

Научная статья

АЛГОРИТМ ПРОГНОЗИРОВАНИЯ ПАРАМЕТРОВ КАЧЕСТВА ВОДНЫХ ОБЪЕКТОВ С ИСПОЛЬЗОВАНИЕМ НЕЙРОННОЙ СЕТИ

М.А. Ширяева^{1,2}, О.О. Сеницына¹, М.В. Пушкарева¹, В.В. Турбинский¹

¹Федеральный научный центр гигиены имени Ф.Ф. Эрисмана, Российская Федерация, 141014, Московская область, г. Мытищи, ул. Семашко, 2

²Российский государственный аграрный университет – Московская сельскохозяйственная академия имени К.А. Тимирязева, Российская Федерация, 127550, г. Москва, ул. Тимирязевская, 49

Чистая и безопасная питьевая вода является фундаментальной необходимостью для здоровья и благополучия людей, а также важным компонентом устойчивого развития экосистем. В последние десятилетия проблемы качества воды стали еще более актуальными в связи с ростом численности населения, расширением промышленной деятельности и изменением климата.

В ряде работ зарубежных исследователей показаны результаты применения нейросетей. Имеются исследования, подтверждающие достоверность результатов прогноза качества воды, сгенерированного нейросетями.

В ходе работы использовались программы Google Earth Pro, Microsoft Excel, датчик расхода воды на базе платы Arduino UNO с авторской доработкой (хвостовым оперением и встроенным плагин для расчета скорости течения), Python, библиотеки Tensorflows keras2.2.0, Scikit-learn, Pandas для машинного обучения и разработки архитектуры нейросети. В данной работе две модели нейросети были объединены для построения гибридной нейросетевой модели прогнозирования параметров качества воды.

Нейросетевые модели предоставляют уникальные возможности для улучшения управления водными ресурсами на различных уровнях, начиная от местных и заканчивая глобальными. Одним из ключевых преимуществ таких моделей является возможность адаптации к конкретным условиям и требованиям, что обеспечивает более точное предсказание и своевременное принятие решений в условиях неопределенности. Актуальность работы обусловлена применением нейросетей для прогнозирования качества воды, что может способствовать улучшению системы раннего предупреждения о загрязнении, оптимизации операционных процессов на водоочистных станциях и разработке эффективных стратегий по управлению водными ресурсами.

В ходе исследований разработана инновационная гибридная нейросетевая модель прогнозирования параметров качества воды, основанная на интеграции глубокой сверточной нейронной сети и двунаправленной рекуррентной нейросети, которая состоит из трех функциональных частей.

Ключевые слова: нейросеть, Tensorflows keras2.2.0, водные объекты, питьевая вода, фактор риска, негативное воздействие, загрязнение водной среды, коэффициент детерминации, алгоритм оптимизации.

Оценка состояния водных ресурсов играет исключительно важную роль в современном обществе, особенно на фоне прогрессирующей антропогенной нагрузки на водные экосистемы и проявлений изменения климата [1–3]. Реки, озера, водохранилища выступают основными источниками централизованного водоснабжения населения питьевой водой, ключевым объектом орошения сельскохозяйствен-

ных угодий, основным источником водных ресурсов для нужд промышленности, а также занимают существенное место в инфраструктуре отдыха и рекреации населения [4, 5].

В последние десятилетия в связи с интенсификацией антропогенного воздействия и увеличением масштабов загрязнения произошло значительное снижение общего качества воды в водных объектах

© Ширяева М.А., Сеницына О.О., Пушкарева М.В., Турбинский В.В., 2024

Ширяева Маргарита Александровна – младший научный сотрудник отдела гигиены воды (e-mail: Shiryaeva.MA@fncg.ru; тел.: 8 (903) 161-14-04; ORCID: <https://orcid.org/0000-0001-8019-1203>).

Сеницына Оксана Олеговна – член-корреспондент РАН, доктор медицинских наук, профессор, заместитель директора по научной работе, директор Института комплексных проблем гигиены (e-mail: sinityna.oo@fncg.ru; тел.: 8 (926) 447-08-74; ORCID: <http://orcid.org/0000-0002-0241-0690>).

Пушкарева Мария Васильевна – доктор медицинских наук, профессор, главный научный сотрудник отдела гигиены воды (e-mail: pushkareva.mv@fncg.ru; тел.: 8 (912) 980-92-74; ORCID: <https://orcid.org/0000-0002-5932-6350>).

Турбинский Виктор Владиславович – доктор медицинских наук, заведующий отделом гигиены воды (e-mail: turbinskii.vv@fncg.ru; тел.: 8 (920) 666-72-73; ORCID: <https://orcid.org/0000-0001-7668-9324>).

[6]. Это обстоятельство обуславливает необходимость разработки и внедрения инновационных подходов к мониторингу и прогнозированию состояния гидросреды, превосходящих традиционные методы по степени точности, надежности и оперативности получения результатов [7].

Одним из наиболее перспективных направлений в этом контексте выступает использование машинного обучения с применением нейронных сетей для моделирования и прогноза динамики факторов, детерминирующих качество воды. Подобные модели способны учитывать сложнейшие нелинейные взаимосвязи многочисленных воздействующих факторов и обладают свойством самообучения, что делает их высокоэффективным инструментом решения поставленных задач [8, 9].

Цель исследования – разработка и применение инновационного алгоритма прогнозирования параметров качества водных объектов с использованием нейронной сети. Достижение этой цели осуществлялось поэтапно путем последовательного выполнения ряда изучающих и совершенствующих этапов. Первостепенной задачей стало всестороннее изучение уже существующих методов оценки качества воды. Детальный анализ научных трудов позволил выявить сильные и слабые стороны разных подходов, оценить их эффективность в различных природных условиях. Благодаря этому удалось определить наиболее перспективные направления для разработки новаторского алгоритма. Следующим этапом явилось проектирование собственной нейронной сети, подобранных входных параметров и обучение ее на реальных данных. При этом учитывался химический состав воды. Благодаря калибровке алгоритма удалось добиться максимальной точности прогнозов. Далее осуществлялась оценка эффективности разработанной нейронной модели. Проводились сравнительные испытания, при которых результаты прогнозирования сопоставлялись с фактическими данными и показателями традиционных подходов. Это позволило установить преимущества нового алгоритма и области его наибольшей целесообразности. Завершающим этапом стало сопоставление нового подхода с классическими методами, посредством чего выявляли достоинства и недостатки предложенной модели. В заключении сформулированы рекомендации по практическому применению алгоритма, определены перспективы дальнейшего совершенствования разработки. Данное исследование позволяет сделать важный вклад в решение стратегической задачи прогнозирования качества водных ресурсов.

Материалы и методы. Исследования, направленные на изучение санитарного состояния и динамики качественных параметров воды, проводились на участке русла реки Оки, географически приуроченном к городской агломерации Рязани, представляющей собой репрезентативный пример взаимодействия крупного водотока с высокоурбанизированной территорией.

Для обеспечения эффективного прогноза качества воды Оки, позволяющего заблаговременно выявлять негативные тенденции и принимать предупредительные меры по предотвращению санитарно-эпидемиологического неблагополучия, необходимо проводить комплексный и последовательный мониторинг на анализируемом водотоке. Ключевую роль в обеспечении высокой эффективности и репрезентативности системы мониторинга играет оптимальное расположение контрольных пунктов для сбора данных, которое должно учитывать как особенности гидрологического режима реки, так и пространственное распределение потенциальных источников загрязнения [10]. На рис. 1 авторами предложена примерная схема размещения модульных метеостанций, интегрированных с системами автоматического отбора проб воды, где красные кружки обозначают расположение измерительных приборов, зеленый сектор охватывает территорию наблюдения в 400 км², а голубой сектор демонстрирует зону перекрытия наблюдений двух соседних станций, что обеспечивает необходимую избыточность данных для повышения достоверности результатов.

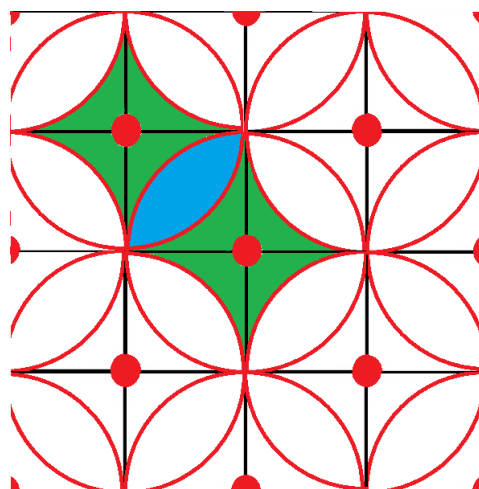


Рис. 1. Схема расположения модульных метеостанций

Предложенная схема расположения контрольно-измерительных пунктов была разработана с целью оптимизации экономической составляющей при сохранении максимально возможного уровня эффективности системы систематического мониторинга, что представляет особую важность в условиях жестких бюджетных ограничений на финансирование природоохранных программ. Применение математического моделирования и методов оптимизации с учетом гидрологической специфики водного объекта, характеристик распределения источников антропогенного воздействия и экономических ограничений позволяет с высокой степенью достоверности определить необходимое и достаточное количество измерительных установок для эффективного охвата всей водосборной площади исследуемого участка русла реки Оки, обеспечивающего систематический

мониторинг и прогнозирование ее санитарного состояния с учетом многофакторной динамики внешних воздействий [11].

Расстояния между модульными станциями составят:

$$l_1 = 21,6\sqrt{2} = 30,5 \text{ км};$$

и

$$l_2 = 21,6 \cdot 2 = 43,2 \text{ км}.$$

Теперь можно определить площадь пересечения наблюдений двух станций:

$$F = 2 \cdot \left(\frac{21,6^2 \cdot \Pi}{4} - \frac{21,6^2}{2} \right) = \\ = \frac{21,6^2 \cdot \Pi}{2} - \frac{2 \cdot 21,6^2}{2} = \frac{21,6^2 (\Pi - 2)}{2} \approx 266 \text{ км}^2.$$

На основе анализа представленной на рисунке концептуальной модели перекрытия зон действия гидрометрических станций можно сделать следующие теоретические умозаключения. Каждая две соседние станции имеют по одной точке пересечения границ их измерительных зон, три станции – две точки пересечения и так далее в соответствии с установленной закономерностью.

Используя данную тенденцию, представляется возможным спрогнозировать необходимое количество наблюдательных постов для полного охвата контрольных измерений по всей площади рассматриваемого водосборного бассейна. Для количественной оценки требуемой величины предлагается построить математическую модель путем постановки уравнения¹, в котором исходным параметром (n) будет численность станций, обеспечивающая покрытие территории площадью 400 км^2 .

Тогда количество точек пересечения зон действия между станциями на участке площадью 266 км^2 можно записать как величину $(n - 1)$. Решив данное уравнение, представляется возможным определить минимально необходимую численность гидрометрических постов для формирования регулярной сетки наблюдений и обеспечения качественного гидрометеорологического мониторинга на всей территории водосбора, а также проектирования оптимальной конфигурации сети наблюдений.

Отсюда получается уравнение:

$$400n + 266(n - 1) = 245\,000 \text{ км}^2,$$

$$n \approx 368 \text{ единиц}.$$

С учетом полученного результата математического моделирования предельно необходимого количества гидрометрических станций, можно сделать

вывод, что для пространственного покрытия всей акватории бассейна реки Оки с целью качественного мониторинга за его климатическими, гидрохимическими и микробиологическими характеристиками будет целесообразно развернуть сеть из 368 единиц универсальных модульных гидрологических постов.

В процессе проведения исследования и сбора первичных данных использовался комплекс современных программных средств и инновационного оборудования, включающий в себя геоинформационную систему Google Earth Pro для пространственного анализа и визуализации данных, табличный процессор Microsoft Excel для статистической обработки и предварительного анализа полученных результатов, надводный беспилотный аппарат авторской разработки, оснащенный системой автоматического отбора проб воды, высокоточный эхолот Garmin Striker Cast GPS для измерения глубин и составления батиметрических карт исследуемого участка реки, а также инновационный датчик расхода воды, созданный на базе микроконтроллерной платы Arduino UNO с авторскими модификациями, включающими оптимизированное хвостовое оперение для стабилизации положения в потоке и интегрированный программный плагин для расчета скорости течения на основе измеренных параметров потока.

Для повышения точности измерений и расширения функциональных возможностей датчика расхода воды, разработанного на базе платы Arduino UNO, был создан специализированный программный плагин, в основу которого легла математически выведенная формула преобразования данных о расходах воды в показатели скорости течения, учитывающая геометрические параметры датчика, в частности, диаметр входного и выходного отверстий, составивший $11,9 \text{ мм}$, что обеспечило оптимальное соотношение между чувствительностью прибора и его устойчивостью к засорению взвешенными частицами.

Соответственно для определения скорости течения (m/c) из расхода воды ($л/с$) в плагин преобразования данных прописаны следующие формулы:

$$V = \frac{4W}{\pi \cdot D^2 \cdot 1000}, \quad (1)$$

$$V = \frac{4W}{\pi \cdot 0,0119^2 \cdot 1000},$$

где $\pi = 3,14$, W – исходные данные датчика расхода воды ($л/с$), D – диаметр сечения входного и выходного отверстия датчика ($мм$).

Разработанная в результате исследования комплексная модель машинного обучения будет в первую очередь способствовать повышению эффектив-

¹ Евграфов А.В. Метрология, стандартизация и сертификация: учебно-методическое пособие. – М.: РГАУ-МСХА, 2015. – 83 с.

ности прогнозирования состояния поверхностных вод как стратегически важного источника водоснабжения, оценивая его соответствие, прежде всего, гигиеническим нормативам, а также предоставлять научно обоснованные рекомендации промышленным предприятиям, сельскохозяйственным комплексам и иным потенциальным источникам загрязнения по разработке и внедрению мер, направленных на минимизацию негативного воздействия на водный объект и снижение объемов поступления загрязняющих веществ [12, 13].

Данные по химическому составу воды были получены в ходе проведения лабораторных исследований.

Предложенный в рамках данного исследования алгоритм комплексного прогнозирования качества воды включает в себя следующие последовательные этапы.

Шаг 1. Очистка данных. Перед непосредственным прогнозированием качества воды применяется метод iForest для выявления аномальных значений в массиве данных о качестве воды $X_{n \times m}$ (где n обозначает количество параметров качества воды, а m – количество групп данных; в контексте данной работы n и m являются постоянными величинами: $n = 9$, $m = 1360$), при этом выявленные аномальные значения заменяются на пустые значения. Впоследствии для заполнения пустых значений используется метод интерполяции Лагранжа, обеспечивающий целостность и непрерывность данных [14, 15].

Шаг 2. Расширение данных. На первом этапе из массива $X_{n \times m}$ удаляется прогнозируемая цель, в результате чего формируется новый массив $X_{n \times (m-1)}$. Учитывая, что данные о качестве воды собираются с интервалом в 4 ч, применяется метод усреднения в скользящем окне с размером окна 6 для формирования набора скользящих средних $Z_{n \times (m-1)}$, что позволяет минимизировать влияние случайных факторов вариации данных о качестве воды и более точно отследить тенденцию суточного изменения параметров качества воды. На втором этапе используется технология анализа главных компонент (PCA) для уменьшения размерности $X_{n \times (m-1)}$ и сохранения двух главных компонент $P_{2 \times m}$. С целью предотвращения переобучения модели на вход модели одновременно подаются $Z_{n \times (m-1)}$, $P_{2 \times m}$ и данные о качестве воды $X_{n \times (m-1)}$ без целевых параметров, в то время как на выходе модели формируется целевое предсказание.

Шаг 3. Обучение модели. Имеющийся массив данных о качестве воды разделяется на обучающий и тестовый наборы в соотношении 8: 2. В рамках данного исследования обучающий набор включал 1100 наборов данных, охватывающих период с 25 июня 2021 г. по 16 февраля 2022 г., в то время как тестовый набор содержал 272 набора данных, собранных в период с 17 февраля 2022 г. по 1 апреля 2022 г. Учитывая долгосрочную зависимость данных о качестве воды от временных факторов, применяется метод скользящего окна [16, 17] для разде-

ления обучающего набора на фиксированные обучающие окна с шагом длиной i во временной последовательности, после чего данные первых j обучающих окон используются для прогнозирования $(j+1)$ -го обучающего окна. При каждом новом цикле обучения наиболее старое окно исключается из рассмотрения, а следующее новое окно включается в анализ, и этот процесс продолжается до достижения последнего окна обучения. Такой подход, предполагающий отбрасывание устаревших данных, способствует обучению модели с учетом будущих тенденций. На заключительном этапе, в соответствии с тестовым набором каждой станции, обученная модель применяется для прогнозирования ключевых параметров качества воды, включая общий азот, общий фосфор и перманганатную окисляемость.

В рамках проведенного комплексного исследования была осуществлена всесторонняя оценка эффективности предложенной гибридной нейросетевой модели прогнозирования параметров качества воды, которая включала в себя сравнительный анализ с референтными методами, применяемыми в данной области. С целью получения количественной характеристики точности прогнозирования исследователи использовали ряд общепринятых в научном сообществе метрик, среди которых: средняя абсолютная ошибка (MAE), отражающая среднее отклонение прогнозируемых значений от фактических; средняя абсолютная процентная ошибка (MAPE), позволяющая оценить относительную величину ошибки прогноза; среднеквадратическая ошибка (RMSE), учитывающая квадрат отклонений и придающая больший вес крупным ошибкам; а также коэффициент детерминации (R^2), характеризующий долю дисперсии зависимой переменной, объясняемую моделью [18].

На начальном этапе исследования с применением метода изоляционного леса (iForest), который является эффективным алгоритмом обнаружения аномалий в многомерных данных, были идентифицированы и количественно оценены выбросы в исходных данных о качестве воды на исследуемых станциях, составившие приблизительно 1,1, 1,7 и 3,2 % от общего объема данных соответственно. После тщательного удаления выявленных выбросов, которые могли бы существенно повлиять на точность модели, оставшиеся пропущенные значения для станций 1–3 составили около 3,9, 4,5 и 5 % соответственно, что потребовало применения методов восстановления данных, в частности, была использована интерполяция Лагранжа, позволяющая восстановить непрерывную функцию по дискретному набору точек.

Для оценки допустимости применения разработанной модели в сравнение приводилась проверка традиционных прогнозных моделей ARIMA и SMA.

Модели ARIMA (Autoregressive Integrated Moving Average) и SMA (Simple Moving Average) являются популярными методами прогнозирования

временных рядов, в том числе и качества воды. ARIMA модели учитывают автокорреляцию и авторегрессию в данных, что позволяет им захватить динамику изменения качества воды во времени. SMA подходит для прогнозирования качества воды с более стабильными временными рядами и меньшим количеством выбросов.

Результаты и их обсуждение. В рамках проведенного комплексного исследования была осуществлена оценка качественных характеристик водных ресурсов реки Оки, рассматриваемой в качестве поверхностного источника водоснабжения, основанная на анализе средних многолетних значений 52 контрольных показателей, включающих органолептические, микробиологические и химические параметры за продолжительный период с 2014 по 2022 г. Результаты исследования показали, что вода, отобранная на Соколовском водозаборе, характеризовалась значительно меньшей средней многолетней концентрацией аммиака, составившей 0,48 мг/л, что оказалось существенно ниже аналогичных показателей, зафиксированных на Окском и Борковском водозаборах, – в 1,6 и 2,1 раза соответственно ($p < 0,05$). При этом следует отметить, что практически в каждой пятой пробе, отобранной в створе Борковского водозабора, были зарегистрированы концентрации аммиака, превышающие предельно допустимую концентрацию (ПДК), в то время как аналогичный показатель для Окского водозабора оказался в 2,8 раза меньше и составил 7,5 %. Примечательно, что ни в одной из разовых проб воды, отобранных в контрольной точке Соколовского водозабора, превышений ПДК ионов аммония зафиксировано не было. В ходе статистического анализа не было выявлено значимых различий в средних многолетних значениях химического потребления кислорода (ХПК) и биохимического потребления кислорода (БПК₅) в водах рассматриваемых водозаборов, при этом процент разовых проб, в которых указанные показатели не соответствовали установленным гигиеническим требованиям, варьировался в пределах 22,7–32,5 и 61,8–75,0 % соответственно. Проведенное исследование также продемонстрировало, что среднее содержание общих колиформных бактерий (ОКБ) в водах Окского и Борковского водозаборов составило 813,3 и 818,9 КОЕ/100 мл соответственно, что оказалось в 1,5 раза выше аналогичного показателя, зафиксированного в контрольном створе Соколовского водозабора ($p < 0,05$). Некоторые результаты анализов показаны на графике (рис. 2).

В рамках данной научной работы была разработана инновационная гибридная нейросетевая модель прогнозирования параметров качества воды, основанная на интеграции глубокой сверточной нейронной сети и двунаправленной (bidirectional) рекуррентной нейросети, которая состоит из трех функциональных частей. На начальном этапе модель применяется для выявления и извлечения потенциальных нелинейных взаимосвязей между дан-

ными временного ряда качества воды реки Оки с целью формирования эффективных низкоразмерных признаков. Далее на основе извлеченных признаков конструируется вектор признаков качества воды, который используется в качестве входного сигнала для глубокой сверточной нейронной сети. В процессе обучения сеть непрерывно осуществляет регулировку весов и смещений, учитывая зависимость краткосрочных, долгосрочных и контекстных атрибутов временного ряда данных для дальнейшей оптимизации информации о качестве воды с целью

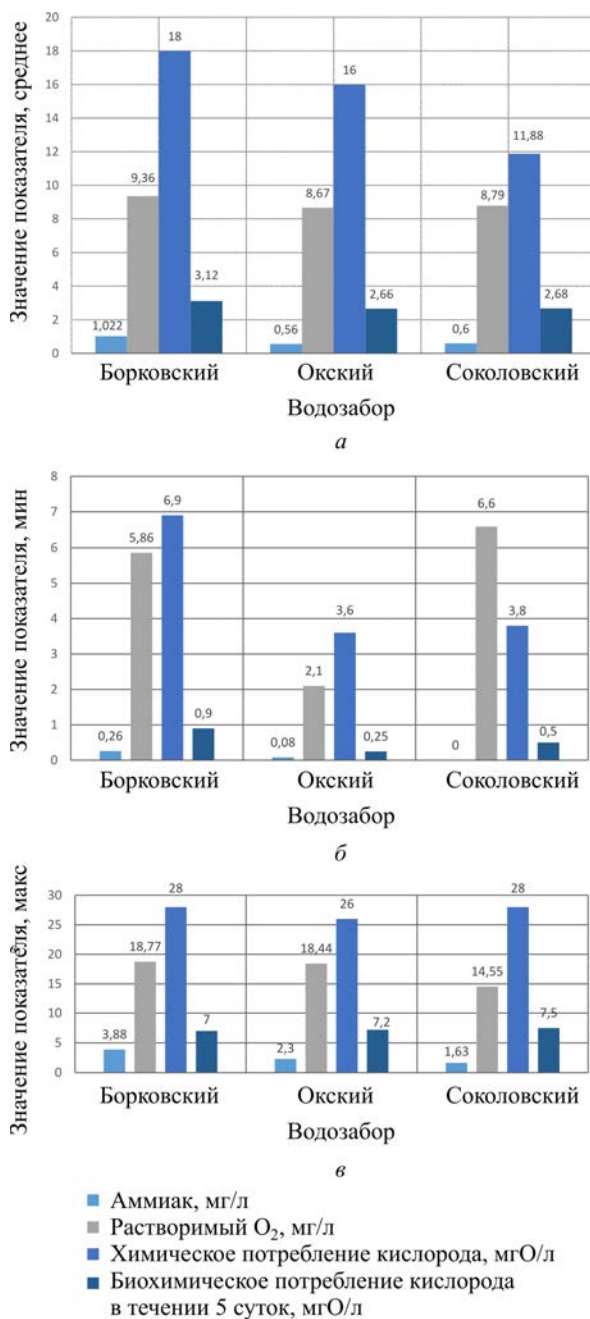


Рис. 2. Некоторые показатели качества воды р. Оки трех исследуемых водозаборов на период 2014–2022 гг.: а – средние значения, б – минимальные значения, в – максимальные значения

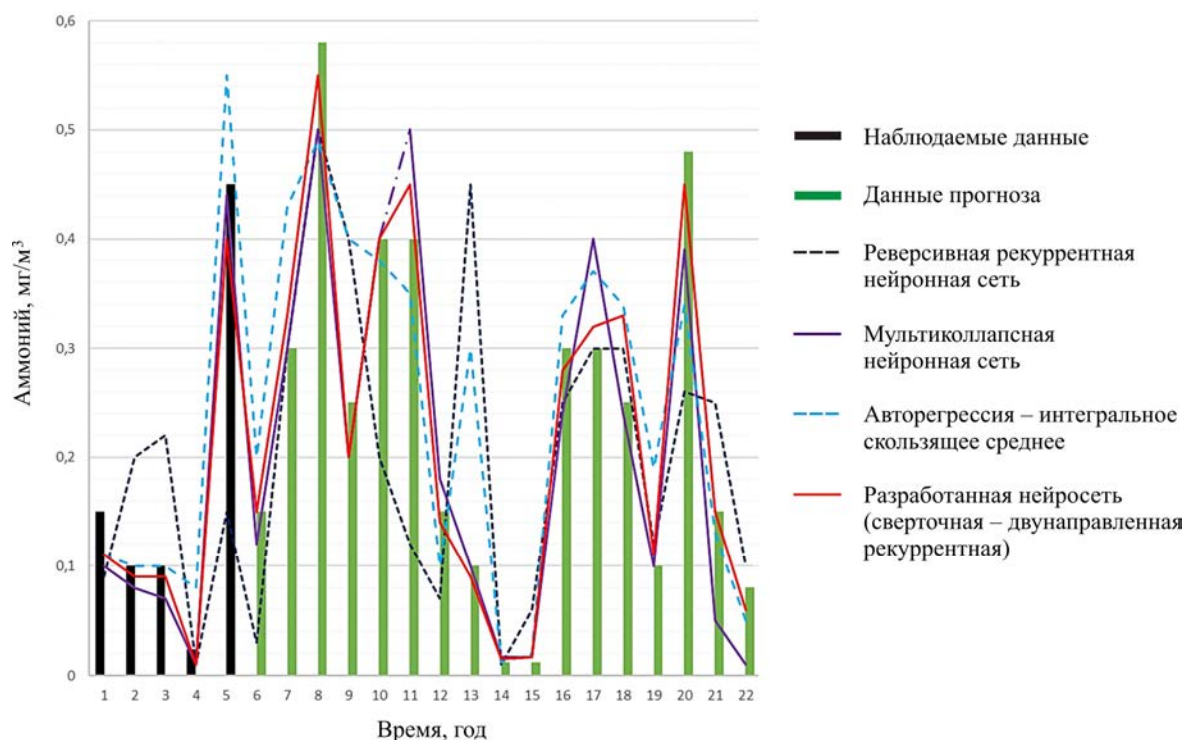
более точного выражения признаков. На заключительном этапе в верхней части модели подключается слой полного соединения, выполняющий функцию выходного слоя для генерации прогнозируемых значений параметров качества воды.

Программная реализация разработанной гибридной нейросетевой модели прогнозирования была осуществлена с использованием высокопроизводительной библиотеки глубокого обучения Tensorflows keras версии 2.2.0, которая предоставляет широкий спектр инструментов для построения и обучения нейронных сетей. Процесс обучения модели проводился на протяжении 50 эпох с использованием 120 временных интервалов, что позволило достичь оптимального баланса между точностью прогнозирования и вычислительными затратами. В качестве алгоритма оптимизации для корректировки весов и смещений модели был применен метод Adam, который сочетает в себе преимущества адаптивных методов градиентного спуска и метода моментов. По достижении сходимости модели, что свидетельствовало о минимизации функции потерь, были получены финальные весовые коэффициенты, которые впоследствии использовались для прогнозирования качества воды на исследуемых станциях водозабора (Соколовский, Окский и Борковский). Архитектура и параметры модели были тщательно подобраны и заданы следующим образом: количество скрытых слоев составило два, что позволило модели эффек-

тивно улавливать сложные нелинейные зависимости в данных; в качестве алгоритма оптимизации был выбран метод сопряженных градиентов, известный своей способностью быстро сходиться к оптимальному решению; минимальное относительное изменение коэффициента ошибки обучения было установлено на уровне 0,001, что обеспечило баланс между точностью модели и предотвращением переобучения.

Это дает возможность использовать представленную нейросетевую модель для восполнения пробелов в данных путем расчета отсутствующих значений концентрации определенных соединений. Некоторые результаты моделирования с применением нейронных сетей, охватывающие 25-летний интервал наблюдений ($n = 25$), демонстрируются в качестве примера на рис. 3 для показателя аммония Окского водозабора. Полученные закономерности могут быть использованы для прогноза динамики исследуемых показателей в перспективе.

Для определения допустимости применения нейросетевой модели разработанная нейросеть сравнивалась с классическим методом оценки: скользящего среднего (рис. 3, observed forecast data). Данный метод выбран, исходя из того, что он является распространенной техникой для анализа временных рядов, таких как концентрация нитратов в воде [19, 20]. Он помогает выявить тренды, уменьшая влияние случайных колебаний и шума в данных.



Метод скользящего среднего (СМ) заключается в том, что для каждого момента времени вычисляется среднее значение некоторого количества предыдущих точек данных. Формула для простого скользящего среднего выглядит следующим образом:

$$SMA_t = \frac{X_{t-n+1} + X_{t-n+2} + \dots + X_t}{n}, \quad (2)$$

где SMA_t – значение скользящего среднего в момент времени t ;

X – наблюдаемые значения (например, концентрация нитратов);

n – количество периодов (интервал) для сглаживания.

Коэффициент α определяет вес, который придается последнему наблюдению: более высокие значения α ведут к более быстрому реагированию на изменения в данных.

В контексте анализа концентрации загрязняющих веществ применялись оба метода для сглаживания данных по временным рядам и выявления долгосрочных трендов:

1. Скользящее среднее делает данные о концентрации более стабильными и поможет визуализировать тенденции изменения во времени.

2. Экспоненциальное сглаживание обеспечит более быстрое реагирование на изменения в концентрации, что особенно полезно, если данные подвержены резким колебаниям.

Помимо SMA_t (рис. 3, *observed forecast data*) применяли классический метод прогнозирования с помощью ARIMA (авторегрессионная интегрированная скользящая средняя или модель статистического анализа, которая использует данные временных рядов для прогноза будущих значений в ряде).

В рамках настоящего исследования предлагается к рассмотрению альтернативный подход к прогнозированию качества воды, базирующийся на применении нейронных сетей для анализа обширных массивов исторических данных, что представляет собой принципиально иной метод по сравнению с традиционными механистическими моделями, широко используемыми в данной области. Механистические модели качества воды, к числу которых относятся такие известные системы, как QUAL, WASP, MIKE, SWAT, BASINS и ряд других, основываются на детальном описании структуры исследуемой водной системы и учете многочисленных ограничений, связанных с комплексом физических, биологических и химических процессов, протекающих в водной среде, что обуславливает их высокую сложность и требует значительного объема исходной информации для создания и последующего решения системы уравнений, описывающих динамику изменения качества воды во времени и пространстве [21–23].

Несмотря на широкое распространение и признание в научном сообществе, механистические модели характеризуются высокой степенью сложности

и требуют значительного объема исходных данных, включающих многочисленные параметры моделирования, условия источников и стоков загрязняющих веществ, а также другие специфические характеристики водной системы, что делает процесс построения таких моделей чрезвычайно трудоемким, а определение оптимальных параметров – затруднительным, существенно ограничивая их применимость для широкого спектра водных объектов, особенно в условиях недостатка детальной информации о их гидрологическом режиме и санитарном состоянии [24, 25].

Рассматриваемая нейронная модель, построенная на основе современных глубоких архитектур, показала высокую эффективность при решении данной задачи. Заложенные в основу модели нелинейные многослойные механизмы обработки данных позволяют выявлять сложные взаимозависимости между показателями качества воды и внешними факторами, формируя при этом статистически значимые предсказания. Проведенные исследования подтвердили высокую степень достоверности прогнозов, что обусловлено способностью модели эффективно анализировать и прогнозировать нелинейные процессы в условиях неопределенности.

Кроме того, модель характеризуется высокой универсальностью, позволяя применять ее для различных типов водных объектов, включая реки, озера и водохранилища. Это значительно расширяет сферу потенциального использования модели в целях мониторинга качества воды и управления водными ресурсами. Несомненным является преимущество предлагаемой нейронной модели по сравнению с традиционными численными алгоритмами, обусловленное более высокой точностью прогнозов и вычислительной эффективностью. Данная модель открывает новые возможности для разработки перспективных подходов в области мониторинга и управления водными ресурсами.

Прогнозирование концентраций нитратов с использованием моделей временных рядов, таких как модель ARIMA (Autoregressive Integrated Moving Average), представляет собой мощный метод анализа данных, который позволяет учитывать тренды, сезонность и автокорреляцию. В данной методике мы рассмотрим шаги, необходимые для реализации модели ARIMA, и представим соответствующие формулы. За последние несколько десятилетий исследования в области прогнозирования временных рядов в основном были сосредоточены на двух подходах [26]. Один подход основан на математической статистике, например, на моделях авторегрессии, интегрированного скользящего среднего (ARIMA). Когда набор данных временных рядов невелик, эти модели часто имеют преимущества, поскольку требуют относительно меньшего объема данных для оценки параметров модели [21].

Собраны исторические данные о концентрациях нитратов в створах реки. Данные представлены в виде временного ряда, где каждый элемент соответ-

стует концентрации в определенный момент времени. ARIMA обозначается как ARIMA (p, d, q), где p – порядок авторегрессии, d – уровень дифференцирования, q – порядок скользящего среднего.

Использовались графики ACF (Autocorrelation Function) и PACF (Partial Autocorrelation Function) для определения значений p и q . ACF показывает автокорреляцию временного ряда на разных лагах. Если ACF быстро убывает, то q может быть мал. PACF показывает частичную автокорреляцию и может помочь определить p .

Далее строилась модель ARIMA [27]:

$$Y_t = \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \dots + \Phi_p Y_{t-p} + \Theta_1 e_{t-1} + \Theta_2 e_{t-2} + \dots + \Theta_q e_{t-q} + e_t, \quad (3)$$

где Φ – коэффициенты модели AR, Θ – коэффициенты модели MA, e_t – шум ошибки.

Параметры модели оценивались с помощью метода максимального правдоподобия. Проверка качества модели осуществлялась с применением статистических критериев AIC (Akaike Information Criterion) и BIC (Bayesian Information Criterion).

Затем проводился этап прогнозирования с помощью оцененной модели ARIMA:

$$\hat{Y}_{t+h} = \hat{Y}_t + \sum_{i=1}^p \Phi_i \hat{Y}_{t-i} - \sum_{j=1}^q \Theta_j e_{t-j}, \quad (4)$$

где h – горизонт прогноза.

Эта методика предоставляет структурированный подход к прогнозированию концентраций загрязнителей в водозаборах реки с использованием модели ARIMA [27, 28]. Управление качеством воды может значительно улучшиться за счет точных прогнозов, что позволит принимать более обоснованные решения для защиты экосистем и здоровья населения.

Чтобы оценить преимущества и недостатки предложенной модели прогнозирования и других эталонных нейросетей, таких как LSTM и реверсивная рекуррентная, сравнивались средняя абсолютная ошибка (MAE), отражающая среднее отклонение

прогнозируемых значений от фактических; средняя абсолютная процентная ошибка (MAPE); среднеквадратическая ошибка (RMSE), учитывающая квадрат отклонений; коэффициент детерминации (R^2) (таблица). В качестве эталонного метода выбрана модель ARIMA. Разработанная модель показала наиболее высокие результаты в сравнении с эталонной моделью ARIMA и нейросетевой реверсивной рекуррентной. Наиболее сильным конкурентом оказалась мультисверточная модель, которая дала осредненный показатель среднеквадратичной ошибки 0,0557, в то время как разработанная модель показала ошибку на 0,0248 ниже (т.е. среднее значение $RMSE = 0,0309$).

Для наглядности построен график сравнения результатов по концентрациям аммония Окского водозабора по классической модели ARIMA и разработанной нейросети (рис. 4).

В обучении на исходных данных за период 2018–2022 гг. разработанная нейросетевая модель занизила показатель на 2021 г. (соответствующий показатель на графике – 4-й год). В сравнении с эталонной моделью ARIMA нейросеть давала заниженные показатели концентраций. Однако точность нейросети выше, чем ARIMA ($RMSE_{ARIMA}$ больше $RMSE_{нейросеть}$ на 1,17).

Выводы. Результаты лабораторных исследований показали, что вода, взятая на Соколовском водозаборе, имела значительно более низкую среднюю многолетнюю концентрацию аммиака, равную 0,48 мг/л. Это значение оказалось существенно ниже, чем аналогичные показатели на Окском и Борковском водозаборах, которые были в 1,6 и 2,1 раза выше соответственно ($p < 0,05$). Не было обнаружено значительных различий в средних многолетних значениях химического потребления кислорода (ХПК) и биохимического потребления кислорода (БПК₅) в водах исследуемых водозаборов. При этом процент разовых проб, в которых указанные показатели не соответствовали установленным гигиеническим нормам, колебался в диапазоне от 22,7 до 32,5 % для ХПК и от 61,8 до 75,0 % для БПК₅.

Сравнение параметров коэффициента детерминации и среднеквадратичной ошибки разработанной нейросетевой модели с существующими для показателей Окского водозабора

| Показатель качества | Параметр модели статистической обработки | Авторегрессия – Интегрированное скользящее среднее (ARIMA) | Нейронная модель | | |
|--------------------------|--|--|---|------------------------|---|
| | | | Реверсивная рекуррентная нейронная сеть | Мульти-сверточная LSTM | Разработанная нейросеть (сверточная + двунаправленная рекуррентная) |
| ХПК, мгО ₂ /л | R^2 | 0,9408 | 0,9920 | 0,9996 | 0,9996 |
| | RMSE | 1,2030 | 0,5360 | 0,0566 | 0,0299 |
| Азот общий, мг/л | R^2 | 0,8760 | 0,9933 | 0,9999 | 0,9996 |
| | RMSE | 1,0000 | 0,5400 | 0,0542 | 0,0315 |
| Аммиак | R^2 | 0,9400 | 0,9945 | 0,9999 | 0,9999 |
| | RMSE | 0,9850 | 0,5466 | 0,0520 | 0,0310 |
| О ₂ , мг/л | R^2 | 0,9308 | 0,9900 | 0,9996 | 0,9999 |
| | RMSE | 1,0000 | 0,5280 | 0,0600 | 0,0312 |

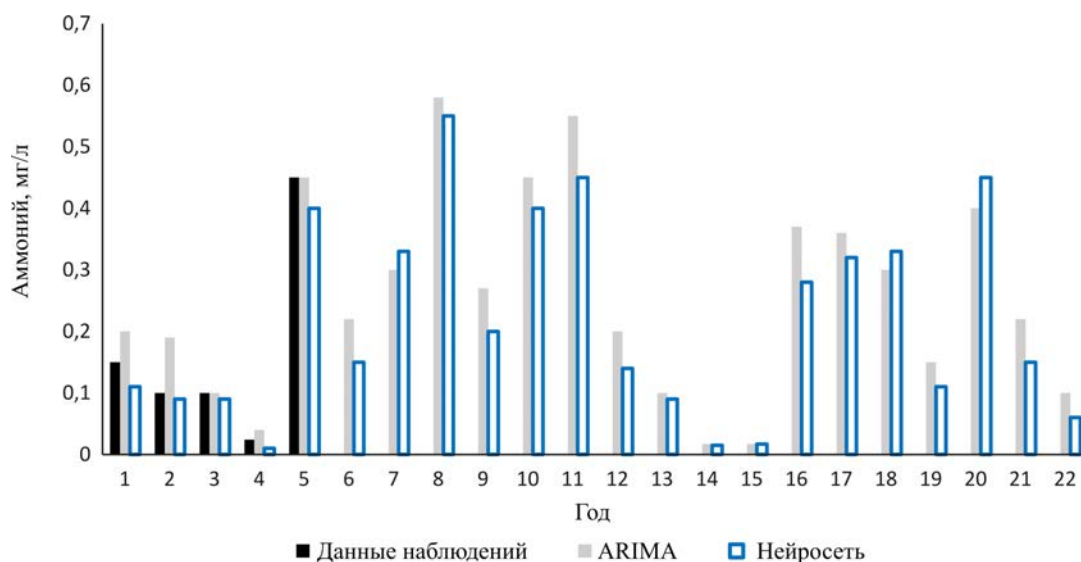


Рис. 4. Сравнение результатов средних концентраций аммония по данным наблюдений (период 1–5 лет, равный 2018–2022 гг.), классического прогноза ARIMA и разработанной нейронной сети для Окского водозабора

На предварительном этапе проведенного исследования с целью обеспечения высокой достоверности исходной информационной базы были использованы передовые подходы к обработке данных, включающие в себя алгоритм изолирующего леса и интерполяционные методы Лагранжа, что позволило не только эффективно повысить целостность информационного массива, но и минимизировать потенциальное влияние погрешностей и аномалий на последующий процесс моделирования. В дополнение к предварительной обработке исходных данных были использованы метод скользящего среднего и метод главных компонент анализа, обеспечившие оптимизацию параметров качества водных ресурсов и предотвращение явления переобученности модели, что является критически важным фактором для обеспечения высокой точности прогнозных расчетов в долгосрочной перспективе. Таким образом, разработана инновационная гибридная нейросетевая модель прогнозирования параметров качества воды, основанная на интеграции глубокой сверточной нейронной сети и двунаправленной (bidirectional) рекуррентной нейросети, которая состоит из трех функциональных частей.

Результаты экспериментальной апробации разработанной модели убедительно демонстрируют высокую устойчивость и обобщающую способность предложенного подхода, что проявляется в значительном снижении погрешности прогнозирования по сравнению с традиционными методиками и открывает новые перспективы для применения данной концепции в прогнозировании одномерных хронологических рядов различных объектов естественно-

научного и технического анализа в сравнении с такими моделями, как авторегрессия – интегрированная скользящее среднее и реверсивная рекуррентная нейронная сеть. Наиболее мощным конкурентом стала мультисверточная модель, которая продемонстрировала среднеквадратичную ошибку на уровне 0,0557. В то же время разработанная модель показала ошибку на 0,0248 меньше, что соответствует среднему значению $RMSE = 0,0309$.

Проведенные исследования в области применения рекуррентных нейросетевых моделей для прогнозирования загрязнения реки Оки, являющейся одним из ключевых водных объектов центральной России, позволяют сделать обоснованный вывод о возможности осуществления точного и своевременного предсказания изменений качества воды, что открывает новые возможности для реализации эффективных мер по охране природы и обеспечению устойчивого развития региона в долгосрочной перспективе. Использование разработанной модели для прогнозирования динамики загрязнения реки Оки на ближайшие два десятилетия предоставляет уникальную возможность для выявления потенциальных внешнесредовых угроз и принятия превентивных мер по их предотвращению, что является значимым шагом в направлении сохранения природных ресурсов, обеспечения санитарно-эпидемиологической безопасности и повышения качества жизни населения в регионе.

Финансирование. Исследование не имело спонсорской поддержки.

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Список литературы

1. An integrated simulation framework for NDVI pattern variations with dual society-nature drives: A case study in Baiyangdian Wetland, North China / Z. Liao, X. Wang, Y. Zhang, H. Qing, C. Li, Q. Liu, J. Cai, C. Wei // *Ecological Indicators*. – 2024. – Vol. 158. – P. 111584. DOI: 10.1016/j.ecolind.2024.111584
2. Карпенко Н.П., Глазунова И.В., Ширяева М.А. Анализ геоэкологических проблем и оценка обеспеченности питьевыми водами Клинского района Московской области // *Природообустройство*. – 2023. – № 5. – С. 88–94. DOI: 10.26897/1997-6011-2023-5-88-94
3. 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. – Vol. 13, № 7. – P. 1772. DOI: 10.3390/en13071772
4. 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 // *Engineering Applications of Artificial Intelligence*. – 2008. – Vol. 21, № 8. – P. 1189–1195. DOI: 10.1016/j.engappai.2008.03.015
5. Towards a time and cost effective approach to water quality index class prediction / J.Y. Ho, H.A. Afan, A.H. El-Shafie, S.B. Koting, N.S. Mohd, W.Z.B. Jaafar, L.S. Hin, M.A. Malek [et al.] // *Journal of Hydrology*. – 2019. – Vol. 575. – P. 148–165. DOI: 10.1016/j.jhydrol.2019.05.016
6. 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 // *Ecological indicators*. – 2016. – Vol. 61. – P. 170–178. DOI: 10.1016/j.ecolind.2015.08.034
7. Розенталь О.М., Федотов В.Х. Идентификация предприятий-загрязнителей воды на основе нейросетевого анализа // *Природообустройство*. – 2023. – № 1. – С. 62–68. DOI: 10.26897/1997-6011-2023-1-62-68
8. Шамсутдинова Т.М. Применение нейросетевого моделирования в задачах прогнозирования уровня паводка рек // *Вестник Новосибирского государственного университета. Серия: Информационные технологии*. – 2023. – Т. 21, № 2. – С. 39–50. DOI: 10.25205/1818-7900-2023-21-2-39-50
9. 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. – Т. 4, № 2. – С. 280–289.
10. Раткович Л.Д., Маркин В.Н., Глазунова И.В. Вопросы рационального использования водных ресурсов и проектного обоснования водохозяйственных систем: монография. – М.: Московский государственный университет природообустройства, 2013. – 256 с.
11. Карпенко Н.П., Ломакин И.М., Дроздов В.С. Вопросы управления геоэкологическими рисками при оценке качества подземных вод на урбанизированных территориях // *Природообустройство*. – 2019. – № 5. – С. 106–111. DOI: 10.34677/1997-6011/2019-5-106-111
12. Сравнительная характеристика показателей качества воды реки Оки в местах водозаборов хозяйственно-питьевой системы водоснабжения города Рязани / А.А. Литвинова, А.А. Дементьев, А.А. Ляпкало, Е.П. Коршунова // *Российский медико-биологический вестник имени академика И.П. Павлова*. – 2022. – Т. 30, № 4. – С. 481–488. DOI: 10.17816/PAVL0VJ89568
13. 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. – Т. 100, № 11. – С. 1192–1197. DOI: 10.47470/0016-9900-2021-100-11-1192-1197
14. Карпенко Н.П., Ширяева М.А. Трёхмерное моделирование как система отображения суммарного химического загрязнения почв // *Природообустройство*. – 2021. – № 1. – С. 6–14. DOI: 10.26897/1997-6011-2021-1-6-14
15. Оценка качества вод Рыбинского водохранилища вследствие снижения уровня вод / Н.В. Лагутина, А.В. Новиков, О.В. Сумарукова, Н.О. Науменко // *Природообустройство*. – 2019. – № 2. – С. 122–126. DOI: 10.34677/1997-6011/2019-2-122-126
16. Науменко Н.О. Введение рационального нормирования на объемы сбросов загрязняющих веществ в водные объекты с целью поддержания устойчивости экосистемы // *Современные проблемы и перспективы развития рыбохозяйственного комплекса: материалы VII Научно-практической конференции молодых ученых с международным участием*. – М.: Всероссийский научно-исследовательский институт рыбного хозяйства и океанографии, 2019. – С. 344–346.
17. Multi-scale quaternion CNN and BiGRU with cross self-attention feature fusion for fault diagnosis of bearing / H. Liu, F. Zhang, Y. Tan, L. Huang, Y. Li, G. Huang, S. Luo, A. Zeng // *Meas. Sci. Technol.* – 2024. – Vol. 35, № 8. – P. 086138. DOI: 10.1088/1361-6501/ad4c8e
18. A deep learning algorithm for multi-source data fusion to predict water quality of urban sewer networks / Y. Jiang, C. Li, L. Sun, D. Guo, Y. Zhang, W. Wang // *Journal of Cleaner Production*. – 2021. – Vol. 318. – P. 128533. DOI: 10.1016/j.jclepro.2021.128533
19. Forecasting models for surface water quality using predictive analytics / G.T.N. Veerendra, B. Kumaravel, P. Kodanda Rama Rao, S. Dey, A.V. Phani Manoj // *Environment, Development and Sustainability*. – 2024. – Vol. 26, № 6. – P. 15931–15951. DOI: 10.1007/s10668-023-03280-3
20. A novel global average temperature prediction model – based on GM-ARIMA combination model / X. Chen, Z. Jiang, H. Cheng, H. Zheng, D. Cai, Y. Feng // *Earth Science Informatics*. – 2023. – Vol. 17, № 1. – P. 853–866. DOI: 10.1007/s12145-023-01179-1
21. Water quality evaluation and prediction based on a combined model / G. Jiao, S. Chen, F. Wang, Z. Wang, F. Wang, H. Li, F. Zhang, J. Cai, J. Jin // *Appl. Sci.* – 2023. – Vol. 13, № 3. – P. 1286. DOI: 10.3390/app13031286
22. Application of data prediction models in a real water supply network: comparison between arima and artificial neural networks / A.C. da Silva, F. das Graças Braga da Silva, V.E. de Mello Valério, A.T.Y. Lima Silva, S.M. Marques, J.A. Tosta dos Reis // *Revista Brasileira de Recursos Hídricos*. – 2024. – Vol. 29. – P. e12. DOI: 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. – Vol. 284. – P. 112051. DOI: 10.1016/j.jenvman.2021.112051

24. Simulation on TN and TP distribution of sediment in Liaohe estuary national wetland park using mike21-coupling model / X. Lu, Y. Dong, Q. Liu, H. Zhu, X. Xu, J. Liu, Y. Wang // *Water*. – 2023. – Vol. 15, № 15. – P. 2727. DOI: 10.3390/w15152727

25. Augmentation of limited input data using an artificial neural network method to improve the accuracy of water quality modeling in a large lake / J. Kim, D. Seo, M. Jang, J. Kim // *Journal of Hydrology*. – 2021. – Vol. 602, № 4. – P. 126817. DOI: 10.1016/j.jhydrol.2021.126817

26. Wongburi P., Park J.K. Prediction of Wastewater Treatment Plant Effluent Water Quality Using Recurrent Neural Network (RNN) Models // *Water*. – 2023. – Vol. 15, № 19. – P. 3325. DOI: 10.3390/w15193325

27. Jaya N.A., Arsyad M., Palloan P. Estimation of Groundwater River Availability in Leang Lonrong Cave Using ARIMA Model and Econophysics Valuation Approach // *Advances in Social Humanities Research*. – 2024. – Vol. 2, № 5. – P. 737–754. DOI: 10.46799/adv.v2i5.240

28. Tiyasha, Tung T.M., Yaseen Z.M. Deep learning for prediction of water quality index classification: tropical catchment environmental assessment // *Natural Resources Research*. – 2021. – Vol. 30, № 6. – P. 4235–4254. DOI: 10.1007/s11053-021-09922-5

Алгоритм прогнозирования параметров качества водных объектов с использованием нейронной сети / М.А. Ширяева, О.О. Синицына, М.В. Пушкарева, В.В. Турбинский // Анализ риска здоровью. – 2024. – № 4. – С. 50–62. DOI: 10.21668/health.risk/2024.4.05

UDC 614.878.086

DOI: 10.21668/health.risk/2024.4.05.eng



Research article

ALGORITHM FOR PREDICTING WATER QUALITY INDICATORS IN WATER BODIES USING A NEURAL NETWORK

M.A. Shiryayeva^{1,2}, O.O. Sinitsyna¹, M.V. Pushkareva¹, V.V. Turbinsky¹

¹F.F. Erisman Federal Scientific Center of Hygiene, 2 Semashko St., Mytishchi, Moscow region, 141014, Russian Federation

²Russian State Agrarian University – Moscow Timiryazev Agricultural Academy, 49 Timiryazeva St., Moscow, 127550, Russian Federation

Clean and safe drinking water is a fundamental necessity for human health and well-being and a critical component in sustainable ecosystem development. In recent decades, water quality issues have become even more urgent due to population growth, industrial expansion and climate change.

A series of works by foreign researchers report results obtained by applying neural networks. There are studies confirming results of water quality prediction generated by neural networks to be quite valid.

In this research, we used Google Earth Pro, Microsoft Excel, water flow sensor based on Arduino UNO board with author's modification (tail feathering and built-in plugin for calculation of flow velocity), Python, Tensorflows keras2.2.0, Scikit-learn, Pandas libraries for machine learning and development of neural network architecture. In this study, two ANNs were combined to build a hybrid neural network model for predicting water quality indicators.

© Shiryayeva M.A., Sinitsyna O.O., Pushkareva M.V., Turbinsky V.V., 2024

Margarita A. Shiryayeva – Junior Researcher of Water Hygiene Department (e-mail: Shiryayeva.MA@fncg.ru; tel.: +7 (903) 161-14-04; ORCID: <https://orcid.org/0000-0001-8019-1203>).

Oxana O. Sinitsyna – Corresponding Member of the Russian Academy of Sciences, Doctor of Medical Sciences, Professor, Deputy Director for Research, Director of the Institute of Complex Hygiene Issues (e-mail: sinitsyna.oo@fncg.ru; tel.: +7 (926) 447-08-74; ORCID: <http://orcid.org/0000-0002-0241-0690>).

Maria V. Pushkareva – Doctor of Medical Sciences, Professor, Chief Researcher of Water Hygiene Department (e-mail: pushkareva.mv@fncg.ru; tel.: +7 (912) 980-92-74; ORCID: <https://orcid.org/0000-0002-5932-6350>).

Viktor V. Turbinsky – Doctor of Medical Sciences, Head of Water Hygiene Department (e-mail: turbinskii.vv@fncg.ru; tel.: +7 (920) 666-72-73; ORCID: <https://orcid.org/0000-0001-7668-9324>).

Neural network models offer unique opportunities to improve water resources management at various levels, ranging from local to global one. A key advantage of such models is a possibility to adapt them to specific conditions and requirements, which provides more accurate prediction and timely decision making under uncertainty. The relevance of the work is determined by application of neural networks for water quality prediction. This can improve systems for early warning about pollution, help optimize operational processes at water treatment plants and develop effective water management strategies.

In this research, an innovative hybrid neural network model has been developed for predicting water quality indicators. It is based on integrating deep convolutional neural network and bidirectional recurrent neural network, which consists of three functional parts.

Keywords: neural network, Tensorflows keras2.2.0, water bodies, drinking water, risk factor, negative impact, water pollution, determination coefficient, optimization algorithm.

References

1. Liao Z., Wang X., Zhang Y., Qing H., Li C., Liu Q., Cai J., Wei C. An integrated simulation framework for NDVI pattern variations with dual society-nature drives: A case study in Baiyangdian Wetland, North China. *Ecological Indicators*, 2024, vol. 158, pp. 111584. DOI: 10.1016/j.ecolind.2024.111584
2. Karpenko N.P., Glazunova I.V., Shiryayeva M.A. Analysis of geo ecological problems and assessment of the availability of drinking water in the Klinsky district of the Moscow region. *Prirodoobustroistvo*, 2023, no. 5, pp. 88–94. DOI: 10.26897/1997-6011-2023-5-88-94 (in Russian).
3. 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, vol. 13, no. 7, pp. 1772. DOI: 10.3390/en13071772
4. 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. *Engineering Applications of Artificial Intelligence*, 2008, vol. 21, no. 8, pp. 1189–1195. DOI: 10.1016/j.engappai.2008.03.015
5. Ho J.Y., Afan H.A., El-Shafie A.H., Koting S.B., Mohd N.S., Jaafar W.Z.B., Hin L.S., Malek M.A. [et al.]. Towards a time and cost effective approach to water quality index class prediction. *Journal of Hydrology*, 2019, vol. 575, pp. 148–165. DOI: 10.1016/j.jhydrol.2019.05.016
6. 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. *Ecological indicators*, 2016, vol. 61, pp. 170–178. DOI: 10.1016/j.ecolind.2015.08.034
7. Rosenthal O.M., Fedotov V.Kh. Identification of water polluting enterprises based on neural network analysis. *Prirodoobustroistvo*, 2023, no. 1, pp. 62–68. DOI: 10.26897/1997-6011-2023-1-62-68 (in Russian).
8. Shamsutdinova T.M. Application of Neural Network Modeling in Problems of Predicting the Level of River Floods. *Vestnik Novosibirskogo gosudarstvennogo universiteta. Seriya: Informatsionnye tekhnologii*, 2023, vol. 21, no. 2, pp. 39–50. DOI: 10.25205/1818-7900-2023-21-239-50 (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, vol. 4, no. 2, pp. 280–289 (in Russian).
10. Ratkovich L.D., Markin V.N., Glazunova I.V. Voprosy ratsional'nogo ispol'zovaniya vodnykh resursov i proektnogo obosnovaniya vodokhozyaistvennykh sistem: monografiya [Issues of rational use of water resources and design substantiation of water management systems: a monograph]. Moscow, K.A. Timiryazev Russian State Agrarian University – K.A. Timiryazev MSHA, 2013, 256 p. (in Russian).
11. Karpenko N.P., Lomakin I.M., Drozdov V.S. Management issues of geoenvironmental risks in the assessment of groundwater quality in urban areas. *Prirodoobustroistvo*, 2019, no. 5, pp. 106–111. DOI: 10.34677/1997-6011/2019-5-106-111 (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 medikobioogicheskii vestnik imeni akademika I.P. Pavlova*, 2022, vol. 30, no. 4, pp. 481–488. DOI: 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*, 2021, vol. 100, no. 11, pp. 1192–1197. DOI: 10.47470/0016-9900-2021-100-11-1192-1197 (in Russian).
14. Karpenko N.P., Shiryayeva M.A. Three-dimensional modeling as a system for displaying total chemical soil pollution. *Prirodoobustroistvo*, 2021, no. 1, pp. 6–14. DOI: 10.26897/1997-6011-2021-1-6-14 (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. *Prirodoobustroistvo*, 2019, no. 2, pp. 122–125. DOI: 10.34677/1997-6011/2019-2-122-126 (in Russian).
16. Naumenko N.O. Vvedenie ratsional'nogo normirovaniya na ob"emy sbrosov zagryaznyayushchikh veshchestv v vodnye ob"ekty s tsel'yu podderzhaniya ustoichivosti ekosistemy [Introduction of rational rationing of the volume of pollutant discharges into water bodies in order to maintain sustainability of an ecosystem]. *Sovremennye problemy i perspektivy razvitiya rybokhozyaistvennogo kompleksa: materialy VII nauchno-prakticheskoi konferentsii molodykh uchenykh s mezhdunarodnym uchastiem*. Moscow, Russian Federal Research Institute of Fisheries and Oceanography Publ., 2019, pp. 344–346 (in Russian).

17. Liu H., Zhang F., Tan Y., Huang L., Li Y., Huang G., Luo S., Zeng A. Multi-scale quaternion CNN and BiGRU with cross self-attention feature fusion for fault diagnosis of bearing. *Meas. Sci. Technol.*, 2024, vol. 35, no. 8, pp. 086138. DOI: 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. *Journal of Cleaner Production*, 2021, vol. 318, pp. 128533. DOI: 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. *Environment, Development and Sustainability*, 2024, vol. 26, no. 6, pp. 15931–15951. DOI: 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 Science Informatics*, 2023, vol. 17, no. 1, pp. 853–866. DOI: 10.1007/s12145-023-01179-1
21. Jiao G., Chen S., Wang F., Wang Z., Wang F., Li H., Zhang F., Cai J., Jin J. Water quality evaluation and prediction based on a combined model. *Appl. Sci.*, 2023, vol. 13, no. 3, pp. 1286. DOI: 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. *Revista Brasileira de Recursos Hídricos*, 2024, vol. 29, pp. e12. DOI: 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, vol. 284, pp. 112051. DOI: 10.1016/j.jenvman.2021.112051
24. Lu X., Dong Y., Liu Q., Zhu H., Xu X., Liu J., Wang Y. Simulation on TN and TP distribution of sediment in Liaohe estuary national wetland park using mike21-coupling model. *Water*, 2023, vol. 15, no. 15, pp. 2727. DOI: 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. *Journal of Hydrology*, 2021, vol. 602, no. 4, pp. 126817. DOI: 10.1016/j.jhydrol.2021.126817
26. Wongburi P., Park J.K. Prediction of Wastewater Treatment Plant Effluent Water Quality Using Recurrent Neural Network (RNN) Models. *Water*, 2023, vol. 15, no. 19, pp. 3325. DOI: 10.3390/w15193325
27. Jaya N.A., Arsyad M., Palloan P. Estimation of Groundwater River Availability in Leang Lonrong Cave Using ARIMA Model and Econophysics Valuation Approach. *Advances in Social Humanities Research*, 2024, vol. 2, no. 5, pp. 737–754. DOI: 10.46799/adv.v2i5.240
28. Tiyasha, Tung T.M., Yaseen Z.M. Deep learning for prediction of water quality index classification: tropical catchment environmental assessment. *Natural Resources Research*, 2021, vol. 30, no. 6, pp. 4235–4254. DOI: 10.1007/s11053-021-09922-5

Shiryayeva M.A., Simitsyna 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, no. 4, pp. 50–62. DOI: 10.21668/health.risk/2024.4.05.eng

Получена: 22.07.2024

Одобрена: 29.11.2024

Принята к публикации: 19.12.2024