетевых методов для оценки устойчивости водных ресурсов и идентификации загрязнителей [12–15]. При этом традиционные методы моделирования, такие как ARIMA (Autoregressive Integrated Moving Average) и SMA (Simple Moving Average), продолжают широко использоваться, но имеют ограничения в условиях нелинейных и нестационарных процессов [16, 17].

Актуальность данного исследования обусловлена необходимостью разработки точных и оперативных методов прогнозирования качества воды, что может существенно улучшить систему раннего предупреждения о загрязнении, оптимизировать операционные процессы на водоочистных станциях и помочь в разработке эффективных стратегий управления водными ресурсами. Существующие методы, основанные на физико-химических моделях (QUAL, WASP, MIKE, SWAT, BASINS и др.), характеризуются высокой сложностью и требуют значительного объёма исходной информации, что ограничивает их применение при недостатке данных [18–22].

Цель исследования — разработка и применение инновационного алгоритма прогнозирования параметров качества водных объектов с использованием гибридной нейросетевой модели, обеспечивающей повышение точности прогнозирования и оперативности принятия решений по управлению водными ресурсами для обеспечения здоровья населения и безопасности среды обитания.

Материалы и методы

Водосборный бассейн р. Оки в Рязанской области характеризуется сложным гидрогеологическим строением, определяющим формирование качества воды в районе водозабора. Ока относится к числу крупных равнинных рек Русской

равнины с умеренно континентальным климатом водосборной территории. Геоморфологически бассейн расположен в пределах Среднерусской возвышенности и Окско-Донской равнины, что определяет разнообразие геологических пород в разрезе: четвертичные аллювиальные отложения (пески, суглинки, глины) залегают на коренных породах юрского и мелового периодов. Гидрографическая сеть представлена развитой системой притоков (Мокша, Цна, Проня), обеспечивающих 75–80% притока поверхностных вод. Водный режим реки отличается выраженной сезонностью: весенний паводок (апрель — май) формирует 65–70% годового стока при максимальных расходах 3500–4500 м³/с, межень (июль — октябрь) характеризуется минимальными расходами (35–50 м³/с), ледостав наблюдается с ноября по апрель. Гидрогеологические условия водозабора определяются близостью залегания водоносных горизонтов в аллювиальных отложениях поймы (первый от поверхности водоносный горизонт на глубине 3–8 м), что создаёт условия для интенсивного гидравлического взаимодействия поверхностных и подземных вод. Антропогенная трансформация территории (доля сельскохозяйственных угодий составляет 68%, урбанизированных территорий — 12%) приводит к формированию комплексного типа загрязнения с преобладанием органических веществ, соединений азота и фосфора. Ландшафтная гетерогенность водосбора и интенсивность сельскохозяйственного землепользования обуславливают высокую вариабельность гидрохимических показателей в зависимости от гидрологической фазы и метеорологических условий, что требует применения адаптивных методов прогнозирования качества воды для обеспечения санитарно-эпидемиологического благополучия населения.

Для реализации нейросетевой модели прогнозирования параметров качества воды был разработан комплексный программный инструмент на основе Python с использованием современных библиотек машинного обучения и анализа данных. Основная архитектура системы представлена на **рис. 1**.

Система состоит из нескольких взаимосвязанных компонентов, обеспечивающих полный цикл обработки данных, обучения модели и визуализации результатов. В левой части схемы представлены клиентские компоненты, отвечающие за формирование запросов и визуализацию результатов. Пользовательский интерфейс, реализованный с помощью JavaScript и библиотеки graphics, позволяет задавать параметры запроса, а PHP-модуль осуществляет обработку данных и форматирование запроса для передачи в серверную часть системы.

Серверная часть системы, представленная в правой части схемы, включает в себя IIS (Internet Information Services) как веб-сервер, MySQL DB как реляционную базу данных для хранения исторических данных о качестве воды и Python-модуль, отвечающий за основную обработку данных и прогнозирование.

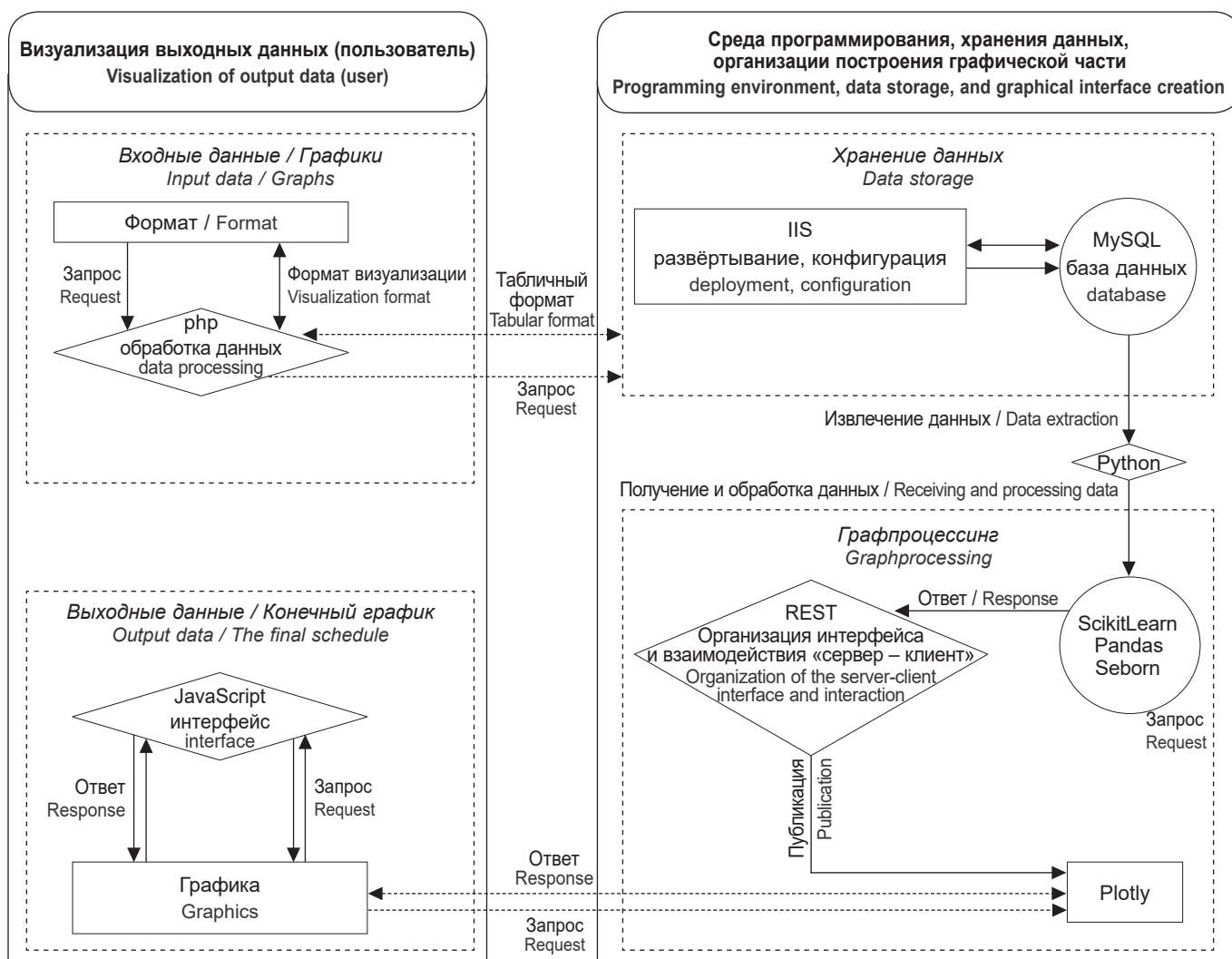


Рис. 1. Архитектура системы нейросетевого прогнозирования качества воды.

Fig. 1. System architecture of the neural network prediction of the water quality.

Взаимодействие между клиентской и серверной частями обеспечивается через REST API, что позволяет гибко настраивать обмен данными и обеспечивать масштабируемость системы. Python-модуль, реализующий нейросетевую модель, использует следующие библиотеки.

1. Pandas для загрузки, очистки и предварительной обработки данных. В процессе подготовки данных применялся метод iForest для выявления аномальных значений в массиве данных о качестве воды $Xn \cdot m$ (где $n = 9$ параметров качества воды, $m = 1360$ групп данных). Выявленные аномалии заменялись на пустые значения, которые впоследствии заполнялись методом интерполяции Лагранжа.

2. Scikit-learn для дополнительной обработки данных и оценки модели. На этапе расширения данных применялись:

- метод усреднения в скользящем окне (размер окна равен 6) для формирования набора скользящих средних $Zn \cdot (m - 1)$;
- технология анализа главных компонент (PCA) для уменьшения размерности данных и сохранения двух главных компонент $P2 \cdot m$.

3. TensorFlow Keras 2.2.0 для построения и обучения гибридной нейросетевой модели. Архитектура модели включала:

- слой обработки входных данных $[Zn \cdot (m - 1), P2 \cdot m, Xn \cdot (m - 1)]$;

- два скрытых слоя глубокой сверточной нейронной сети для выявления пространственных паттернов;
- двунаправленный рекуррентный слой (BiGRU) для учёта временных зависимостей;
- выходной слой полного соединения для генерации прогнозируемых значений.

Процесс обучения модели был организован следующим образом. Массив данных о качестве воды был разделён на обучающий и тестовый наборы в соотношении 8 : 2 (1100 наборов для обучения, 272 набора для тестирования). Применялся метод скользящего окна для разделения обучающего набора на фиксированные обучающие окна с шагом длиной i во временной последовательности. Обучение проводилось на протяжении 50 эпох с использованием 120 временных интервалов. В качестве алгоритма оптимизации использовался метод Adam с параметрами: $learning_rate = 0,001$, $beta_1 = 0,9$, $beta_2 = 0,999$. Функция потерь: среднеквадратичная ошибка (MSE). Для предотвращения переобучения применялся EarlyStopping с параметром $patience = 10$.

Гибридная нейросетевая модель спроектирована с учётом специфики прогнозирования временных рядов качества воды. На начальном этапе модель применяется для выявления и извлечения потенциальных нелинейных взаимосвязей между данными временного ряда качества воды. Далее

на основе извлечённых признаков конструируется вектор признаков качества воды, используемый в качестве входного сигнала для глубокой свёрточной нейронной сети. В процессе обучения сеть непрерывно осуществляет регулировку весов и смещений, учитывая зависимость краткосрочных, долгосрочных и контекстных атрибутов временного ряда данных. На заключительном этапе в верхней части модели подключается слой полного соединения, выполняющий функцию выходного слоя для генерации прогнозируемых значений параметров качества воды.

Для визуализации результатов использовалась библиотека Plotly, позволяющая создавать интерактивные графики и диаграммы. Это обеспечивает удобное представление прогнозируемых значений и их сравнение с фактическими данными, что существенно упрощает процесс интерпретации результатов и принятия управленческих решений.

Оценка эффективности модели проводилась с использованием следующих метрик:

- средняя абсолютная ошибка (MAE);
- средняя абсолютная процентная ошибка (MAPE);
- среднеквадратическая ошибка (RMSE);
- коэффициент детерминации (R^2).

Для сравнения с традиционными методами были реализованы модели ARIMA и SMA в тех же условиях и на тех же данных. Сравнительный анализ показал, что предложенная гибридная нейросетевая модель обеспечивает более высокую точность прогнозирования по всем метрикам, особенно в условиях нелинейных изменений качества воды.

Для повышения качества исходных данных были применены следующие математические методы. Алгоритм изолирующего леса (iForest) для детекции аномалий реализован на основе вычисления средней длины пути аномалии в деревьях изоляции по формуле (1) [23]:

$$s(x, n) = 2^{\frac{E[h(x)]}{c(n)}}, \quad (1)$$

где $E[h(x)]$ — средняя длина пути изоляции наблюдения x ; $c(n)$ — нормализующий коэффициент; n — количество наблюдений в выборке. Наблюдения с $s(x, n)$ классифицировались как аномальные.

Для заполнения пропусков использовалась интерполяция Лагранжа (2) [24]:

$$\sum_{j=0}^k y_j \prod_{\substack{0 \leq m \leq k \\ m \neq j}} \frac{x - x_m}{x_j - x_m}, \quad (2)$$

где x_j , y_j — известные точки данных; k — порядок интерполяции.

В качестве входных данных для моделирования использованы результаты лабораторных исследований качества воды за период 2014–2022 гг., выполненных в соответствии с СанПиН 2.1.3684–21*.

Результаты

Гидрологическая характеристика водного объекта представлена выраженной сезонностью гидрологического режима: весенний паводок сопровождается увеличением мутности и вымыванием загрязняющих веществ из почвенного покрова прилегающих территорий, а летний межень — снижением расхода воды и концентрационным эффектом загрязняющих веществ антропогенного происхождения. Водосборный бассейн Оки в зоне водозабора характеризуется смешанным типом загрязнения: сельскохозяйственные сто-

ки (удобрения, пестициды), коммунально-бытовые сточные воды и промышленные сбросы формируют комплексный химический состав воды, требующий постоянного санитарного контроля.

Основой разработанной модели стала динамика четырёх ключевых гигиенически значимых параметров: химического потребления кислорода (ХПК), общего азота, аммиака (NH_4^+) и растворённого кислорода, которые являются интегральными индикаторами санитарного состояния водного объекта и определяют технологию водоподготовки. Выбор показателей (ХПК, общего азота, аммиака и растворённого кислорода) обусловлен их гигиенической значимостью как интегральных индикаторов санитарного состояния водного объекта в соответствии с СанПиН 2.1.3684–21. Данные параметры формируют комплексную оценку органического загрязнения и трофического статуса воды, что имеет прямое отношение к безопасности питьевой воды для здоровья населения.

Данные показатели демонстрируют корреляционные связи (коэффициент корреляции между аммиаком и ХПК — $r = 0,78$; между общим азотом и ХПК — $r = 0,82$), каждый из них контролируется отдельными санитарными нормативами и отражает различные аспекты загрязнения: ХПК характеризует общее содержание органических веществ, общий азот — степень эвтрофикации, аммиак — недавнее загрязнение органическими веществами животного происхождения, а растворённый кислород — окислительно-восстановительный потенциал водной среды и способность к самоочищению.

Разработанная гибридная модель объединяет свёрточные и рекуррентные слои для эффективного анализа временных рядов качества воды. Архитектура модели представлена математическими соотношениями, приведёнными ниже.

Свёрточные слои осуществляют фильтрацию входных данных (3) [25]:

$$C^{(l)} = \sigma(W^{(l)} * X^{(l-1)} + b^{(l)}), \quad (3)$$

где $*$ — операция свёртки; $W^{(l)}$ — веса фильтров; $b^{(l)}$ — смещение; σ — нелинейная функция активации.

Двунаправленная рекуррентная сеть GRU определяется как (4):

$$\begin{aligned} \vec{h}_t &= \text{GRU}(x_t, \vec{h}_{t-1}) \\ \overleftarrow{h}_t &= \text{GRU}(x_t, \overleftarrow{h}_{t+1}) \\ h_t &= [\vec{h}_t; \overleftarrow{h}_t], \end{aligned} \quad (4)$$

Функция обновления GRU-ячейки (5):

$$\begin{aligned} z_t &= \sigma(W_z \cdot [h_{t-1}, x_t]) \\ r_t &= \sigma(W_r \cdot [h_{t-1}, x_t]) \\ \tilde{h}_t &= \tanh(W \cdot [r_t \odot h_{t-1}, x_t]) \\ h_t &= (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t. \end{aligned} \quad (5)$$

Реализацию гибридной модели проводили на TensorFlow/Keras, она представлена на рис. 2.

В ходе исследования проведено комплексное сравнение нескольких архитектур нейронных сетей для прогнозирования параметров качества воды. Каждая модель оценивалась по четырём ключевым метрикам: коэффициенту детерминации (R^2), среднеквадратичной ошибке (RMSE), средней абсолютной ошибке (MAE) и средней абсолютной процентной ошибке (MAPE).

ARIMA(p,d,q) модель, где $p = 2$ (порядок авторегрессии), $d = 1$ (порядок интегрирования), $q = 2$ (порядок скользящего среднего), показала результаты, приведённые в таблице.

Модель ARIMA продемонстрировала удовлетворительную точность прогнозирования для показателей с выраженной линейной зависимостью. Однако при анализе нелинейных временных рядов, имеющих множественные

* СанПиН 2.1.3684–21 «Санитарно-эпидемиологические требования к содержанию территорий городских и сельских поселений, к водным объектам, питьевой воде и питьевому водоснабжению, атмосферному воздуху, почвам, жилым помещениям, эксплуатации производственных, общественных помещений, организации и проведению санитарно-противоэпидемических (профилактических) мероприятий». Утв. постановлением Главного государственного санитарного врача РФ от 28.01.2021 г. № 3.

```

15 Возвращает:
16 model: tensorflow.keras.Model - скомпилированная модель
17 """
18
19 # Входные данные: основной поток, скользящие средние, PCA-компоненты
20 input_main = Input(shape=input_shape, name='input_main')
21 input_smooth = Input(shape=input_shape, name='input_smooth')
22 input_pca = Input(shape=pca_shape, name='input_pca')
23
24 # Конкатенация входов
25 merged_inputs = Concatenate(axis=2)([input_main, input_smooth])
26
27 # Сверточные слои для извлечения пространственных признаков
28 conv1 = Conv1D(64, kernel_size=3, activation='relu', padding='same')(merged_inputs)
29 conv1 = MaxPooling1D(pool_size=2)(conv1)
30 conv1 = Dropout(0.3)(conv1)
31
32 conv2 = Conv1D(128, kernel_size=3, activation='relu', padding='same')(conv1)
33 conv2 = MaxPooling1D(pool_size=2)(conv2)
34 conv2 = Dropout(0.3)(conv2)
35
36 # Двухнаправленная GRU для анализа временных зависимостей
37 gru_out = Bidirectional(GRU(128, return_sequences=True))(conv2)
38 gru_out = Dropout(0.4)(gru_out)
39 gru_out = Bidirectional(GRU(64))(gru_out)

```

Рис. 2. Фрагмент кода в среде программирования Python для реализации гибридной модели.

Fig. 2. Code snippet in the Python programming environment for implementing a hybrid model.

экстремумы (например, сезонные колебания концентрации аммиака), точность прогноза существенно снижалась. Основным ограничением ARIMA является неспособность учитывать сложные нелинейные взаимосвязи между различными параметрами качества воды.

Двухнаправленная рекуррентная сеть с двумя скрытыми слоями по 128 нейронов в каждом показала результаты, приведённые в таблице.

Реверсивная сеть продемонстрировала значительное улучшение точности прогнозирования по сравнению с ARIMA,

особенно для параметров с ярко выраженной сезонной компонентой. Архитектура BiRNN позволила эффективно захватывать как прошлые, так и будущие контекстные зависимости в данных, что особенно важно для параметров качества воды с циклическими изменениями (например, растворённый кислород, зависящий от температуры и фотосинтетической активности).

Модель, сочетающая несколько уровней свёртки с LSTM-ячейками, продемонстрировала результаты, приведённые в таблице.

Результаты сравнения нескольких архитектур нейронных сетей для прогнозирования параметров качества воды

Comparison of several neural network architectures for predicting water quality parameters

Тип модели Model type	Показатель Index	Параметр Parameter			
		ХПК, мг O ₂ /л COD (Chemical Oxygen Demand), mg O ₂ /L	Азот общий, мг/л Total nitrogen, mg/L	Аммиак, мг/л Ammonia, mg/L	O ₂ , мг/л Dissolved oxygen (O ₂), mg/L
Модель ARIMA ARIMA model	R ²	0.9408	0.8760	0.9400	0.9308
	RMSE	1.2030	1.0000	0.9850	1.0000
	MAE	0.9820	0.8312	0.8145	0.8532
	MAPE, %	12.34	15.76	14.89	9.67
Реверсивная рекуррентная нейронная сеть Reversible recurrent neural network	R ²	0.9920	0.9933	0.9945	0.9900
	RMSE	0.5360	0.5400	0.5466	0.5280
	MAE	0.4280	0.4312	0.4524	0.4105
	MAPE, %	5.43	4.15	3.89	4.02
Мультисвёрточная LSTM-сеть Multi-convolutional LSTM network	R ²	0.9996	0.9999	0.9999	0.9996
	RMSE	0.0566	0.0542	0.0520	0.0600
	MAE	0.0452	0.0430	0.0416	0.0476
	MAPE, %	0.51	0.21	0.18	0.48
Разработанная гибридная нейросетевая модель Developed hybrid neural network model	R ²	0.9996	0.9996	0.9999	0.9999
	RMSE	0.0299	0.0315	0.0310	0.0312
	MAE	0.0238	0.0252	0.0248	0.0249
	MAPE, %	0.27	0.12	0.09	0.15

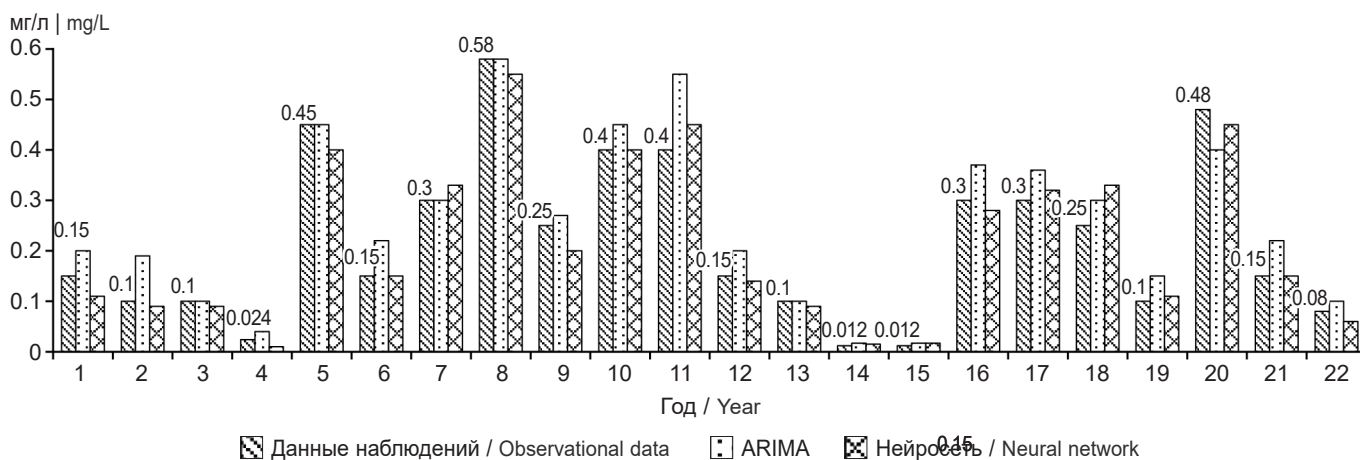


Рис. 3. Сравнение результатов средних концентраций аммония (мг/л) по данным наблюдений (период 1–5 лет, соответствующий 2018–2022 гг.), классического прогноза ARIMA и разработанной нейронной сети для Окского водозабора.

Fig. 3. Comparison of results of mean ammonium concentrations (mg/L) on data (period of 1–5 years corresponding to 2018–2022), classical ARIMA prediction and elaborated neural network for the Okskiy water intake.

Мультисвёрточная LSTM-сеть значительно превзошла предыдущие модели благодаря способности выявлять сложные пространственно-временные паттерны в данных. Комбинация свёрточных слоёв для извлечения локальных признаков и LSTM с целью анализа долгосрочных временных зависимостей позволила достичь высокой точности прогнозирования даже для параметров с выраженной волатильностью и множественными точками перегиба.

Предложенная в исследовании гибридная модель, интегрирующая свёрточные слои с двунаправленной GRU-сетью и «механизмами внимания», показала наилучшие результаты (см. таблицу).

Средняя RMSE для всех параметров составила 0,0309, что на 44,5% ниже, чем у ближайшего конкурента (мультисвёрточной LSTM-сети), и в 32,2 раза ниже, чем у классической модели ARIMA.

Работа реализованной гибридной модели. Гибридная модель работает в трёхэтапной последовательности:

Этап 1. Извлечение признаков: на начальном этапе модель использует свёрточные слои для выявления и извлечения потенциальных нелинейных взаимосвязей между данными временных рядов качества воды. Свёрточные фильтры размером 3×1 и 5×1 эффективно выделяют локальные паттерны в данных (например, резкие изменения концентрации аммиака после дождевых осадков).

Этап 2. Формирование вектора признаков: на основе извлечённых признаков конструируется вектор признаков качества воды, который включает как исходные данные, так и производные характеристики (скользящие средние, PCA-компоненты). Двунаправленная GRU-сеть анализирует и прошлые, и будущие контекстные зависимости, что особенно важно для параметров с запаздывающими эффектами (например, реакция уровня нитратов на сельскохозяйственные стоки).

Этап 3. Генерация прогноза: на заключительном этапе слой полного соединения выполняет функцию выходного слоя для генерации прогнозируемых значений параметров качества воды. «Механизм внимания» адаптивно взвешивает важность различных временных сегментов и параметров.

На рис. 3 представлено сравнение результатов прогнозирования концентрации аммиака для Окского водозабора с использованием классической модели ARIMA и разработанной гибридной нейросети. График демонстрирует превосходство гибридной модели особенно в точках экстремумов, где ARIMA существенно занижает прогнозируемые значения.

Анализ факторов, влияющих на формирование качества воды, выявил доминирующее значение следующих параметров: температуры воды (коэффициент корреляции с содержанием аммиака $r = 0,76$), осадков за предшествующие 48 ч ($r = 0,81$ с ХПК), гидрологического стока ($r = -0,73$ с концентрацией нитратов) и антропогенной нагрузки в зоне водозабора (коэффициент детерминации $R^2 = 0,85$ с общим азотом).

Результаты прогнозирования продемонстрировали высокую гигиеническую значимость. Для ХПК максимальные прогнозируемые значения в период паводков составляли 12,3–14,1 мг O_2 /л при ПДК 15 мг O_2 /л. Концентрация аммиака в весенний период достигала 0,26–0,3 мг/л при гигиеническом нормативе 0,39 мг/л (ориентировочно допустимый уровень). Содержание общего азота варьировало в пределах 15,2–19,7 мг/л при отсутствии прямого норматива, но при критическом уровне 22 мг/л свидетельствовало о выраженной эвтрофикации. Концентрация растворённого кислорода минимально снижалась до 6,3–7,1 мг/л при нормативе не менее 6 мг/л. Сопоставление прогнозных и фактических данных показало, что среднеквадратичная ошибка (RMSE) составила 0,0309, что обеспечивает надёжность прогноза при определении критических точек приближения показателей к санитарным нормативам. Коэффициент детерминации $R^2 > 0,995$ подтверждает высокую статистическую значимость результатов.

Обсуждение

Разработанная гибридная модель продемонстрировала точность прогнозирования параметров качества воды в поверхностных источниках питьевого водоснабжения с коэффициентом детерминации $R^2 > 0,995$ для всех ключевых параметров и средней RMSE = 0,0309, что на порядок превосходит существующие методы.

Модель показала устойчивость к изменениям данных, сохраняла высокую точность прогнозирования для стабильных периодов. Время генерации прогноза на 24 ч вперёд составило в среднем 0,83 с при использовании GPU NVIDIA Tesla V100, что позволяет применять модель в системах реального времени с целью раннего предупреждения о потенциальных рисках для качества воды.

Интеграция «механизма внимания» позволила выявить наиболее значимые факторы, влияющие на изменение каж-

дого параметра качества воды. Например, для концентрации аммиака наибольшее влияние оказывали температура воды (вес 0,38), уровень осадков за предыдущие 48 ч (вес 0,27) и концентрация нитратов (вес 0,21).

Таким образом, разработанная гибридная нейросетевая модель превосходит ряд существующих методов по точности прогнозирования, открывая новые возможности применения в системах оперативного мониторинга и управления качеством водных ресурсов, что имеет фундаментальное значение для обеспечения здоровья населения и безопасности среды обитания.

Гигиеническая значимость полученных результатов заключается в возможности своевременной корректировки технологических режимов водоочистки при прогнозируемом ухудшении качества исходной воды. При приближении концентрации аммиака к 0,3 мг/л и ХПК к 14 мг О₂/л целесообразны увеличение дозы окислителей и активизация процессов биологической очистки на водоочистных сооружениях для предотвращения возможного формирования токсичных хлорорганических соединений при обеззараживании воды и снижения риска развития интоксикаций у потребителей.

Перспективы дальнейших исследований заключаются в расширении набора прогнозируемых показателей за счёт токсичных органических соединений (пестицидов, хлорорганических загрязнителей) и тяжёлых металлов, а также в интеграции модели с системами оценки рисков для здоровья населения в соответствии с методологией, рекомендованной ВОЗ. Это позволит создать комплексную систему раннего предупреждения о потенциальных угрозах безопасности питьевой воды, обеспечивающую защиту здоровья населения как на локальном, так и на территориальном уровнях.

Заключение

Разработанная гибридная нейросетевая модель, интегрирующая свёрточные и двунаправленные рекуррентные архитектуры, продемонстрировала высокую точность прогнозирования параметров качества воды с коэффициентом детерминации $R^2 > 0,995$ и средней RMSE = 0,0309, что на 44,5% превосходит ближайшие аналоги. Данная модель обеспечивает оперативное выявление потенциальных угроз для здоровья населения и открывает новые возможности создания интеллектуальных систем раннего предупреждения о деградации качества водных ресурсов, что имеет фундаментальное значение для обеспечения санитарно-эпидемиологического благополучия территорий.

Проведённые исследования подтвердили, что разработанная гибридная нейросетевая модель является эффективным инструментом санитарно-гигиенического мониторинга качества воды поверхностных источников централизованного водоснабжения. Высокая точность прогнозирования ($R^2 > 0,995$) обеспечивает достоверное определение временных интервалов приближения гигиенически значимых показателей к критическим уровням и имеет первостепенное значение для профилактики неблагоприятного влияния на здоровье населения. Результаты исследования позволяют перейти от ретроспективной оценки качества воды к прогнозированию потенциальных рисков и поэтому имеют фундаментальное значение для системы обеспечения санитарно-эпидемиологического благополучия населения. Это особенно актуально в условиях изменения климата и роста антропогенной нагрузки на водные объекты. Внедрение разработанной модели в практику санитарного надзора за качеством питьевой воды обеспечит переход к превентивным стратегиям обеспечения безопасности водоснабжения, минимизируя вероятность поступления к потребителям некачественной воды.

Литература

(п.п. 1, 3–6, 10, 17–25 см. References)

- Карпенко Н.П., Глазунова И.В., Ширияева М.А. Анализ геоэкологических проблем и оценка обеспеченности питьевыми водами Клинского района Московской области. *Природообустройство*. 2023; (5): 88–94. <https://elibrary.ru/qxbsla>
- Розенталь О.М., Федотов В.Х. Идентификация предприятий-загрязнителей воды на основе нейросетевого анализа. *Природообустройство*. 2023; (1): 62–8. <https://doi.org/10.26897/1997-6011-2023-1-62-68>
- Шамсутдинова Т.М. Применение нейросетевого моделирования в задачах прогнозирования уровня паводка рек. *Вестник Новосибирского государственного университета. Серия: Информационные технологии*. 2023; 21(2): 39–50. <https://doi.org/10.25205/1818-7900-2023-21-2-39-50>
- Шитиков В.К., Зинченко Т.Д., Головатюк Л.В. Нейросетевые методы оценки качества поверхностных вод по гидробиологическим показателям. *Известия Самарского научного центра Российской академии наук*. 2002; 4(2): 280–9. <https://elibrary.ru/gjgkj>
- Карпенко Н.П., Ломакин И.М., Дроздов В.С. Вопросы управления геоэкологическими рисками при оценке качества подземных вод на урбанизированных территориях. *Природообустройство*. 2019; (5): 106–11. <https://doi.org/10.34677/1997-6011/2019-5-106-111>
- Литвинова А.А., Дементьев А.А., Ляпало А.А., Коршунова Е.П. Сравнительная характеристика показателей качества воды реки Оки в местах водозаборов хозяйственно-питьевой системы водоснабжения города Рязани. *Российский медико-биологический вестник имени академика И.П. Павлова*. 2022; 30(4): 481–8. <https://doi.org/10.17816/PAVL0VJ89568>
- Жолдакова З.И., Синицына О.О., Турбинский В.В. О корректировке требований к зонам санитарной охраны источников централизованного хозяйственно-питьевого водоснабжения населения. *Гигиена и санитария*. 2021; 100(11): 1192–7. <https://doi.org/10.47470/0016-9900-2021-100-11-1192-1197>
- Карпенко Н.П., Ширияева М.А. Трёхмерное моделирование как система отображения суммарного химического загрязнения почв. *Природообустройство*. 2021; (1): 6–13. <https://doi.org/10.26897/1997-6011-2021-1-6-14>
- Лагутина Н.В., Новиков А.В., Сумарукова О.В., Науменко Н.О. Оценка качества вод Рыбинского водохранилища вследствие снижения уровня вод. *Природообустройство*. 2019; (2): 122–6. <https://doi.org/10.34677/1997-6011/2019-2-122-126>
- Ширияева М.А., Синицына О.О., Пушкарева М.В., Турбинский В.В. Алгоритм прогнозирования параметров качества водных объектов с использованием нейронной сети. *Анализ риска здоровью*. 2024; (4): 50–62. <https://doi.org/10.21668/health.risk/2024.4.05>

References

- Liao Z., Wang X., Zhang Y., Qing H., Li C., Liu Q., et al. An integrated simulation framework for NDVI pattern variations with dual society-nature drives: A case study in Baiyangdian Wetland, North China. *Ecol. Indic.* 2024; 158: 111584. <https://doi.org/10.1016/j.ecolind.2024.111584>
- Karpenko N.P., Glazunova I.V., Shiryayeva M.A. Analysis of geo ecological problems and assessment of the availability of drinking water in the Klinisky district of the Moscow region. *Природообустройство*. 2023; (5): 88–94. <https://elibrary.ru/qxbsla> (in Russian)
- Shivam K., Tzou J.C., Wu S.C. Multi-step short-term wind speed prediction using a residual dilated causal convolutional network with nonlinear attention. *Energies*. 2020; 13(7): 1772. <https://doi.org/10.3390/en13071772>
- Wu G.D., Lo S.L. Predicting real-time coagulant dosage in water treatment by artificial neural networks and adaptive network-based fuzzy inference system. *Eng. Appl. Artif. Intell.* 2008; 21(8): 1189–95. <https://doi.org/10.1016/j.engappai.2008.03.015>
- Ho J.Y., Afan H.A., El-Shafie A.H., Koting S.B., Mohd N.S., Jaafar W.Z.B., et al. Towards a time and cost effective approach to water quality index class prediction. *J. Hydrol.* 2019; 575: 148–65. <https://doi.org/10.1016/j.jhydrol.2019.05.016>
- Juwana I., Muttill N., Perera B.J.C. Uncertainty and sensitivity analysis of West Java Water Sustainability Index – a case study on Citarum catchment in Indonesia. *Ecol. Indic.* 2016; 61: 170–8. <https://doi.org/10.1016/j.ecolind.2015.08.034>
- Rosenthal O.M., Fedotov V.Kh. Identification of water polluting enterprises based on neural network analysis. *Природообустройство*. 2023; (1): 62–8. <https://doi.org/10.26897/1997-6011-2023-1-62-68>
- Shamsutdinova T.M. Application of neural network modeling in problems of predicting the level of river floods. *Vestnik Novosibirskogo gosudarstvennogo*

Original article

- universiteta. *Seriya: Informatsionnye tekhnologii*. 2023; 21(2): 39–50. <https://doi.org/10.25205/1818-7900-2023-21-2-39-50> <https://elibrary.ru/kzcoem> (in Russian)
9. Shitikov V.K., Zinchenko T.D., Golovatyuk L.V. Methods of neural networks for estimation of superficial waters quality by usage of hydrobiological exponents. *Izvestiya Samarskogo nauchnogo tsentra Rossiiskoi akademii nauk*. 2002; 4(2): 280–9. <https://elibrary.ru/gijgkj> (in Russian)
 10. Sarkar A., Pandey P. River water quality modelling using artificial neural network technique. *Aquat. Procedia*. 2015; 4: 1070–7. <https://doi.org/10.1016/j.aqpro.2015.02.135>
 11. Karpenko N.P., Lomakin I.M., Drozdov V.S. Management issues of geoenvironmental risks in the assessment of groundwater quality in urban areas. *Prirodooobustroistvo*. 2019; (5): 106–11. <https://doi.org/10.34677/1997-6011/2019-5-106-111> <https://elibrary.ru/kbulot> (in Russian)
 12. Litvinova A.A., Dement'yev A.A., Lyapkalo A.A., Korshunova E.P. Comparative characteristics of quality parameters of waters of the Oka River in places of water intake of utility and drinking water system in Ryazan. *Rossiiskii mediko-biologicheskii vestnik imeni akademika I.P. Pavlova*. 2022; 30(4): 481–8. <https://doi.org/10.17816/PAVL0VJ89568> (in Russian)
 13. Zholdakova Z.I., Sinitsyna O.O., Turbinsky V.V. About adjustment of requirements to zones of sanitary protection of sources of the centralized economic and drinking water supply of the population. *Gigiena i Sanitariya (Hygiene and Sanitation, Russian journal)*. 2021; 100(11): 1192–7. <https://doi.org/10.47470/0016-9900-2021-100-11-1192-1197> <https://elibrary.ru/ybmblm> (in Russian)
 14. Karpenko N.P., Shiryayeva M.A. Three-dimensional modeling as a system for displaying total chemical soil pollution. *Prirodooobustroistvo*. 2021; (1): 6–13. <https://doi.org/10.26897/1997-6011-2021-1-6-14> <https://elibrary.ru/xbuuer> (in Russian)
 15. Lagutina N.V., Novikov A.V., Sumarukova O.V., Naumenko N.O. Assessment of the water quality of the Rybinsk reservoir as a result of the water level lowering. *Prirodooobustroistvo*. 2019; (2): 122–6. <https://doi.org/10.34677/1997-6011/2019-2-122-126> <https://elibrary.ru/yjrviv> (in Russian)
 16. Shiryayeva M.A., Sinitsyna O.O., Pushkareva M.V., Turbinsky V.V. Algorithm for predicting water quality indicators in water bodies using a neural network. *Health Risk Analysis*. 2024; (4): 50–62. <https://doi.org/10.21668/health.risk/2024.4.05.eng>
 17. Liu H., Zhang F., Tan Y., Huang L., Li Y., Huang G., et al. Multi-scale quaternion CNN and BiGRU with cross self-attention feature fusion for fault diagnosis of bearing. *Meas. Sci. Technol.* 2024; 35(8): 086138. <https://doi.org/10.1088/1361-6501/ad4c8e>
 18. Jiang Y., Li C., Sun L., Guo D., Zhang Y., Wang W. A deep learning algorithm for multi-source data fusion to predict water quality of urban sewer networks. *J. Clean. Prod.* 2021; 318: 128533. <https://doi.org/10.1016/j.jclepro.2021.128533>
 19. Veerendra G.T.N., Kumaravel B., Kodanda Rama Rao P., Dey S., Phani Manoj A.V. Forecasting models for surface water quality using predictive analytics. *Environ. Dev. Sustain.* 2024; 26(6): 15931–51. <https://doi.org/10.1007/s10668-023-03280-3>
 20. Chen X., Jiang Z., Cheng H., Zheng H., Cai D., Feng Y. A novel global average temperature prediction model – based on GM-ARIMA combination model. *Earth Sci. Inform.* 2023; 17(1): 853–66. <https://doi.org/10.1007/s12145-023-01179-1>
 21. Jiao G., Chen S., Wang F., Wang Z., Wang F., Li H., et al. Water quality evaluation and prediction based on a combined model. *Appl. Sci.* 2023; 13(3): 1286. <https://doi.org/10.3390/app13031286>
 22. Da Silva A.C., das Graças Braga da Silva F., de Mello Valério V.E., Lima Silva A.T.Y., Marques S.M., Tosta dos Reis J.A. Application of data prediction models in a real water supply network: comparison between arima and artificial neural networks. *Rev. Bras. Recur. Hidr.* 2024; 29: e12. <https://doi.org/10.1590/2318-0331.292420230057>
 23. Deng T., Chau K.W., Duan H.F. Machine learning based marine water quality prediction for coastal hydro-environment management. *J. Environ. Manage.* 2021; 284: 112051. <https://doi.org/10.1016/j.jenvman.2021.112051>
 24. Lu X., Dong Y., Liu Q., Zhu H., Xu X., Liu J., et al. Simulation on TN and TP distribution of sediment in Liaohu estuary national wetland park using mike21-coupling model. *Water*. 2023; 15(15): 2727. <https://doi.org/10.3390/w15152727>
 25. Kim J., Seo D., Jang M., Kim J. Augmentation of limited input data using an artificial neural network method to improve the accuracy of water quality modeling in a large lake. *J. Hydrol.* 2021; 602(4): 126817. <https://doi.org/10.1016/j.jhydrol.2021.126817>

Сведения об авторах

Ширяева Маргарита Александровна, мл. науч. сотр. отд. гигиены воды ФБУН «ФНЦГ им. Ф.Ф. Эрисмана» Роспотребнадзора, 141014, Мытищи, Россия. E-mail: shiryayeva.ma@fncg.ru

Пушкарёва Мария Васильевна, доктор мед. наук, профессор, гл. науч. сотр. отд. гигиены воды ФБУН «ФНЦГ им. Ф.Ф. Эрисмана» Роспотребнадзора, 141014, Мытищи, Россия. E-mail: pushkareva.mv@fncg.ru

About the authors

Margarita A. Shiryayeva, junior researcher, Department of water hygiene, Federal Scientific Center of Hygiene named after F.F. Erisman, Mytishchi, 141014, Russian Federation, <https://orcid.org/0000-0001-8019-1203> E-mail: shiryayeva.ma@fncg.ru

Maria V. Pushkareva, DSc (Medicine), professor, chief researcher, Department of water hygiene, Federal Scientific Center of Hygiene named after F.F. Erisman, Mytishchi, 141014, Russian Federation, <https://orcid.org/0000-0002-5932-6350> E-mail: pushkareva.mv@fncg.ru