Magazine of Civil Engineering. 2025. 18(6). Article No. 13806

Magazine of Civil Engineering

ISSN 2712-8172

journal homepage: http://engstroy.spbstu.ru/

Research article

UDC 69

DOI: 10.34910/MCE.138.6

# Neural network modeling for real-time water quality assessment

***A.A. Naumova***[1] ✉, ***V.V. Ilinich***[2], ***M.A. Shiryaeva***[3]

[1] *Russian State Agrarian University – Moskow Timiryazev Agricultural Academy, Moscow, Russian Federation*

[2] *A.N. Kostyakov Federal Scientific Center for Hydraulic Engineering and Land Reclamation, Moscow, Russian Federation*

[3] *Federal Budgetary Institution of Science Federal Scientific Center of Hygiene named after F.F. Erisman of the Federal Service for Supervision of Consumer Rights Protection and Human Well-Being, Mytishchi, Moscow Region, Russian Federation*

✉ *koshevaya81@mail.ru*

**Keywords:** neural network, Tensorflows Keras 2.2.0, watershed, water, water flow, negative impact, water quality, water resources, optimization's algorithm.

**Abstract.** In recent decades, water quality problems have become even more pressing due to population growth, industrial expansion, and climate change. A number of studies by foreign researchers have shown the results of applying neural networks. There are studies confirming the reliability of water quality prediction results generated by neural networks. During the work, OpenAI Earth Pro, Microsoft Excel, a water flow sensor based on the Arduino UNO board with author's modifications (tail feathers and a built-in plugin for calculating flow velocity), Python, Tensorflows Keras 2.2.0, Scikit-learn, Pandas libraries for machine learning and developing the neural network architecture were used. Two neural network models were combined to build a hybrid neural network model for predicting water quality parameters in the research. Neural network models provide unique opportunities to improve water resource management at various levels, from local to global. One of the key advantages of such models is the ability to adapt to specific conditions and requirements, providing more accurate predictions and timely decision-making in the face of uncertainty. The relevance of the work is due to the application of neural networks for predicting water quality can contribute to improving the early warning system for pollution, optimizing operational processes at water treatment plants, and developing effective strategies for water resource management. During the research, an innovative hybrid neural network model for predicting water quality parameters was developed, based on the integration of a deep convolutional neural network and a bidirectional recurrent neural network, which consists of three functional parts.

**Citation:** Naumova, A.A., Ilinich, V.V., Shiryaeva, M.A. Neural network modeling for real-time water quality assessment. Magazine of Civil Engineering. 2025. 18(6). Article no. 13806. DOI: 10.34910/MCE.138.6

## *1. Introduction*

Water resources assessment is critical in contemporary society, especially against progressive anthropogenic pressure on aquatic ecosystems and climate change manifestations. Rivers, lakes, and reservoirs are the main centralized sources of drinking water supply to the population, the key irrigation object for agriculture, the main provider of water resources for the industry, as well as an essential part of the recreation and leisure infrastructure for the population [1].

Due to the increasing human impact and pollution, there has been a significant degradation in the global watersheds' water quality in recent decades. This circumstance necessitates the development and implementation of innovative approaches to monitoring and forecasting the state of the hydrological environment that exceed conventional methods in terms of accuracy, reliability and efficiency in obtaining results [2, 3].

One of the most prominent trends in this regard is the use of machine learning and neural networks for modelling and predicting factors determining the water quality dynamics [4–6]. Similar models are able to take into account the complex non-linear interrelationships of numerous factors and are self-learning, making them a highly effective tool for solving the assigned tasks.

Surface water quality represents a fundamental determinant of aquatic ecosystem health and the sustainability of water supply systems. Contamination by nutrients (e.g., ammonium, nitrates, phosphorus), organic pollutants, and pathogens can lead to eutrophication, hypoxia, and significant disruption of biogeochemical cycles, adversely affecting biodiversity and ecosystem services [7].

Recent assessments indicate a progressive decline in the chemical, microbiological, and physicochemical status of surface waters in regions with high anthropogenic pressure, including industrial, agricultural, and urban activities [8]. Agricultural runoff, in particular, contributes to elevated concentrations of nitrogenous compounds and pesticides, which exacerbate nutrient loading and pose challenges to conventional water treatment processes [9].

Effective management of water resources under changing climatic conditions requires accurate, timely, and predictive assessments of water quality. Traditional mechanistic models (e.g., QUAL2K, SWAT) often demand extensive input data and computational resources, limiting their adaptability to real-time monitoring and forecasting [10–13]. Machine learning approaches, particularly hybrid neural network models, offer an alternative framework capable of capturing nonlinear interactions among hydrological, chemical, and environmental variables, thereby enhancing predictive accuracy [3–5, 14].

Consequently, there is a demonstrable need for research integrating long-term monitoring data with advanced analytical methods to develop robust predictive models for surface water quality. Such studies are critical for informed decision-making in water resource management, pollution control, and ecological conservation.

The objective of the research was to develop and apply an innovative neural network-based algorithm for predicting key water quality parameters. To evaluate the effectiveness of the proposed approach, its performance was compared with existing statistical and neural network models: autoregressive integrated moving average (ARIMA), recurrent neural, long short-term memory (LSTM) networks.

**Object of the study:** A section of the Oka River channel within the Ryazan urban agglomeration. This section was selected as representative of the interactions between a major watercourse and surrounding anthropogenic influences.

**Subject of the study:** Temporal and spatial variations of key water quality parameters (ammonium, total nitrogen, chemical oxygen demand (COD), biochemical oxygen demand ($BOD_5$), total coliform bacteria (TBA)) and their prediction using machine learning models.

A review of existing approaches to water quality prediction was conducted to identify their methodological limitations and potential for improvement. On this basis, a neural network model was developed. The model structure, input parameters, and training dataset were defined using historical water chemistry data from the Oka River. Calibration of the algorithm was performed to optimize prediction accuracy.

The performance of the developed neural model was evaluated through comparative tests. Predicted values were compared with observed measurements as well as with the results of reference models (ARIMA, recurrent, and LSTM networks). This procedure allowed assessment of the relative advantages of the proposed approach and identification of conditions, under which its application is most effective.

The final stage was the comparison between the new approach and classical methods, through which the advantages and disadvantages of the proposed model were revealed. Finally, recommendations on the algorithm's practical application were formulated, and the prospects for subsequent improvement of the development were identified. The study is a significant contribution to the solution of the strategic problem of water resources quality forecasting.

## 2. Methods

The study employed a combination of observational, statistical, and computational methods:

1. **Data Collection:** water samples were collected at three intake stations (Sokolovskiy, Okskiy, Borkovskiy) of the Oka River from 2014 to 2022. Measured parameters included ammonium, total nitrogen, COD, $BOD_5$, and TBA, along with organoleptic and microbiological indicators.

2. **Statistical Analysis:** long-term averages, standard deviations, and exceedances relative to regulatory thresholds were calculated. Differences between intake stations were evaluated using appropriate statistical tests (e.g., t-test, ANOVA, $p < 0.05$).
3. **Neural Network Modeling:** a hybrid model combining convolutional neural networks (CNNs) and bidirectional recurrent neural networks (Bi-RNNs) was developed. Input data included historical water quality parameters and temporal hydrological features. The model was trained and validated using TensorFlow Keras, with hyperparameters optimized to minimize prediction error: root mean square error (RMSE), mean absolute error (MAE), the coefficient of determination (R-squared). Predictions were compared with benchmark models (ARIMA, LSTM, reverse recurrent networks) to assess relative performance.
4. **Data Visualization:** Spatial analysis and visualization were performed using Google Earth Pro. Temporal trends and model predictions were visualized using spreadsheet programs and standard plotting software.

## *2.1. The Research Object and Hydrological Monitoring*

Research into the ecological status and dynamics of water quality parameters was carried out on a section of the Oka River channel located within the Ryazan urban agglomeration [6, 16]. This section serves as a representative model for the interaction of a major watercourse with an urbanized environment.

Effective prediction of the Oka River water quality, enabling timely detection of negative trends and the implementation of measures to prevent sanitary-epidemiological and environmental crises, requires comprehensive and systematic monitoring [17, 18].

A key factor in enhancing the efficiency and representativeness of the monitoring system is the optimal placement of data collection control points. This requires consideration of both the characteristics of the river's hydrological regime and the spatial distribution of potential pollution sources [19].

The presented figure (Fig. 1) proposes an approximate layout for modular weather stations integrated with automatic water sampling systems. Red circles indicate the measuring instruments, the green sector represents the observation area (400 km$^2$), and the blue sector indicates the zone where observations from two adjacent stations overlap. This overlap ensures the necessary data redundancy, which increases the reliability of the results.

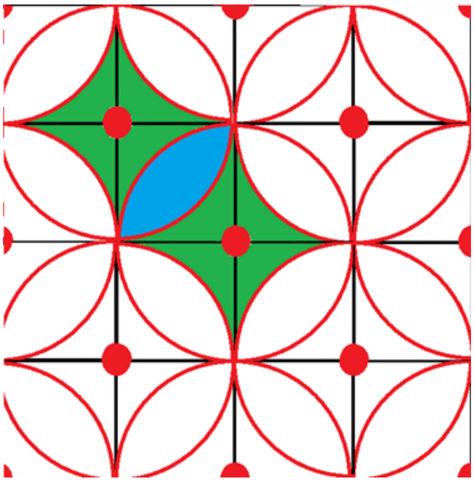

**Figure 1. Location scheme of module meteorological stations (created by authors).**

The proposed layout for the monitoring stations was designed to optimize economic considerations while maintaining the highest possible level of efficiency for the systematic monitoring system. This is particularly crucial given the stringent budgetary constraints often imposed on environmental protection programs [4, 20–22]. Employing mathematical modeling and optimization methods, considering the hydrological specifics of the water body, characteristics of anthropogenic impact distribution, and economic limitations, allows for a highly accurate determination of the necessary and sufficient number of measuring installations to effectively cover the entire watershed area of the investigated section of the Oka River.

This ensures systematic monitoring and prediction of its ecological status, taking into account the multi-factor dynamics of external influences.

The distances between the modular stations will be:

$$l_1 = 21.6\sqrt{2} = 30.5\ \text{km}; \quad l_2 = 21.6 \cdot 2 = 43.2\ \text{km}.$$

The overlapping area of the observations of the two stations can be determined:

$$F = 2 \cdot \left( \frac{21.6^2 \cdot P}{4} - \frac{21.6^2}{2} \right) = \frac{21.6^2 \cdot P}{2} - \frac{2 \cdot 21.6^2}{2} = \frac{21.6^2 (P-2)}{2} \approx 266 \text{ km}^2.$$

Every two neighbouring stations have one point of overlap of their measuring zones, three stations have two points of overlap and so on in accordance with the established pattern.

Using this trend, it is possible to predict the required number of observation stations for complete coverage of control measurements over the whole area of the considered catchment. To quantify the required value, we propose to build a mathematical model by setting the equation, in which the initial parameter $(n)$ will be the number of stations providing coverage of the territory of 400 km$^2$.

Then, the number of intersection points between stations in a 266 km$^2$ area can be written as $(n-1)$. Having solved this equation, it is possible to determine the minimum necessary number of hydrometric stations to form a regular observation grid and ensure quality hydrometeorological monitoring over the entire catchment area, as well as to design the optimal configuration of the observation network.

Therefore, the following equation is obtained:

$$400n + 266(n-1) = 245000 \text{ km}^2, \quad n \approx 368 \text{ units.}$$

## *2.2. Water Sampling*

During the research and primary data collection, the following software and tools were used: Google Earth Pro for spatial analysis and data visualization, a spreadsheet program for statistical processing and preliminary analysis of the obtained results, and other standard software for data management and visualization.

To improve the measurement accuracy and expand the functional capabilities of the water flow sensor developed on the basis of the Arduino UNO board, a specialized software plug-in was created, which was based on a mathematically derived formula for converting data on water flow rate into flow velocity indicators, taking into account the geometrical parameters of the sensor, in particular, the diameter of the inlet and outlet holes, which was 11.9 mm, which provided an optimal ratio between the sensitivity of the device and its resistance to clogging by suspended particles.

Consequently, to determine the flow velocity (m/s) from the water flow rate (l/s), the following formulas were written into the data conversion plugin:

$$V = \frac{4W}{\pi \cdot D^2 \cdot 1000}, \quad V = \frac{4W}{\pi \cdot 0.0119^2 \cdot 1000}. \tag{1}$$

where $\pi$ = 3.14, $W$ – initial data of the water flow sensor in (l/s), $D$ – cross-sectional diameters of sensor input and output port.

Fig. 2 shows the design of a high-precision water flow measurement sensor designed to provide continuous remote monitoring of the dynamics of the velocity characteristics of the flow at different depths of the watercourse by cyclic longitudinal scanning of its channel part. The sensor is equipped with a stability system in the guided steering wheel, allowing to automatically correct its attitude depending on the current vector parameters of water mass transport both in the stream thickness and directly on its surface.

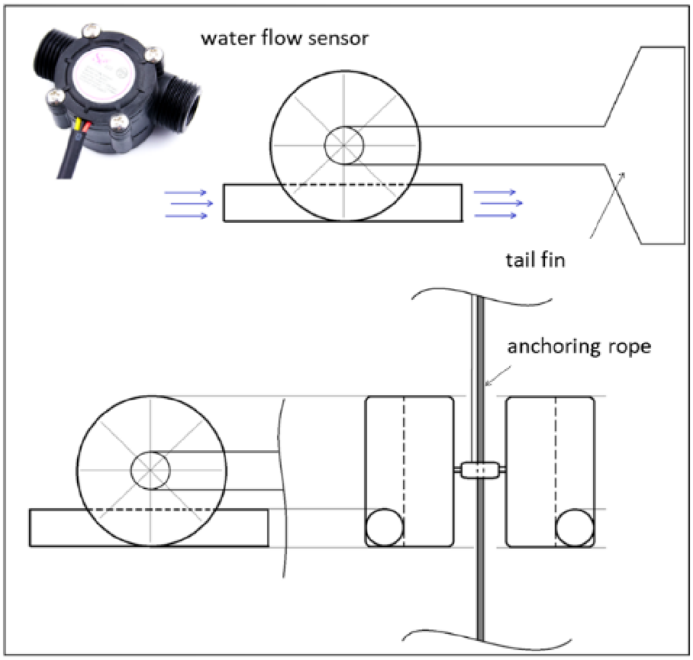


**Figure 2. Device of the water flow sensor.**

It is assumed that to ensure a stable drift position of the sensor at a given depth and within the required sector of the channel, it will be attached directly to the anchor cable of the unmanned hydrological probe using a flexible ring or hoop mount that provides compensation for its movements. This assembly scheme will make it possible to arrange permanent cyclic real-time hydraulic parameter tracking with high spatial and temporal resolution (Fig. 2).

This sensor sample had a tail fin made of polyvinyl chloride, an environmentally safe and durable thermoplastic material. The material ensured the reliability of the structure in watercourse conditions and minimized the impact on the hydro-ecosystem.

During laboratory approval of the device, it was revealed that at flow velocities over 5.5 m/s, there are limitations in qualitative transmission of measured parameters due to increased impact on the rotating elements of the sensor. Thus, at the rotational speeds of the blades, which ensure the collection of data on velocity, the microprocessor system of the measuring unit becomes overloaded, which complicates the processing of information.

Taking into account this factor, the limit of applicability of the tested model was determined at the level of 5.5 m/s. Further improvement of accuracy and operating range is possible by installing more advanced and expensive components that meet the stringent requirements of high-speed hydrometrics. However, implementation of such measures will require additional financial expenditures for design optimization.

### *2.3. Machine Learning and Neural Network*

The developed machine learning model demonstrated high accuracy in forecasting key water quality parameters of surface water, which is a critical source for water supply. Comparative evaluation against benchmark models (ARIMA, recurrent, and LSTM networks) confirmed its improved predictive performance, indicating that the model can be effectively used for operational monitoring and management of water resources.

Data on the chemical composition of water were obtained in the course of laboratory studies.

The algorithm for integrated water quality prediction proposed in this study includes the following sequential steps:

Step 1: Data cleaning. Before direct water quality prediction, the isolation forest (iForest) method is applied to identify anomalous values in the water quality data set $X_{n\times m}$ (where $n$ denotes the number of water quality parameters and $m$ denotes the number of data groups; in the context of this paper, $n$ and $m$ are constant values: $n$ = 9, $m$ = 1360), and the identified anomalous values are replaced with empty values. Subsequently, the Lagrangian interpolation method is used to fill in the empty values to ensure data integrity and continuity.

Step 2: Data expansion. In the first step, the predicted target is removed from the $X_{n\times m}$ array, resulting in a new array $X_{n\times(m-1)}$. Considering that the water quality data are collected at 4-hour intervals, a sliding window averaging technique with a window size of 6 is applied to form a set of moving averages $Z_{n\times(m-1)}$, which minimises the influence of random factors of variation in water quality data and traces the trend of daily variation of water quality parameters more accurately. In the second step, principal component analysis (PCA) technique is used to reduce the dimensionality of $X_n\times_{(m-1)}$ and retain two principal components $P_{2\times m}$. In order to prevent model overtraining, $Z_{n\times(m-1)}$, $P_{2\times m}$, and $X_{n\times(m-1)}$ water quality data without target parameters are simultaneously fed to the model input, while the target prediction is generated at the model output.

Step 3: Training the model. The available water quality data set is divided into training and test sets in the ratio of 8:2. For this study, the training set included 1100 datasets covering the period from 25 June 2021 to 16 February 2022, while the test set contained 272 datasets collected between 17 February 2022 and 1 April 2022. Considering the long-term dependence of water quality data on temporal factors, a sliding window method [19, 20] is applied to divide the training set into fixed training windows with step length *i* in the time sequence, after which the data from the first *j* training windows are used to predict the (*j*+1)-th training window. At each new training cycle, the oldest window is discarded and the next new window is included in the analysis, and this process continues until the last training window is reached. This approach of discarding outdated data favors training the model with future trends. In the final step, according to each station's test set, the trained model is applied to predict key water quality parameters including total nitrogen, total phosphorus and permanganate oxidizability.

As part of the comprehensive study, a detailed evaluation of the effectiveness of the proposed hybrid neural network model for predicting water quality parameters was carried out, including a benchmarking analysis with reference methods used in the field. In order to obtain a quantitative characterization of forecasting accuracy, the researchers used a number of metrics generally accepted in the scientific community, including: MAE, reflecting the average deviation of predicted values from actual values; MAPE, which allows estimating the relative magnitude of the forecast error; RMSE, which takes into account the square of deviations and gives greater weight to large errors; and R-squared, characterizing the proportion of variance over the forecasted values.

In the initial phase of the study, outliers in the raw water quality data at the study stations were identified and quantified at approximately 1.1, 1.7, and 3.2 % of the total data, respectively, using the iForest method, which is an efficient algorithm for detecting anomalies in multivariate data. After careful removal of identified outliers that could significantly affect the accuracy of the model, the remaining missing values for stations 1–3 were approximately 3.9, 4.5, and 5 %, respectively, which necessitated the application of data reconstruction techniques, in particular, Lagrange interpolation was used to recover a continuous function from a discrete set of points.

# *3. Results and Discussion*

## *3.1. Chemical Analysis*

Long-term monitoring data (2014–2022) were analyzed for three intake stations of the Oka River: Sokolovskiy, Okskiy, and Borkovskiy. The dataset included 52 parameters, among them organoleptic, microbiological, and chemical indicators.

The analysis showed that average ammonium concentrations at the Sokolovskiy intake were 0.48 mg/L, significantly lower than those recorded at the Okskiy and Borkovskiy intakes by 1.6 and 2.1 times, respectively ($p < 0.05$). Exceedances of the maximum permissible concentration (MPC) for ammonium were detected in nearly 20 % of samples at Borkovskiy, compared with 7.5 % at the Oka intake, while no exceedances were found at Sokolovskiy.

For COD and $BOD_5$, long-term averages did not differ significantly among the three sites. However, 22.7–32.5 % of samples exceeded regulatory limits for COD and 61.8–75.0 % for $BOD_5$, reflecting persistent oxygen regime disturbances. The mean content of TBA at the Okskiy and Borkovskiy intakes (813.3 and 818.9 CFU/100 ml, respectively) exceeded the values at Sokolovskiy by 1.5 times ($p < 0.05$).

Some results are summarized in Fig. 3, which presents mean, maximum, and minimum values for selected indicators across the three intake points.

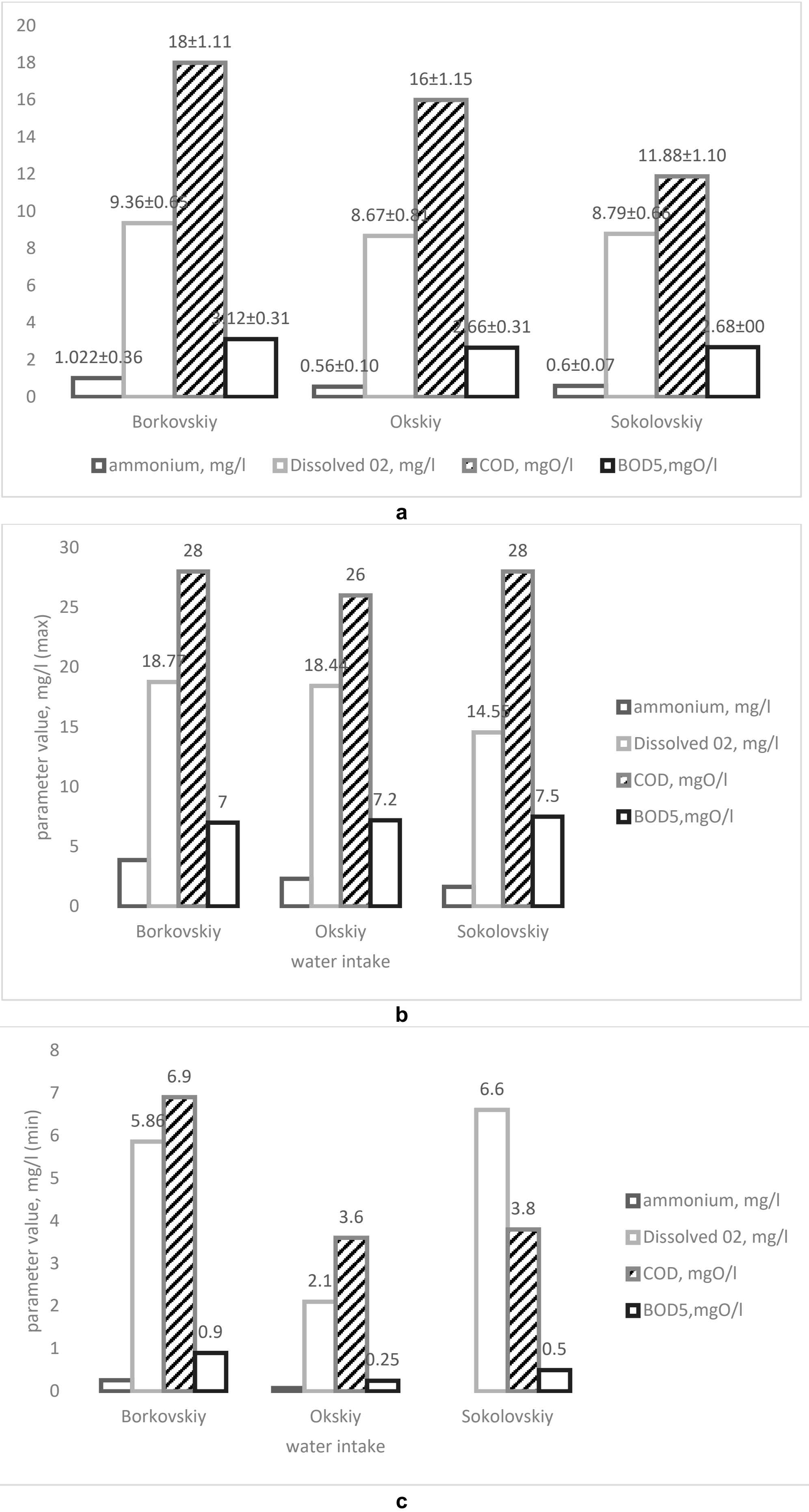


**Figure 3. Some water quality indicators of the Oka River of three researched water intakes for the period 2014–2022, where: a – the mean values, b – the maximum values, c – the minimum values.**

Overall, the data demonstrate spatial heterogeneity in water quality across intake sites and frequent exceedances of regulatory standards, particularly for oxygen-demanding substances and microbial contamination. These findings highlight the need for continuous monitoring and reliable prediction tools to support timely water quality management.

### *3.2. The Neural Network Development*

Within the current research work, an innovative hybrid neural network model for predicting water quality parameters based on the deep CNN integration and Bi-RNN has been developed, which consists of three functional parts. At the initial stage, the model is applied to identify and extract potential non-linear relationships between the Oka River water quality time series data in order to generate effective low-dimensional features. Then, based on the extracted features, a vector of water quality features is constructed and used as input to a deep CNN. During the training process, the network continuously adjusts the weights and biases by considering the dependency of short-term, long-term, and contextual attributes of the time series data to further optimize the water quality information for more accurate feature expression. In the final step, a full connection layer is connected at the top of the model to act as an output layer for generating predicted values of water quality parameters.

The software implementation of the developed hybrid neural network prediction model was carried out using the high-performance deep learning library Tensorflows Keras version 2.2.0, which provides a wide range of tools for building and training neural networks. The model training process was carried out over 50 epochs using 120-time intervals to achieve an optimal balance between prediction accuracy and computational cost. The Adam method, which combines the advantages of adaptive gradient descent and method of moments methods, was applied as an optimization algorithm to adjust the weights and biases of the model. Once the convergence of the model was achieved, which indicated that the loss function was minimized, the final weights were obtained and subsequently used to predict water quality at the studied intake stations (Sokolovskiy, Okskiy, and Borkovskiy). The model architecture and parameters were carefully selected and set as follows: the number of hidden layers was two, which allowed the model to effectively capture complex non-linear dependencies in the data; the conjugate gradient method, known for its ability to quickly converge to an optimal solution, was chosen as the optimization algorithm; the minimum relative change in the learning error rate was set at 0.001, which provided a balance between model accuracy and the prevention of overtraining.

This makes it possible to use the presented neural network model to fill data gaps by calculating missing concentration values of certain compounds. Some results of the neural network modelling covering a 25-year observation interval ($n$ = 25) are demonstrated as an example in Graph 3 for the ammonium indicator of the Oka water intake. The obtained regularities can be used for forecasting the dynamics of the studied indicators in the future.

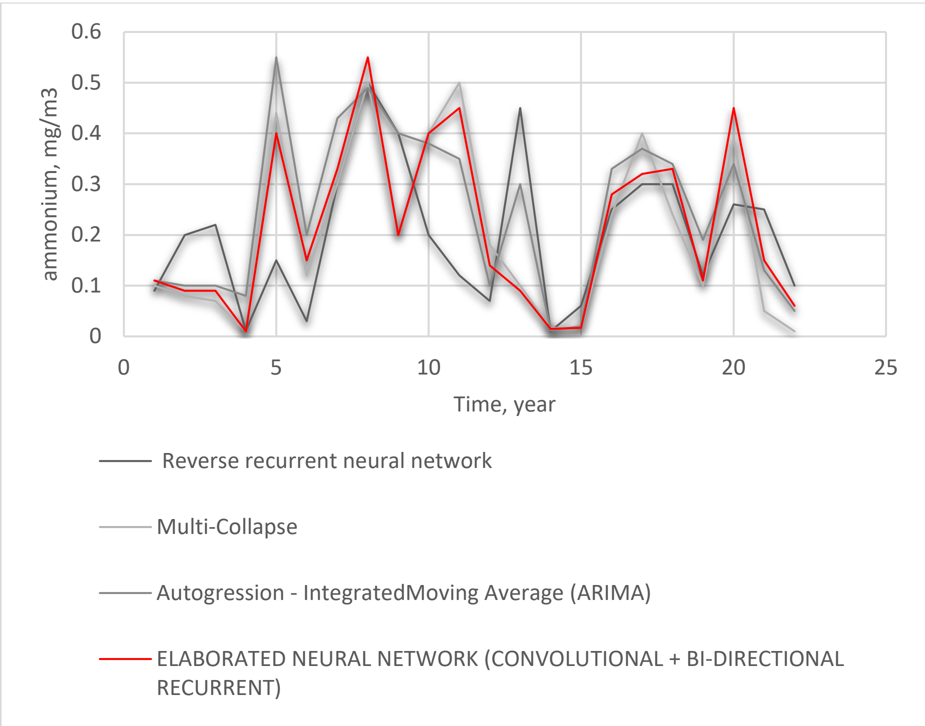


**Figure 4. Predictive results of the Okskiy water intake based on neural network on the content of ammonium in comparison with existing neural network models.**

This study proposes an alternative approach to water quality prediction based on the application of neural networks to analyze large historical data sets, which is a fundamentally different method from the traditional mechanistic models widely used in the field. Mechanistic models of water quality, which include such well-known systems as QUAL, WASP, MIKE, SWAT, BASINS and a number of others, are based on a detailed description of the structure of the water system under study and taking into account numerous constraints associated with the complex of physical, biological, and chemical processes occurring in the aquatic environment, which causes their high complexity and requires a significant amount of input information for the creation and subsequent solution of a system of equations describing the dynamics of changes in water quality in time and space [23, 24].

Despite their widespread use and acceptance in the scientific community, mechanistic models are characterized by a high degree of complexity and require a significant amount of input data, including numerous modelling parameters, conditions of pollutant sources and effluents, as well as other specific characteristics of the water system, which makes the process of building such models extremely time-consuming and the determination of optimal parameters difficult, significantly limiting their applicability to a wide range of watersheds.

The considered neural model, built on the basis of modern deep architectures, has shown high efficiency in solving this problem. Nonlinear multilayer data processing mechanisms underlying the model make it possible to identify complex interdependencies between water quality indicators and external factors, while forming statistically significant predictions. The conducted studies have confirmed a high degree of reliability of predictions, which is due to the model's ability to effectively analyze and predict non-linear processes under conditions of uncertainty.

Furthermore, the model is strongly multi-purpose, permitting its application to different watershed types, including rivers, lakes, and reservoirs. This significantly expands the potential use of the model for water quality monitoring and water resources management. There is an undeniable advantage of the proposed neural model over traditional numerical algorithms due to its higher prediction accuracy and computational efficiency. This model opens new perspectives for the development of promising approaches in the field of water resources monitoring and management.

### *3.3. Neural Network's Benefit for Crop Yields*

Combining CNN with Bi-RNN offers a powerful approach to analyzing river parameters, which can significantly benefit agriculture in various ways. Here's how the integration of these neural network architectures can support agricultural practices in the context of monitoring water quality and river health:

1. **Data Fusion and Multi-Scale Feature Extraction**

– **Convolutional Layers**: CNNs are excellent at capturing spatial features from high-dimensional data, such as images of rivers, satellite imagery, or time-series data related to river conditions. They can analyze patterns in the data, such as water color changes, sediment transport, or algal blooms.
– **Bi-RNNs**: Once the CNN has extracted spatial features, the Bi-RNN can analyze sequential data by processing it in both forward and backward directions. This dual processing helps capture temporal dependencies and patterns over time, such as changes in water quality or flow rates that occur seasonally or due to specific weather events.

2. **Enhanced Prediction of Water Quality**

– **Water Quality Metrics**: By processing data related to water quality parameters (e.g., pH, turbidity, temperature, and chemical pollutants), the combined model can predict how these factors may evolve over time. This predictive capability is crucial for early detection of issues, such as contamination or harmful algal blooms, which can directly affect agricultural practices.
– **Impact on Crop Health**: Understanding the temporal trends and spatial distribution of river parameters allows farmers to make informed decisions about irrigation practices, such as when to water crops or when to take remedial actions (e.g., treating water or altering planting schedules) to mitigate adverse effects.

3. **Forecasting Hydrological Patterns**

– **Flow and Flood Predictions**: The model was applied to forecast river flow patterns and potential flooding events using historical and current hydrological data. These forecasts provide quantitative information on flow dynamics and water level variations, which can support operational water management and early warning systems for flood risk mitigation.
– **Irrigation Management**: Understanding river flow patterns and water availability assists in optimizing irrigation practices, thereby conserving water and improving crop yields. Accurate

predictions can help determine the best times to irrigate, reducing waste and ensuring adequate moisture for crops.

4. **Integration of Remote Sensing Data**

- **Satellite Imagery Analysis**: CNNs processes satellite images to monitor changes in land use, vegetation, and water bodies. When combined with Bi-RNNs for time-series analysis, the model can reveal temporal patterns and dependencies between key water quality parameters, providing insights into the dynamics and interactions of the river ecosystem.
- **Environmental Monitoring**: By continuously integrating and analyzing real-time data from various sources (e.g., weather stations, sensor networks), this combined model can help monitor the ecological health of river systems and the surrounding agricultural lands.

5. **Adaptive Management Strategies**

- **Decision Support Systems**: The insights derived from this model can feed into decision support systems for farmers, offering recommendations based on predictive analytics. For example, farmers can receive alerts about potential water quality issues or optimal irrigation timings based on predicted river conditions.
- **Sustainability Practices**: By understanding how agricultural practices impact river health and vice versa, stakeholders can implement more sustainable practices that balance agricultural productivity with environmental conservation.

### *3.4. Statistical Parameters*

To evaluate the advantages and disadvantages of the proposed prediction model and other benchmark neural networks, such as ARIMA, LSTM, and reverse recurrent, the MAE, reflecting the average deviation of predicted values from actual values; MAPE, RMSE, accounting for the square of deviations; and R-squared were compared (Table 1). The developed model showed the strongest performance in comparison with the ARIMA reference neural network model and the reversible recurrent model. The strongest competitor was the multi-convergent model, which gave an average RMSE of 0.0557, while the developed model showed an error 0.0248 lower (i.e., average RMSE=0.0309).

***Table 1. Table captions should be placed above the tables.***

| Water quality | Statistical processing model parameters | Neural network type | | | |
|---|---|---|---|---|---|
| | | ARIMA | Reverse recurrent | LSTM | Elaborated (CNN + Bi-RNN) |
| COD, mgO2/l | R-squared | 0.9408 | 0.9920 | 0.9996 | 0.9996 |
| | RMSE | 1.2030 | 0.5360 | 0.0566 | 0.0299 |
| Total nitrogen, mg/l | R-squared | 0.8760 | 0.9933 | 0.9999 | 0.9996 |
| | RMSE | 1.0000 | 0.5400 | 0.0542 | 0.0315 |
| ammonium | R-squared | 0.9400 | 0.9945 | 0.9999 | 0.9999 |
| | RMSE | 0.9850 | 0.5466 | 0.0520 | 0.0310 |
| O2, mg/l | R-squared | 0.9308 | 0.9900 | 0.9996 | 0.9999 |
| | RMSE | 1.0000 | 0.5280 | 0.0600 | 0.0312 |

## *4. Conclusion*

Laboratory results indicated that water from the Sokolovskiy water intake had a significantly lower long-term average concentration of total ammonia equal to 0.48 mg/l. This value was significantly lower than similar values at the Okskiy and Borkovskiy water intakes, which were 1.6 and 2.1 times higher, respectively ($p < 0.05$). This value was significantly lower than similar values at the Okskiy and Borkovskiy water intakes, which were 1.6 and 2.1 times higher, respectively ($p < 0.05$). No significant differences were found in the long-term average values of COD and $BOD_5$ in the waters of the investigated water intakes. At the same time, the percentage of single samples, in which these parameters did not meet the established hygienic standards, ranged from 22.7 to 32.5 % for COD and from 61.8 to 75.0 % for $BOD_5$, respectively.

The developed neural network model was applied for the prediction of key water quality parameters (COD, total nitrogen, ammonium, and dissolved oxygen) at monitoring sites of the Oka River. Input variables included hydrological and meteorological data, as well as historical records of water chemistry. The model was trained and tested on time series from 2014–2022.

The evaluation demonstrated that the neural network provides reliable short-term forecasts of water quality. In particular, the combined CNN–Bi-RNN architecture achieved higher accuracy than ARIMA, simple recurrent, and LSTM models (Table 1). For example, RMSE values for ammonium prediction were reduced to 0.0310, while R-squared values approached 0.9999.

These results indicate that the proposed model can effectively capture nonlinear dependencies in hydrological and chemical datasets. The approach enables early detection of deviations from regulatory thresholds and can be integrated into monitoring systems for operational water quality control.

In order to ensure high reliability of the initial information base, advanced approaches to data processing, including the iForest algorithm and Lagrangian interpolation methods, were used at the preliminary stage of the study to not only effectively improve the integrity of the information set but also minimize the potential impact of errors and anomalies on the subsequent modelling process. In addition to pre-processing of the input data, the moving average method and PCA method were used to optimize water quality parameters and prevent the phenomenon of model overfitting, which is a critical factor for ensuring high accuracy of forecast calculations in the long term. Thus, we developed an innovative hybrid neural network model for predicting water quality parameters based on the integration of a deep CNN and a Bi-RNN, which consists of three functional parts.

The combination of CNNs and Bi-RNNs provides a robust framework for analyzing temporal and spatial patterns of key river water quality parameters. By leveraging spatial and temporal data, this integrated CNN–Bi-RNN approach improves the accuracy of water quality predictions and enables more comprehensive analysis of river dynamics, supporting effective monitoring and water management.

## References


1. Liao, Z., Wang, X., Zhang, Yu., Qing, H., Li, Ch., Liu, Q., Cai, J., Wei, Ch. An integrated simulation framework for NDVI pattern variations with dual society-nature drives: A case study in Baiyangdian Wetland, North China. Ecological Indicators. 2024. 158. Article no. 111584. DOI: 10.1016/j.ecolind.2024.111584
2. Talal, M., Alamoodi, A.H., Albahri, O.S., Albahri, A.S., Pamucar, D. Evaluation of remote sensing techniques-based water quality monitoring for sustainable hydrological applications: an integrated FWZIC-VIKOR modelling approach. Environment, Development and Sustainability. 2024. 26(8). Pp. 19685–19729. DOI: 10.1007/s10668-023-03432-5
3. Shivam, K., Tzou, J.-Ch., Wu, Sh.-Ch. Multi-step short-term wind speed prediction using a residual dilated causal convolutional network with nonlinear attention. Energies. 2020. 13(7). Article no. 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. 21(8). Pp. 1189–1195. DOI: 10.1016/j.engappai.2008.03.015
5. Ho, J.Yu., Afan, H.A., El-Shafie, A.H., Koting, S.B., Mohd, N.S., Jaafar, W.Z.B., Hin, L.S., Malek, M.A., Ahmed, A.N., Mohtar, W.H.M.W., Elshorbagy, A., El-Shafie, A. Towards a time and cost-effective approach to water quality index class prediction. Journal of Hydrology. 2019. 575. Pp. 148–165. DOI: 10.1016/j.jhydrol.2019.05.016
6. Juwana, I., Muttil, 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. 61(2). Pp. 170–178. DOI: 10.1016/j.ecolind.2015.08.034
7. Shamsutdinova, T.M. Application of Neural Network Modeling in Problems of Predicting the Level of River Floods. Vestnik NSU. Series: Information technologies. 2023. 21(2). Pp. 39–50. DOI: 10.25205/1818-7900-2023-21-2-39-50
8. Shitikov, V.K., Zinchenko, T.D., Golovatyuk, L.V. Neural network methods of surface water quality assessment by hydrobiological indicators. Izvestia Samara Scientific Centre of the Russian Academy of Sciences. 2002. 4(2). Pp. 280–289.
9. Noor, S.S.M., Saad, N.A., Akhir, M.F.M., Rahimet, M.S.A. QUAL2K water quality model: A comprehensive review of its applications and limitations. Environmental Modelling & Software. 2024. 184. Article no. 106284. DOI: 10.1016/j.envsoft.2024.106284
10. Vorobyev, S.N., Pokrovsky, O.S., Korets, M., Shirokova, L.S. A snap-shot assessment of carbon emission and export in a pristine river draining permafrost peatlands (Taz River, Western Siberia). Frontiers in Environmental Science. 2022. 10. Article no. 987596. DOI: 10.3389/fenvs.2022.987596
11. Kaal, J., González-Pérez, J.A., San Emeterio, L.M. Fingerprinting macrophyte Blue Carbon by pyrolysis-GC-compound specific isotope analysis (Py-CSIA). Science of the Total Environment. 2025. 836. Article no. 155598. DOI: 10.1016/j.scitotenv.2022.155598
12. Rosenthal, O.M., Fedotov, V.Kh., Identification of water polluting enterprises based on neural network analysis. Prirodoobustrojstvo. 2023. 1. Pp. 62–68. DOI: 10.26897/1997-6011-2023-1-62-68
13. Shamsutdinova, T.M. Application of Neural Network Modeling in Problems of Predicting the Level of River Floods. Vestnik NSU. Series: Information Technologies. 2023. 21(2), Pp. 39–50. DOI: 10.25205/1818-7900- 2023-21-2-39-50
14. Shitikov, V.K., Zinchenko, T.D., Golovatiyk, L.V. Methods of Neural Networks for Estimation of Superficial Waterís Quality by Usage of Hydrobiological Exponents. Institute of Ecology of the Volga River Basin of Russian Academy of Sciences. 2002. 4(2). Pp. 280– 289.
15. 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. Hygiene and Sanitation. 2021. 100(11). Pp. 1192–1197. DOI: 10.47470/0016-9900-2021-100-11-1192-1197
16. Jiang, Y., Li, Ch., Sun, L., Guo, D., Zhang, Yi., Wanget, W. A deep learning algorithm for multi-source data fusion to predict water quality of urban sewer networks. Journal of Cleaner Production. 2021. 318. Article no. 128533. DOI: 10.1016/j.jclepro.2021.128533

17. 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. 602. Article no. 126817. DOI: 10.1016/j.jhydrol.2021.126817
18. Liu, H., Zhang, F., Tan, Y., Huang, L., Li, Y., Huang, G., Luo, Sh., Zeng, A. Multi-scale quaternion CNN and BiGRU with cross self-attention feature fusion for fault diagnosis of bearing. Measurement Science and Technology. 2024. 35(8). Article no. 086138. DOI: 10.1088/1361-6501/ad4c8e
19. Lu, X., Dong, Yu, Liu, Q., Zhu, H., Xu, X., Liu, J., Wang, Yi. Simulation on TN and TP Distribution of Sediment in Liaohe Estuary National Wetland Park Using MIKE21-Coupling Model. 2023. Water. 15(15). Article no. 2727. DOI: 10.3390/w15152727
20. Tiyasha, Tung, T.M., Yaseen, Z.M. Deep Learning for Prediction of Water Quality Index Classification: Tropical Catchment Environmental Assessment. Natural Resources Research. 2021. 30(6). Pp. 4235–4254. DOI: 10.1007/s11053-021-09922-5
21. Wongburi, P., Park, J.K., 2023. Prediction of Wastewater Treatment Plant Effluent Water Quality Using Recurrent Neural Network (RNN) Models. Water. 15(19). Article no. 3325. DOI: 10.3390/w15193325
22. Zhu, M., Wang, J., Yang, X., Zhang, Y., Zhang, L., Ren, H., Wu, B., Ye, L. A review of the application of machine learning in water quality evaluation. Eco-Environment & Health. 2022. 1(2). Pp. 107–116. DOI: 10.1016/j.eehl.2022.06.001
23. Gao, L., Biderman, S., Black, S., Golding, L., Hoppe, T., Foster, C., Phang, J., He, H., Thite, A., Nabeshima, N., Presser, Sh., Leahy, C. The Pile: An 800GB Dataset of Diverse Text for Language Modeling. arXiv. 2020. arXiv:2101.00027v1. DOI: 10.48550/arXiv.2101.00027
24. Da Silva, A.C., das Graças Braga da Silva, F., de Mello Valério, V.E., Silva, A.T.Y.L., Marques, S.M., dos Reis, J.A.T. Application of data prediction models in a real water supply network: comparison between arima and artificial neural networks. Revista Brasileira de Recursos Hídricos. 2024. 29. Article no. e12. DOI: 10.1093/nar/gks1219


***Information about the authors:***


***Anna Naumova***,
*ORCID: https://orcid.org/0000-0002-0373-8655*
*E-mail: koshevaya81@mail.ru*

***Vitaliy Ilinich***, *PhD in Technical Sciences*
*ORCID: https://orcid.org/0000-0003-2094-2882*
*E-mail: vilinitch@gmail.com*

***Margarita Shiryaeva***,
*ORCID: https://orcid.org/0000-0001-8019-1203*
*E-mail: margaretshiryaeva@gmail.com*


*Received 19.11.2024. Approved after reviewing 10.09.2025. Accepted 10.09.2025.*