Eco. Env. & Cons. 30 (3) : 2024; pp. (468-482)
Copyright@ EM International
ISSN 0971–765X

Integrated Ecological Assessment of A River Environment Based on Water Quality Criteria and Pollution Indicators Analysis

Kateryna Berezenko[1], Dmytro Sopov[2], Svitlana Melnychuk[3], Nataliya Gurets[3], Olena Girzheva[3], Oleksandr Marynets[3], Oleksandr Ignatenko[4], Iryna Dzhygyrey[5], Iryna Cherådnychenko[6] and Oleksii Prokopenko[7]

[1]Department of Horticulture and Ecology Luhansk Taras Shevchenko National University, Ukraine
[2]Department of Chemistry, Geography and Earth SciencesLuhansk Taras Shevchenko National University, Ukraine
[3]Department of Ecology and Environmental Technology, Admiral Makarov National University of Shipbuilding, Ukraine
[4]Department of Public Management and Administration State University of Information and Ñommunication Technologies, Ukraine
[5]Educational and Research Institute for Applied System Analysis National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine
[6]Department of Chemistry, Geography and Earth SciencesLuhansk Taras Shevchenko National University, Ukraine
[7]Department of Political Science and Cultural Studies Eastern Ukrainian National University named after V. Dalya, Ukraine

(Received 21 April, 2023; Accepted 4 June, 2024)

ABSTRACT

This study presents an integrated ecological assessment of water quality in the Southern Bug River basin, employing analyses of water quality criteria and pollution indicators. Through meticulous examination, the research identifies various factors pivotal to the river's water quality. Key findings underscore the significant influence of physicochemical parameters on the river's water resources. Two distinct water quality prediction models were developed: a traditional model and a hydrochemical regime-incorporating model. While the traditional model generally exhibits superior accuracy, the hydrochemical regime model demonstrates heightened precision, particularly in scenarios characterized by abrupt environmental changes. Although the hydrochemical regime model predicts with slightly lower accuracy (70-75%) compared to the traditional model, it excels during sudden anthropogenic alterations in water resources, achieving accuracy levels of 80-85%. These results underline the substantial impact of the hydrochemical regime on prediction accuracy and emphasize its crucial role in evaluating water quality. Moreover, the study addresses pollution prediction in the Southern Bug River environment, facilitating proactive responses to potential threats to aquatic ecosystems and public health. An integrated approach to water quality analysis, considering various factors and developing a spatial model of river flow, significantly enhances precision in identifying and understanding variations in water quality, which is imperative for effective water resource management. The insights gleaned from this research provide valuable information for policymakers, stakeholders, and environmental managers tasked with preserving and managing water resources sustainably. By shedding

***Corresponding author's e-mail:** dmitrosopov@gmail.com

light on the complex dynamics of river basins, this study contributes to the advancement of scientific knowledge and the development of strategies for safeguarding freshwater ecosystems.

***Key words**: Pollution indicators,Prediction model, Accuracy, River hydrochemical regime, Quality criteria*

Introduction

Amidst the backdrop of mounting global ecological challenges, ensuring the quality of water resources emerges as a paramount scientific concern. Excessive regulation, population growth, and the exploitation of natural resources have led to a significant deterioration in water quality and the condition of aquatic ecosystems. In this context, integrated ecological assessment becomes an indispensable component of strategies aimed at conserving and restoring water resources(Reid *at al*., 2018).

The preservation of viable aquatic ecosystems stands as a critical objective, as they not only constitute key elements of biodiversity but also play a pivotal role in meeting the needs of both populations and economic sectors for clean and safe water. The examination of water quality criteria and pollution serves as a crucial step in identifying tangible threats and formulating effective management measures (Hossain *at al*., 2020).

The imperative for conducting research on this subject is underscored by the menacing impact of both anthropogenic and natural factors on water resources. This relevance is underscored by several factors. Firstly, a multitude of natural and anthropogenic factors exert influence over the state of aquatic ecosystems. Changes in water quality and pollution levels serve as vital indicators of the ecological health of river systems. Secondly, water pollution directly impacts biota and ecosystem health, while also posing significant risks to human health through its use in households and agriculture.

This underscores the urgency of investigating the worsening water quality and ecological resilience of rivers amidst anthropogenic pressures and ecological crises. The analysis of water quality criteria represents a crucial stage in comprehending the ecological condition of the region.

A systematic analysis of water quality criteria and pollution indicators serves as a prerequisite for the development of effective strategies aimed at conserving and restoring ecosystems, as well as ensuring the sustainable utilization of water resources within the region.

Such research transcends mere academic interest, intersecting with practical environmental protection measures and water resource management practices. The resolution of river ecological status challenges and the formulation of effective protective measures are determined not only by the imperatives of nature but also by the harmonious preservation of water resources to meet societal needs.

The insights garnered from such research are indispensable in practice, as they not only delineate the state of river water environments but also furnish the basis for devising effective water resource management strategies. The assessment of pollution and identification of key influencing factors will facilitate the identification of critical aspects necessitating immediate action to safeguard biodiversity.

The data derived from this research will aid in formulating specific measures to protect water resources, enhancing ecological monitoring, and augmenting the effectiveness of strategies aimed at restoring river ecosystems.

Scientific inquiry into the integrated ecological assessment of river water environments, with a focus on water quality criteria and pollution indicators, represents a crucial stride towards ensuring sustainable water utilization and ecosystem preservation. Its outcomes will serve as a pivotal resource in addressing environmental challenges and enhancing overall water quality.

Therefore, research endeavors dedicated to assessing the pollution status of rivers remain pertinent, as they furnish the groundwork for formulating specific measures aimed at safeguarding water resources, refining ecological monitoring practices, and implementing more efficacious strategies for the restoration of river ecosystems. Such research endeavors will contribute to ensuring sustainable water utilization and ecosystem preservation, both of which are indispensable in tackling environmental challenges and enhancing water quality on a broader scale.

Literature Review and Problem Statement

Developing an effective and precise water quality model proves to be a significant yet challenging task due to variations in river hydrology and the presence of anthropogenic influences in real-world conditions. Traditional data processing methods and statistical forecasting models often fail to address these complexities, especially given the ambiguity and non-linearity of water quality parameters.

Deterministic models aimed at reproducing chemical and physical processes have their limitations due to the need for simplifications in decision-making. While statistical models seek general rules based on empirical data, they can be intricate and require substantial data volumes.

Given the limitations of traditional methods, artificial intelligence (AI) is often considered as an alternative. Models developed based on the correlation of input and output data enable overcoming some constraints associated with conventional methods.

Thus, conducting research devoted to the development and implementation of AI models for effective modeling of complex nonlinear systems, such as assessing water quality in rivers, is deemed worthwhile. The application of AI can aid in addressing the challenges posed by variability and ambiguity in real-world conditions, ensuring accurate and reliable results.

The use of artificial intelligence opens new possibilities for accurate modeling of complex river water quality systems. The application of AI in ecological research can help overcome challenges associated with variability and non-linearity of parameters in real-world conditions.

Study (Lu *et al.*, 2023) analyzes spatial and temporal variations in water quality in the middle and lower reaches of the Han River in China. The study employs Water Quality Index (WQI) and Minimum WQI methods to comprehensively analyze the state of water resources. The uniqueness of the article lies in the wide range of parameters used for analysis, as well as the comparison between the traditional WQI method and the enhanced WQImin method.

However, the study leaves several unresolved questions. Particularly, there is a lack of detailed analysis of the causes of seasonal variations in water quality and their impact on the environment and society. Additionally, the study does not provide clear conclusions and recommendations for further actions, which may complicate the practical implementation of the obtained results. Objective reasons for these unresolved issues may include limited resources for research, insufficient data availability, and the complexity of analyzing seasonal variations.

Study (Yang *et al.*, 2021) of spatiotemporal variations in water quality in the Nanfei River establishes links between various water quality parameters and their impact on the overall water quality index. However, there are several aspects that remain unresolved in this study or could be further elaborated.

Firstly, the article does not consider the impact of geographical factors on water quality. Although it is noted that water from different sources has different qualities, the specific influence of geographical conditions on pollution is not addressed. This limitation may be due to limited resources for research or a lack of access to sufficient data. Another limitation is the limited list of water quality parameters analyzed. While the article emphasizes the importance of considering various parameters such as COD, TN, TP, etc., it may be beneficial to expand the list of analyzed parameters or explore their interaction in greater depth.

A third aspect is the absence of specific recommendations regarding water quality management strategies. Although the authors emphasize the importance of considering various factors when making decisions to reduce pollution, the article does not provide specific recommendations on how this can be done.

It is likely that these points were left unresolved due to limitations in the methods of analysis. Objective reasons for these unresolved issues may include limited data availability, particularly regarding specific sources of pollution and their impact on water quality. Additionally, there may be difficulties in determining the relationships between different factors, such as hydrological conditions and pollution levels.

Of particular interest is the study (Kim *et al.*, 2017), which evaluates changes in water quality in a monitoring network using exploratory factor analysis and empirical orthogonal function.

The main advantages of this study lie in the use of innovative analysis methods such as exploratory factor analysis and empirical orthogonal function, which allow for a more accurate identification and understanding of variations in water quality. Additionally, the use of these methods can contribute to improving the monitoring of water resources and the adoption of water quality management measures.

However, there are aspects that remain unresolved in this study. Primarily, there is no clear methodology for selecting the monitoring network and water quality parameters, which may lead to the misrepresentation of water systems. Additionally, the study does not investigate the impact of external factors such as climate change or anthropogenic pollution on water quality, which could affect the reliability of the obtained results.

For future research, it is important to address these issues and take appropriate measures to resolve them. For example, it may be necessary to develop a standardized methodology for selecting monitoring points and water quality parameters based on the diversity of geographical and anthropogenic factors.

As we can see, traditional studies (Lu *et al.*, 2023; Yang *et al.*, 2021; Kim *et al.*, 2017) have limitations in terms of accuracy due to data constraints. Indeed, conducting continuous monitoring is costly, labor-intensive, and in some cases not rational, yet it is essential for identifying current dynamics. Addressing this issue could involve the application of machine learning, which can generate additional necessary data for assessment and precise forecasting. Therefore, research based on AI plays a significant role in this research direction.

The authors of article (Zhu *et al.*, 2019) propose a model based on Extreme Learning Machine (ELM) for predicting daily water temperature in rivers. Air temperature (Ta), discharge (Q), and day of the year (DOY) are used as predictors. Data from three rivers with different hydrological conditions were used to test the models. The results showed that the inclusion of all three input parameters (Ta, Q, and DOY) yielded the best modeling accuracy for all developed models. Additionally, it was found that Q played a minor role, while Ta and DOY were the most important explanatory variables for predicting water temperature in rivers.

Positive aspects of the study include the innovative approach using ELM for predicting water temperature in rivers. The results showed that ELM and MLPNN models outperformed the MLR model. Additionally, it was found that sigmoid and radial basis activation functions performed best for predicting water temperature in rivers. However, despite its high accuracy, the conducted research is extremely labor-intensive.

Article (Abba *et al.*, 2020) examines the application of various models, such as Genetic Programming (GP), Extreme Learning Machine (ELM), and XGBoost (XGB), for predicting Water Quality Index (WQI) in the Kinabatangan River basin, Malaysia. The authors compare the results of the constructed models with the classical method of linear regression (LR). Positive aspects include the efficiency of GP and XGB as tools for selecting input variables, as well as improved prediction accuracy using ELM. However, using LR after variable selection may degrade prediction accuracy. A negative factor is the limitation of the study: the use of only one data source and the absence of analysis of the impact of anthropogenic activity on water quality. To overcome this, it is possible to expand the data scope, utilize diverse sources of information, and consider the impact of human activity on water resources in more detail.

The methodology described in article (Kim *et al.*, 2015) is based on the combination of clustering methods and Artificial Neural Networks (ANN) for predicting water quality in rivers. The main idea is to use clustering to construct balanced data sets for training ANN and thus improve prediction accuracy.

There are certain advantages to using this methodology, such as reducing modeling errors through the use of balanced data sets and improving model accuracy, as demonstrated compared to results of non-clustered ANN. However, there are several unresolved issues that require further research.

Firstly, the impact of the number of clusters on modeling results has not been investigated. Additionally, the stability of the methodology's results under different conditions is not considered. It is important to determine how different clustering parameters may affect the effectiveness of the methodology.

Furthermore, the methodology should be compared with other modern water quality prediction methods to obtain more objective results. Finally, there is a need to study the influence of clustering parameters on modeling results and address unresolved issues.

For further research, it is recommended to conduct a systematic analysis of methodology parameters. It is also crucial to expand the data scope to enhance stability and generalizability of results. Additionally, it is necessary to compare it with other methods and develop new clustering approaches specifically adapted for water quality data.

Article (Asadollah *et al.*, 2021) addresses the effectiveness of three machine learning models (Decision Tree Regression, Support Vector Regression, and Extra Tree Regression) in predicting Water Quality Index (WQI). As a practical example, the study focuses on the Lam Tsuen River in Hong Kong based on water quality indicators measured monthly. The research demonstrates that the Extra Tree Regression model with ten input parameters achieves the highest accuracy in predicting WQI. Moreover, it shows the capability to significantly reduce computational load compared to other methods and potentially provide effective real-time water quality monitoring. However, the study may be limited by the selected set of chemical and physical parameters, as well as data sampling requirements. Additionally, the article does not account for potential differences in climatic and hydrological conditions of other rivers.

Interest in prediction under data scarcity is highlighted in study (Kouadri *et al.*, 2021). The authors evaluate groundwater quality to ensure safe drinking water sources. To address data limitations, the authors utilized 8 artificial intelligence algorithms, including Multilinear Regression (MLR), Random Forest (RF), M5P tree, Random Subspace (RSS), Additive Regression (AR). They also employed Locally Weighted Linear Regression (LWLR), Artificial Neural Network (ANN), and Support Vector Regression (SVR) to predict WQI in the Illizi region, southeastern Algeria. The article discusses the results and discussion of the research on water resources, particularly the chemical composition of groundwater. The study is based on the investigation of 114 groundwater samples collected from 57 developed wells across 6 different layers. However, a drawback of the study is the neglect of well locations. It should be understood that water bodies are not independent; they interact, forming a complex network of hydrological processes and cycles. Therefore, their hydrochemical regime should be considered not only separately but also in the context of their interaction to gain a fuller understanding of aquatic ecosystems and ensure effective water resource management.

Overall, according to analytical works devoted to AI in water resources (Tiyasha, Minh Tung *et al.*, 2020; Yan *et al.*, 2024), there are some unresolved issues:

Model accuracy issues: Analysis of articles reveals that although artificial intelligence models such as Artificial Neural Networks (ANN) and logic-based reasoning demonstrate high accuracy in predicting water quality in rivers, they may inadequately reflect complex interactions between various water parameters and environmental factors. For instance, they may underestimate the impact of certain pollutants or fail to account for heterogeneity in water resources. This could be an objective reason related to the complexity of physicochemical processes in the aquatic environment, which are challenging to model.

Insufficient attention to forecasting future changes: Many studies focus on predicting the current state of water quality in rivers. However, little attention is paid to forecasting future changes, such as the impact of climate change or human activities on water quality. This could be an objective reason, as forecasting future changes requires a significant amount of additional data and complex models, which may be inaccessible or underestimated.

Inadequate consideration of geographical and cultural specificity: Studies are conducted in various geographical areas and cultural contexts, which may lead to ambiguous results or inconsistent models. This could be a subjective reason, as considering geographical and cultural characteristics can be a challenging task that requires deeper analysis and consideration of local conditions.

We may conclude that there is a notable dearth of studies assessing the holistic impact of anthropogenic pressures on rivers. While numerous investigations delve into specific facets of aquatic environments, they often fail to account for interrelations and the synergistic influence of diverse factors on ecosystems. Consequently, it is imperative to undertake research aimed at scrutinizing the comprehensive ecological state of the Southern Bug River.

Hence, our study proposes conducting an integrated ecological appraisal of the aquatic milieu through a synthesis of water quality criteria analysis and computational modeling, aiming for a more profound comprehension of the multifaceted impacts on riverine ecosystems. Notably, conventional methodologies frequently fall short of meeting the demands of water quality modeling.

The deployment of artificial intelligence (AI) holds promise as a viable remedy to circumvent these limitations, serving as a potent tool for yielding precise and credible outcomes. Therefore, it is recommended to embark on investigations dedicated to crafting AI models for the comprehensive ecological assessment of water quality in rivers. However, the challenge persists in delineating optimal parameters and tailoring AI models to the specific hydrochemical dynamics of the river.

Materials and Methods

Research Objectives and Goals

The aim of this research is to conduct a comprehensive ecological assessment of the riverine environment, utilizing an analysis of water quality criteria and pollution indicators. The study seeks to enhance understanding of the anthropogenic pressures on the river ecosystem and to identify key aspects of pollution and their interrelationships. The outcomes will contribute to the development of effective conservation and restoration strategies for water resources, fostering sustainable water management in the region.

To achieve this goal, the following objectives were set:

To investigate the impact of physico-chemical water quality criteria of the river, focusing on indicators reflecting the ecological status of the river system.

To develop a water quality prediction model taking into account the hydrochemical regime of the Southern Bug River.

Study method

The object of the study is the basin of the Southern Bug River within the Vinnytsia region. The research hypothesis assumes that anthropogenic factors and natural elements significantly interact, leading to changes in water quality and causing imbalances in the river's aquatic ecosystems.

The underlying concept of the study is based on the spatial structure of the input-output model framework.

To develop a water quality assessment model for the Southern Bug River, key input parameters that have a substantial impact on the river's water quality were identified. Considering a critical literature review, the following parameters were determined for modeling the water quality of the Southern Bug River: ammonium ion content, five-day biochemical oxygen demand, suspended solids, and dissolved oxygen. Additionally, nitrate ions, nitrite ions, sulfate ions, phosphate ions (polyphosphates), and chloride ions were taken into account. The selection of these input parameters is based on prior knowledge and statistical analysis of potential inputs.

Furthermore, the proposed model includes a parameter that considers the influence of the hydrochemical regime of the Southern Bug River. Existing models do not account for internal processes; instead, they develop models through the correlation of input and output data based on the value of the same parameter at the previous monitoring point. However, typically, discharge through a local area from an upstream monitoring point affects water pollution relating to a downstream point. Therefore, the proposed model focuses on the possibility of considering the influence of water quality parameters at an upstream monitoring point.

Thus, the developed water quality prediction model at each monitoring point incorporates not only nine input parameters - chemical water indicators measured at monitoring points, but also an additional indicator - water quality at previous points. At the previous points (upstream), the predicted water pollution index (WPI) is calculated according to equation (1):

$$WPI = \frac{1}{9}\sum_{i=1}^{n}\frac{c_i}{PC_i} = \frac{c_1}{2} + \frac{c_2}{3} + \frac{c_3}{9} + \frac{4}{c_4} + \frac{c_5}{45} + \frac{c_6}{3.3} + \frac{c_7}{500} + \frac{c_8}{3.5} + \frac{c_9}{500}, \quad ..\ (1)$$

where c_i–represents the concentration of the normative component, c_1–concentration NH_4^+, c_1–concentration of biological oxygen demand (BOD5), c_2 –indicates the concentration of suspended solids, c_3–signifies the concentration of dissolved oxygen O_2, c_4 – concentration of nitrates NO_3^-, c_5–NO_2^- - concentration of phosphates, c_6–reflects the concentration of sulfates SO_4^{2-}, c_7–represents the concentration of ammonia PO_4^{3-}, c_8–the concentration of chloride Cl; PC_i– signifies the established concentration value of the component for the respective water body type, measured in mg/dm^3; n – denotes the number of indicators utilized in calculating the WPI.

Figure 1 provides a schematic depiction of the parameter selection approach for the models.

This procedure is repeated, incorporating the predicted WPI, for the second, third, and fourth monitoring points downstream.

As a practical implementation, a ANN model was developed for four monitoring points along the Southern Bug River. The model was constructed using the Neural Network Toolbox tool within the Matlab environment. The accuracy of the model was

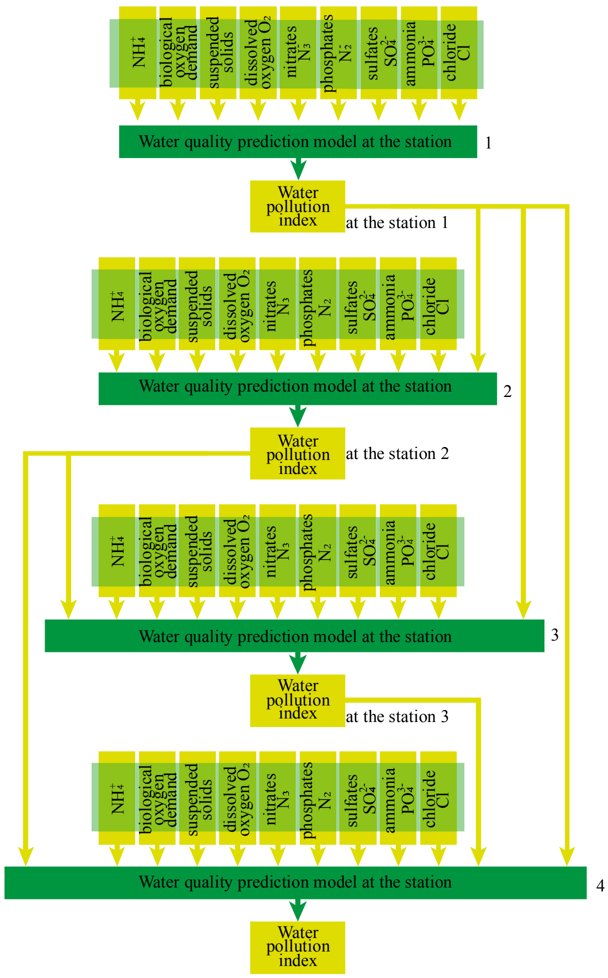

Fig. 1. Schematic representation of parameter selection for water quality prediction models of the Southern Bug River

evaluated using the coefficient of determination (R^2) and the mean squared error (MSE).

Results

Impact of physico-chemical water quality criteria on the Southern Bug River

During the investigation of the influence of physico-chemical water quality criteria on the Southern Bug River, an analysis of indicators was conducted at four monitoring points. The results indicate significant differences in the levels of individual indicators between monitoring points, suggesting diverse influences of anthropogenic and natural factors in different areas. Specifically, the comparison of average indicator values at different monitoring points indicates that the Khmilnyk and Hushchyntsi points have higher average levels of ammonium ions, BOD5, and nitrates compared to other monitoring points. This may indicate possible sources of industrial or agricultural pollution in these areas.

On the other hand, the Mankivka point is characterized by lower levels of ammonium ions and BOD5, suggesting a lesser influence of anthropogenic emissions sources in this basin area.

In particular, ammonium ions have an average value ranging from 0.42 to 0.72 mg/dm³, with the maximum level recorded at the Khmilnyk point and the minimum at the Mankivka point. BOD5 ranges from 5.25 to 7.00 mgO_2/dm^3, with the highest recorded value at the Khmilnyk point.

The main statistical parameters (mean (μ), minimum (min), maximum (max), standard deviation (σ), arithmetic mean ($\bar{x}$), and coefficient of variation (CV)) for input and output parameters are presented in Table 1.

Average values and range indicators provide an overview of substance concentrations in the waters of the Southern Bug River at different monitoring points.

To determine the degree of concentration variation of parameters, the coefficient of variation (CV) is used. Large CV values (ranging from 109.59% to 355.84%) indicate significant diversity in substance concentrations in the water, which may be associated with different sources of pollution and spatial non-uniformity in their distribution.

The general trend shows that nitrate, nitrite, and phosphate levels exhibit significant variability among monitoring points. Sulfate ions and chloride ions also demonstrate variability in distribution along different stretches of the river.

The significant differences in parameter concentrations suggest various pollution sources across the extensive geographical range of the river basin. The river channel passes through different settlements, tributaries, and sewage channels that discharge wastewater into the main river channel.

For modeling efficiency, a statistical correlation analysis was used between input and output parameters. For example, concentrations of ammonium ions and BOD5 exhibit a high degree of correlation.

The most significant influence on forecasting results comes from oxygen-containing nitrates and phosphates. Correlation analysis was conducted by constructing a correlation matrix (Table 2). This correlation matrix represents a table of correlation coef-

Table 1. Main physico-chemical indicators of the Southern Bug River at monitoring points Khmilnyk, Hushchyntsi, Sabarivske, and Mankivka

Indicator	Unit	μ	max	min	σ	x̃	CV
Monitoring Point 1 (Khmilnyk)							
NH_4^+	mg/dm³	0,72	0,00	3,30	0,79	0,72	109,59 %
BOD5	mg O_2/dm³	6,57	3,70	13,30	2,22	6,57	33,77 %
Suspended solids	mg/dm³	11,50	5,00	25,00	5,07	11,50	44,13 %
O_2	mg O_2/dm³	6,91	2,98	11,30	2,09	6,91	30,26 %
NO_3^-	mg/dm³	1,99	0,00	7,71	2,05	1,99	102,82 %
NO_2^-	mg/dm³	0,32	0,00	6,79	1,13	0,32	355,84 %
SO_4^{2-}	mg/dm³	33,53	21,90	47,30	7,05	33,53	21,02 %
PO_4^{3-}	mg/dm³	0,64	0,03	3,05	0,73	0,64	113,75 %
Cl^-	mg/dm³	39,09	22,70	61,10	6,78	39,09	17,35 %
Monitoring Point 2 (Hushchintsy)							
NH_4^+	mg/dm³	0,72	0,00	3,00	0,75	0,72	104,08 %
BOD5	mg O_2/dm³	6,05	3,50	12,20	1,99	6,05	32,85 %
Suspended solids	mg/dm³	11,44	5,00	25,00	4,80	11,44	41,93 %
O_2	mg O_2/dm³	7,08	3,40	13,90	2,09	7,08	29,48 %
NO_2^-	mg/dm³	1,95	0,00	8,56	2,33	1,95	119,29 %
NO_3^-	mg/dm³	0,11	0,00	0,63	0,12	0,11	103,67 %
SO_4^{2-}	mg/dm³	33,73	21,20	57,17	8,00	33,73	23,70 %
PO_4^{3-}	mg/dm³	0,53	0,03	3,05	0,68	0,53	128,98 %
Cl^-	mg/dm³	38,72	22,60	49,50	5,60	38,72	14,46 %
Monitoring Point 3 (Sabarivske)							
NH_4^+	mg/dm³	0,70	0,10	3,62	0,75	0,70	107,21 %
BOD5	mg O_2/dm³	7,00	3,80	19,00	3,25	7,00	46,48 %
Suspended solids	mg/dm³	11,06	4,00	24,00	4,51	11,06	40,80 %
O_2	mg O_2/dm³	7,04	1,70	14,50	2,85	7,04	40,42 %
NO_3^-	mg/dm³	2,11	0,00	13,79	3,00	2,11	142,39 %
NO_2^-	mg/dm³	0,11	0,01	0,55	0,12	0,11	110,02 %
SO_4^{2-}	mg/dm³	34,77	18,40	60,15	9,15	34,77	26,33 %
PO_4^{3-}	mg/dm³	0,50	0,04	2,39	0,60	0,50	119,92 %
Cl^-	mg/dm³	37,81	28,90	53,80	4,69	37,81	12,41 %
Monitoring Point 4 (Mankivka)							
NH_4^+	mg/dm³	0,42	0,10	1,30	0,27	0,42	63,82 %
BOD5	mg O_2/dm³	5,25	2,40	11,30	1,96	5,25	37,38 %
Suspended solids	mg/dm³	7,59	4,00	19,00	3,26	7,59	42,91 %
O_2	mg O_2/dm³	8,69	2,50	15,70	3,14	8,69	36,17 %
NO_3^-	mg/dm³	1,72	0,00	4,79	1,47	1,72	85,10 %
NO_2^-	mg/dm³	0,10	0,00	0,78	0,14	0,10	132,91 %
SO_4^{2-}	mg/dm³	35,33	22,30	53,10	6,42	35,33	18,17 %
PO_4^{3-}	mg/dm³	0,46	0,03	3,50	0,60	0,46	129,95 %
Cl^-	mg/dm³	44,52	23,50	57,50	8,86	44,52	19,91 %

ficients between various chemical parameters in the water. Each element of the matrix indicates the degree of correlation between two parameters. The correlation coefficient can range from -1 to 1, where -1 indicates a complete negative correlation, 1 indicates a complete positive correlation, and 0 indicates no correlation.

The following parameters were analyzed: concentrations of ammonium ions, biochemical oxygen demand for five days, suspended solids, dissolved oxygen, nitrate ions, nitrite ions, sulfate ions, phosphate ions (polyphosphates), and chloride ions.

Regarding the correlation between BOD5 and nitrate ion content, a weak positive correlation (0.37) was found, suggesting that an increase in nitrate ions may lead to an increase in biochemical oxygen

Table 2. Correlation Matrix

Indicator	NH_4^+	BOD5	Suspended	O_2	NO_2^-	NO_2^-	SO_4^{2-}	PO_4^{3-}	Cl^-
NH_4^+	1	0,11	0,01	–0,15	0,37	0,25	0,08	0,57	0,13
BOD5	0,11	1	0,36	–0,36	–0,1	–0,06	–0,29	0,09	0,03
Suspended solids	0,01	0,36	1	–0,21	–0,21	–0,07	–0,01	0,07	–0,05
O_2	–0,15	–0,36	–0,21	1	0,25	–0,03	0,36	–0,14	–0,05
NO_3^-	0,37	–0,1	–0,21	0,25	1	0,23	0,16	0,09	0,01
NO_2^-	0,25	–0,06	–0,07	–0,03	0,23	1	0,02	0,07	–0,05
SO_4^{2-}	0,08	–0,29	–0,01	0,36	0,16	0,02	1	–0,15	–0,03
PO_4^{3-}	0,57	0,09	0,07	–0,14	0,09	0,07	–0,15	1	0,06
Cl^-	0,13	0,03	–0,05	–0,05	0,01	–0,05	–0,03	0,06	1

demand.

The relationship between suspended solids and dissolved oxygen indicates a weak negative correlation (-0.21), implying that an increase in suspended solids may result in a decrease in dissolved oxygen.

There is a moderate positive correlation (0.57) between phosphate ion content and nitrate ions, indicating that an increase in phosphate ions may influence an increase in nitrate ions. A weak positive correlation (0.36) exists between sulfate ions and dissolved oxygen, suggesting that an increase in sulfate ions may contribute to an increase in dissolved oxygen.

This analysis aids in identifying optimal input parameters for the ANN model, relying on statistical relationships. Although the correlation dependency does not essentially imply a physical meaning according to the weight percentage, it indicates that all variables significantly influenced the evaluation of variable outcomes. The contribution of each indicator ranged from 5 to 14%.

Construction of water quality prediction model considering the hydrochemical regime of the river

Two approaches to the structure of ANN models were used for comparison. The first approach involved the construction of a traditional ANN model for rivers: comparing solely physico-chemical indicators and water quality indicators of a single monitoring point (Mankivka). In the second approach (proposed), in addition to physico-chemical indicators and water quality indicator of monitoring point No. 4, water quality indicators of monitoring points No. 3, No. 2, and No. 1 were also included as input parameters.

For practical implementation of the proposed method, ANN modeling was performed for monitoring point Mankivka. The influence of water quality indicators from Khmilnyk, Hushchintsy, and Sabarivske monitoring points was also considered.

Models for monitoring point Mankivka were constructed based on ANN with architecture. For the first variant, a model with dimensions 10-8-10 was used, where 10 nodes corresponded to input parameters, 8 nodes in the hidden layer, and 10 nodes in the output layer representing target variables. For the second variant, a model with architecture 13-8-10 was applied.

Sigmoid and linear activation functions were used

Graphs displaying the overall data distribution of the first variant ANN model with architecture 10–8–10 during training, testing, and validation stages are shown in Fig. 2, a. Correspondingly, Fig. 2, b illustrates the graphs of the model for the second variant of ANN construction with architecture 13-8-10.

Data on water quality and physico-chemical parameters of the river were divided into training, validation, and test sets. Training data were used to train the network, validation for tuning, and test data for evaluating the model's accuracy. The selected ratio was 70%:15%:15%.

The selection of the optimal number of neurons in

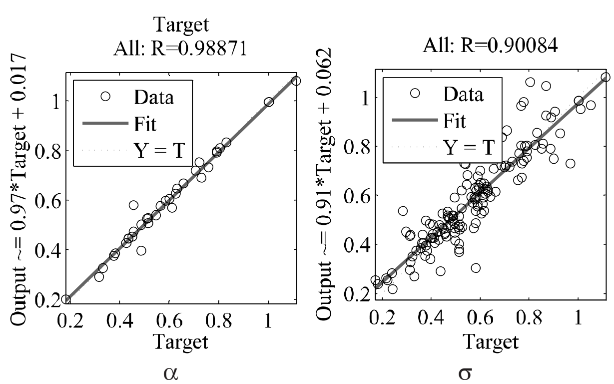

Fig. 2. Regression Plots: a - Artificial Neural Network with architecture 10-8-10; b - Artificial Neural Network with architecture 13-8-10

the hidden layer was carried out to achieve the best accuracy. Several tests with different numbers of neurons were conducted, choosing the option that showed the best results.

Coefficient of determination (R2) and mean squared error (MSE) were used to assess accuracy. The optimal model was chosen based on these metrics and tested on the test dataset.

To assess the adequacy of the models, we analyzed their convergence on separate sections according to known data by comparing actual and predicted values determined using the ANN method. The adequacy of the models was assessed by evaluating convergence on sections with actual data as well as at points of maximum and minimum. Deviations between actual data and predicted values generated using the ANN method were analyzed and compared in Table 3.

Table 3 demonstrates that the first variant of ANN performs better. The highest error within the average data range (trial 2) calculated by ANN amounts to 1.41%; the lowest error is -1.09%. However, during sharp changes in conditions, such as minimal decrease in parameter concentration (trial 1) or sharp increase (trial 3), the first ANN displays significant error. On average, this error is mitigated as abnormal parameter changes are less frequent within the input data range.

Assessing the second variant of ANN constructed based on water quality parameters from preceding points, a considerably higher prediction error is observed within the average data range (1.74%). Nonetheless, it's worth noting that the error size remains almost unchanged when predicting data model variations (0.76%).

Table 3 provides a comparative analysis of the results between two model types: the traditional model and the model accounting for hydrochemical regime. Various chemical indicators of the aquatic environment measured in three trials are used for analysis.

One of the indicators is NH4+, which reflects the ammonia concentration in water. Compared to known data, both models exhibit absolute and relative deviations. For this indicator, it is noteworthy that the traditional model demonstrates less relative deviation in trial 2 than the model considering the hydrochemical regime (-1.81% versus 6.30%, respectively).

Another indicator, BOD5, reflecting the amount of organic substances in water, is also analyzed. In this case, the model considering the hydrochemical regime shows better results, reducing the relative deviation in trial 2 by 6.25% compared to the traditional model, which has a relative deviation of -27.05%.

Additionally, the dissolved oxygen level (O2), indicating the oxygen regime of the aquatic environment, is analyzed. In this case, the model considering the hydrochemical regime proves to be more accurate again, showing less relative deviation in all three trials compared to the traditional model.

Based on the assessment of the determination coefficient error and mean square error, it can be concluded that the proposed ANN models are adequate. Therefore, it can be said that the model with the architecture 13-8-10 demonstrates good predictive characteristics in the occurrence of new local sources of pollution, while the ANN model with the architecture 10-8-10 demonstrates extremely high accuracy under stable conditions.

Discussion

One of the key findings of the analysis is the identification of a strong positive correlation between nitrate-ion and phosphate-ion (polyphosphates) concentrations. The correlation coefficient value of 0.57 indicates that with an increase in nitrate concentrations, the concentration of phosphates is likely to increase, and vice versa. This strong positive correlation suggests a possible interaction between these substances in the aquatic ecosystem, possibly due to a common source of origin, such as the formation of nitrogen-phosphorus compounds.

A moderate positive correlation was also found between ammonium-ion concentration and BOD5, although it is less pronounced with a correlation coefficient of 0.11. This correlation may indicate the potential influence of ammonium ions on the activity of bacteria and other microorganisms that utilize oxygen for their life processes.

On the other hand, a weak negative correlation between suspended solids and dissolved oxygen (correlation coefficient -0.21) may indicate that an increase in the concentration of one parameter leads to a decrease in the other. This could be related to physicochemical processes in the water, where an elevated concentration of suspended solids may affect oxygen solubility.

An important aspect is the absence of a significant correlation between dissolved oxygen concen-

Table 3. Deviation of data predicted by the model from actual data

Indicator	Model	Traditional Model			Model Incorporating Hydrochemical Regime		
	Trial	1	2	3	1	2	3
	Model Errors						
R		0,98871			0,90084		
MSE		0,0159			0,0215		
Gradient error		0,0009			1.09-11		
Mu		1-5			1-8		
	KD*	0,11	1,60	3,30	0,11	1,60	3,30
Y1 - NH_4^+	MD**	0,12	1,63	3,63	0,12	1,50	3,53
	?Error***	-0,01	-0,03	-0,33	-0,01	0,10	-0,23
	RAError****	-12,83	-1,81	-10,07	-6,60	6,30	-6,90
Y2 - BOD5	KD	5,00	5,00	5,00	5,00	5,00	5,00
	MD	6,35	4,92	4,70	5,31	4,59	5,02
	?Error	-1,35	0,08	0,30	-0,31	0,41	-0,02
	RAError	-27,05	1,61	6,02	-6,25	8,25	-0,40
Y3 - Suspended solids	KD	6,00	10,00	9,00	6,00	10,00	9,00
	MD	6,09	10,13	9,74	6,76	11,12	10,15
	AError	-0,09	-0,13	-0,74	-0,76	-1,12	-1,15
	RAError	-1,45	-1,27	-8,27	-12,67	-11,24	-12,83
Y4 - O_2	KD	9,30	7,10	6,20	9,30	7,10	6,20
	MD	10,26	7,20	6,60	10,80	8,00	7,14
	?Error	-0,96	-0,10	-0,40	-1,50	-0,90	-0,94
	RAError	-10,27	-1,41	-6,39	-16,14	-12,64	-15,22
Y5 - NO_3^-							
KD	1,01	4,20	3,60	1,01	4,20	3,60	
	MD	0,74	4,25	3,90	1,15	4,54	3,86
	?Error	0,27	-0,05	-0,30	-0,14	-0,34	-0,26
	RAError	26,35	-1,09	-8,47	-14,07	-8,15	-7,13
Y6 - NO_2^-							
KD	0,11	0,09	0,12	0,11	0,09	0,12	
	MD	0,11	0,09	0,11	0,12	0,10	0,13
	?Error	0,00	0,00	0,01	-0,01	-0,01	-0,01
	RAError	-3,85	-1,25	4,37	-7,47	-7,87	-6,88
Y7 - SO_4^{2-}							
KD	34,90	34,90	0,00	34,90	34,90	0,00	
	MD	49,91	35,54	0,05	37,74	39,10	0,00
	?Error	-15,01	-0,64	-0,05	-2,84	-4,20	0,00
	RAError	-43,02	-1,83	-100,00	-8,15	-12,03	0,00
Y8 - PO_4^{3-}							
KD	0,19	1,32	3,05	0,19	1,32	3,05	
	MD	0,23	1,34	3,71	0,22	1,49	3,45
	?Error	-0,04	-0,02	-0,66	-0,03	-0,17	-0,40
	RAError	-20,39	-1,65	-21,49	-14,22	-12,88	-13,02
Y9 - Cl^-	KD	54,20	41,80	45,30	54,20	41,80	45,30
	MD	49,41	42,25	44,78	58,45	45,92	48,57
	?Error	4,79	-0,45	0,52	-4,25	-4,12	-3,27
	RAError	8,83	-1,07	1,15	-7,85	-9,85	-7,22
Y10 - *WPI*							
KD	0,60	0,13	1,20	0,60	0,13	1,20	
	MD	0,61	0,13	1,28	0,66	0,14	1,33
	?Error	-0,01	0,00	-0,08	-0,06	-0,01	-0,13
	RAError	-1,37	-1,35	-6,46	-9,48	-7,85	-11,11

*Known Data; **Model Data; ***Absolute Error; ****Relative Absolute Error,%

tration and other parameters, except for a weak negative correlation with BOD5. This may indicate that although dissolved oxygen is an important factor for the existence of aquatic organisms, its concentration is less sensitive to changes than other chemical parameters.

Overall, the obtained results of the water quality criteria's impact on the Southern Bug River indicate that there is a systematic increase in suspended solids concentration in the water from the river's sources downstream. This leads to an increase in the BOD5 indicator and a decrease in dissolved oxygen levels.

It has been established that various sources of pollution contribute to suspended solids, including liquid waste, discharges, and livestock waste. Additionally, unauthorized discharges from enterprises and domestic sewage have been identified. These discharges may include chemical substances and various polluting components, exacerbating water quality issues.

The increase in suspended solids leads to a decrease in dissolved oxygen in the water and an increase in the BOD5 indicator.

Analyzing water quality parameters, particularly BOD5, provides an overview of the overall state of water resources. A significant increase in the BOD5 indicator indicates an increase in water pollution and suggests the presence of organic substances subject to biological decomposition. This phenomenon may be associated with increased discharges of untreated wastewater or excessive discharge of organic pollutants into the water body.

Regarding the BOD5 indicator, its increase can serve as an indicator of the intensity of the biological process of water pollution. High values of the indicator indicate intensive decomposition of organic substances in water, which may be caused by high levels of bacterial activity, leading to oxygen consumption by water. In turn, this phenomenon results in ecosystems downstream facing the problem of insufficient oxygen to support aquatic life.

The hypothesis regarding the effectiveness of constructing a prediction model considering the hydrochemical regime of the river has been practically verified in the study.

Two ANN models have been constructed: using the traditional method based on monitoring point data, which is evaluated, and using the method considering the hydrochemical regime of the river with data from previous points.The results of comparing two types of models, traditional and hydrochemical regime-aware models, are presented for ten indicators (Y1-Y10), which include various chemical substances in rivers (Table 1). For each indicator, both known and model-predicted data are provided, along with their absolute and relative deviations. The analysis shows that the effectiveness of the models varies depending on the specific indicator; however, generally, the traditional model often demonstrates better accuracy, although the hydrochemical regime-aware model may be more accurate in certain cases.

It should be noted that despite the first model impressing with its high accuracy in predicting the water quality of the Southern Bug River, the second model has practical advantages. The second model, although less accurate with an error of 0.93, appears less accurate only in terms of overall error. It is noteworthy that the second model considers a greater number of parameters, including water quality parameters at previous monitoring points from the estuary. This gives the model a unique ability to more accurately predict sharp changes in water quality, such as the entry of chemical substances or other pollutants (Table 3). This capability of the model finds practical application in situations where it is important not only to determine the overall water quality but also to accurately predict specific changes for effective water quality management and environmental conservation.

Therefore, despite being somewhat less accurate (by 7%) in overall forecasts, the second model proves to be more practical in real conditions, providing practical application for the effective detection and prediction of specific water pollutants.

The advantages of the presented study over studies (Lu *et al.*, 2023; Yang *et al.*, 2021; Kim *et al.*, 2017) lie in its comprehensive approach to analyzing water quality in the Southern Bug River. It includes water quality analysis at various monitoring points, allowing for the identification of different sources of pollution in the river and establishing correlations between water quality parameters. Such an approach allows for a more complete picture of the state of water resources.

In the proposed study, predictive models of water quality considering the hydrochemical regime of the river have been constructed for different monitoring points. This allows for water quality forecasts to be made considering changes in hydrological and hydrochemical conditions of the river, which is important for effective water resource management. This indicates that the model considering the hydrochemical regime is more accurate and reliable for predicting water quality.

This approach allows for more accurate detection and understanding of variations in water quality, which is crucial for effective water resource management. Additionally, the application of machine learning methods in the study allows for the generation of necessary data for assessing and predicting water quality, avoiding limitations associated with data collection. Such an approach leads to a more comprehensive and reliable analysis of the state of water resources, thereby contributing to making more informed decisions in water system management.

Compared to studies focusing on the WQI methodology (Zhu *et al.*, 2019; Abba *et al.*, 2020; Kim*at al.*, 2015; Asadollah *et al.*, 2021; Kouadri *et al.*, 2021), the proposed study addresses some drawbacks inherent in other studies in this field.

Firstly, compared to the article (Zhu *et al.*, 2019), which uses an extreme learning machine-based model to predict water temperature, the study on the Southern Bug River considers the analysis of various water quality parameters at different monitoring points, providing a more comprehensive understanding of pollution.

Studies (Abba *et al.*, 2020; Kim *et al.*, 2015) use different models to predict the water quality index in the river basin; however, as noted, these studies have limitations regarding the use of only one data source and the lack of analysis of the impact of anthropogenic activities on water quality. In contrast, the proposed study analyzes water quality parameters considering various pollution sources in the river, providing a more objective understanding of the situation and the stability of the methodology's results under different conditions. In comparison, the study on the Southern Bug River is conducted on a specific water body considering various pollution sources, providing a more detailed understanding of the situation and the effectiveness of control measures.

The limitations of the study are its territorial attachment. This means that since the seasonal dynamics of the hydrochemical regime of rivers vary, the model can only be applied to the Southern Bug River. For a general case, it is advisable to apply only the ANN modeling method. Addressing this limitation involves conducting a large-scale analysis of various rivers. However, in this case, the forecasting efficiency will suffer.The study's limitation lies in the insufficient consideration of the soil influence factor. Since the obtained data fall within the distribution range of two types of soil - chernozem and gray forest soils - which during the period of maximum moisture saturation in the watershed, the input of a small amount of salts is relatively insignificant. This means that they do not have a significant impact on the hydrochemical regime of the river in this area.

Therefore, there is a need to expand the study to understand how infiltration waters entering the river network through heavy clayey and medium-clayey saline chernozems affect it.

The proposed methodology for studying the Southern Bug River and its aquatic ecosystems proves to be superior due to several features. Firstly, considering the influence of water parameters at previous monitoring points allows predicting the water pollution level along the river under the influence of any anthropogenic factor.

For example, if there is a random discharge of pollutants due to a technological accident or waste mismanagement at the upper monitoring point, this impact can be accounted for using the proposed methodology. This will enable a more accurate assessment of water quality at downstream river sections and, consequently, provide the opportunity to react promptly to potential threats to aquatic ecosystems and public health.

Secondly, the spatial structure of the input-output model allows analyzing the river's flow and predicting possible changes in pollution levels in case of pollutant entry. For instance, if there is an industrial zone along a certain stretch of the river, the methodology can help determine the impact of discharges from this zone on water quality downstream.

Thus, the ability to predict the river's pollution level depending on anthropogenic interventions becomes a crucial aspect, giving this methodology a specific advantage over traditional approaches.

Conclusion

The study of the influence of physico-chemical water quality criteria on the Southern Bug River has yielded significant results, reflecting key parameters of the aquatic environment. Examination of the physico-chemical water quality criteria of the Southern Bug River has revealed substantial differences in the levels of individual indicators between various monitoring points. These differences indicate diverse influences of anthropogenic and natural factors in different areas of the basin. For example, monitoring points in Khmilnyk and Hushchynets demonstrate higher average levels of ammonium ions, BOD5, and nitrates compared to other points.

Analysis of the concentrations of various substances in the water of the Southern Bug River at different monitoring points has shown significant variability, which may be associated with different sources of pollution and spatial unevenness in their distribution. In particular, the oxygen-containing nitrates and phosphates have the most significant impact on prediction results.

Correlation analysis has revealed statistical relationships between various chemical parameters in the water. For instance, a weak positive correlation between nitrate ion content and BOD5 has been identified, indicating a possible increase in biochemical oxygen demand with an increase in nitrate ion content in the water.

The obtained results allow understanding the complex influence of various factors on the water quality of the Southern Bug River and indicate the need for a systematic approach to water resource management, considering their geographical location and pollution specifics.

The main quantitative parameters of the Southern Bug River at different monitoring points have been investigated. For example, the average concentration of ammonium ions ranges from 0.42 to 0.72 mg/dm^3, with the highest level recorded at the Khmilnyk point and the lowest at the Mankivka point. As for BOD5, it varies from 5.25 to 7.00 mgO/dm^3, with the highest value at the Khmilnyk point. Additionally, concentrations of other substances such as nitrate ions, nitrite ions, sulfate ions, phosphate ions, and chloride ions also differed between different monitoring points, allowing for an accurate assessment of water quality and identification of potential pollution problems. Considering these criteria, effective management of water resources can be achieved, and measures for the preservation and restoration of riverine ecosystems can be developed.

Further examination of physico-chemical water parameters has also allowed understanding the relationship between different factors and their impact on water quality. For instance, it has been found that an increase in water temperature may lead to a decrease in dissolved oxygen concentration, negatively impacting the aquatic ecosystem and biodiversity. These findings underscore the importance of continuous monitoring of water quality parameters and the development of strategies for their improvement.

It is also necessary to note that knowledge of the physico-chemical water criteria of the Southern Bug River is crucial for making decisions regarding the conservation of water resources and ensuring their sustainable use. The research results can be used in developing strategies for environmental protection and reducing the impact of anthropogenic factors on aquatic ecosystems. Such an approach will help ensure ecological resilience and preserve the biodiversity of water resources of the Southern Bug River for the future.

A water quality prediction model has been developed, considering the hydrochemical regime of the river, yielding important results that open up new perspectives in the study and management of water resources. In particular, the developed model allows for accurate prediction of water pollution levels in the Southern Bug River, taking into account hydrochemical parameters.

The obtained results demonstrate the high effectiveness of the model considering the hydrochemical regime in predicting water quality. Comparative assessments indicate that this model exhibits an accuracy level of 70-75%, providing sufficient information for management decisions. However, it is worth noting that during sharp anthropogenic changes in water resources, the accuracy of this model increases to 80-85%, which is highly significant in the context of prompt response to negative impacts.

These results indicate the advantage of using the model considering the hydrochemical regime compared to traditional prediction methods. They also demonstrate the advantage of employing a comprehensive approach, which considers not only the physico-chemical indicators of water at a single monitoring point but also data from other river segments. Consideration of hydrochemical parameters allows for a more precise assessment of water resource status and contributes to the development of effective strategies for managing aquatic ecosystems.

Interpretation of the obtained results lies in the inclusion of additional water quality parameters from other monitoring points, which enhances the predictive characteristics of the model and enables more accurate forecasting of river water quality.

Quantitative evaluations of the result show that the second approach provides less deviation of predicted data from actual data compared to the traditional approach. For example, the average relative error of the second approach is 1.74%, which is lower than the error of the traditional approach in calculating relative deviations within the average data range (27.05%). Such comparative assessments confirm the effectiveness of the second approach in modeling water quality predictions considering the hydrochemical regime.

Thus, the developed prediction model is an important tool for ensuring sustainable use of water resources of the Southern Bug River and maintaining ecological balance in this region.

Conflict of interest

The authors declare no conflict of interest.

References

Abba, S.I., Hadi, S.J., Sammen, S.S., Salih, S.Q., Abdulkadir, R.A., Pham, Q.B. and Yaseen, Z.M. 2020. Evolutionary computational intelligence algorithm coupled with self-tuning predictive model for water quality index determination. *Journal of Hydrology*. 587: 124974. https://doi:10.1016/j.jhydrol.2020.124974

Asadollah, S.B.H.S., Sharafati, A., Motta, D. and Yaseen, Z.M. 2021. River water quality index prediction and uncertainty analysis: A comparative study of machine learning models. *Journal of Environmental Chemical Engineering*. 9(1): 104599. https://doi.org/10.1016/j.jece.2020.104599

Hossain, M. and Patra, P.K. 2020. Water pollution index-A new integrated approach to rank water quality. *Ecological Indicators*. 117: 106668. https://doi.org/10.1016/j.ecolind.2020.106668

Kim, S.E. and Seo, I.W. 2015. Artificial Neural Network ensemble modeling with conjunctive data clustering for water quality prediction in rivers. *Journal of Hydro-Environment Research*. 9(3): 325-339. https://doi.org/10.1016/j.jher.2014.09.006

Kim, S.E., Seo, I.W. and Choi, S.Y. 2017. Assessment of water quality variation of a monitoring network using exploratory factor analysis and empirical orthogonal function. *Environmental Modelling & Software*. 94: 21-35. https://doi.org/10.1016/j.envsoft.2017.03.035

Kouadri, S., Elbeltagi, A. and Islam, A.R.M.T. 2021. Performance of machine learning methods in predicting water quality index based on irregular data set: application on Illizi region (Algerian southeast). *Applied Water Science*. 11: 190 https://doi.org/10.1007/s13201-021-01528-9

Lu, J., Jiarong G., Jinyang H., Jun X., Yi L., Gengmin J. and Yifeng Zh, 2023. Evaluation of Spatiotemporal Patterns and Water Quality Conditions Using Multivariate Statistical Analysis in the Yangtze River, China. *Water*. 15(18): 3242. https://doi.org/10.3390/w15183242

Reid, A.J., Carlson, A.K., Creed, I.F., Eliason, E.J., Gell, P.A., Johnson, P.T.J. and Cooke, S. J. 2018. Emerging threats and persistent conservation challenges for freshwater biodiversity. *Biological Reviews*. https://doi.org/10.1111/brv.12480

Tiyasha, Minh Tung, T. and Mundher Yaseen, Z. 2020. A survey on river water quality modelling using artificial intelligence models: 2000-2020. *Journal of Hydrology*. 124670. https://doi.org/10.1016/j.jhydrol.2020.124670

Yan, X., Zhang, T., Du, W., Meng, Q., Xu, X. and Zhao, X. 2024. A Comprehensive Review of Machine Learning for Water Quality Prediction over the Past Five Years. *Journal of Marine Science and Engineering*. 12(1): 159. https://doi.org/10.3390/jmse12010159

Yang, S., Liang, M. and Qin, Z. 2021. A novel assessment considering spatial and temporal variations of water quality to identify pollution sources in urban rivers. *Scientific Reports*. 11: 8714. https://doi.org/10.1038/s41598-021-87671-4

Zhu, S., Heddam, S., Wu, S., Dai, J. and Jia, B. 2019. Extreme learning machine-based prediction of daily water temperature for rivers. *Environmental Earth Sciences*. 78(6). https://link.springer.com/article/10.1007/s12665-019-8202-7