

Forecasting Water Pollution Indicators Using the STL-ARIMA Model: A Case Study of the Dniester River

Marusic Galina¹, Braniște Rodica^{2*}, Kulev Mihail³, Istrati Daniela⁴, Șevcenco Nina⁵

^{1,2,3,4} Department of Computer Science, Technical University of Moldova, Chisinau, Moldova

⁵ Department of Pediatric Dentistry, Testemitanu State University of Medicine and Pharmacy, Chisinau, Moldova

*Corresponding Author Email: rodica.braniste@iis.utm.md

Abstract

Regular monitoring and forecasting of water quality indicators are important for a proper management of aquatic ecosystems and for the protection of public health. This article analyzes the temporal evolution of two important indicators of nitrogen pollution: nitrite nitrogen and nitrate nitrogen. This analysis is based on data collected during 2019–2024 in the Olănești locality, on the Dniester River, Republic of Moldova.

The time series were processed using a hybrid STL-ARIMA approach. This approach combines the Seasonal-Trend decomposition using Loess (STL) with the AutoRegressive Integrated Moving Average (ARIMA) model. The STL method separated the series into trend, seasonal and residual components. It provides a clearer picture of the evolution of each indicator. After the separation of series, the ARIMA models were used. These models were applied to forecast future concentrations for the period 2025–2027. The results of the models were verified by residual analysis and by using forecast accuracy indicators.

The results indicated that both indicators will have seasonal variations, and each indicator differs. Nitrate nitrogen has more pronounced fluctuations than nitrite nitrogen. The STL-ARIMA method captured the main characteristics of the analyzed series, and the developed forecasts do not indicate a pronounced long-term increase in concentrations during the period 2025–2027.

The obtained results indicate the usefulness of the proposed methodology for monitoring water quality. Forecasting is the method that can be useful to identify in advance the change of the pollution level.

Keywords

ARIMA model, Dniester River, time series forecasting, water pollution.

INTRODUCTION

Water pollution is considered to be one of the biggest global environmental problems. It can affect aquatic and terrestrial ecosystems, as well as biodiversity, and human health. Rivers are important routs for transporting nutrients and organic matter. The rivers are very sensitive to human activities such as agricultural runoff, urban wastewater, and industrial effluents that effect water quality negatively in many regions. This situation shows that it is necessary to monitor water quality continuously and to have reliable models that can predict key pollution indicators. [1], [2].

The Dniester River is one of the most important freshwater resources in Eastern Europe. It supplies water for domestic use, agriculture, and is important to the local ecosystems. Earlier research conducted in the Dniester River basin showed the presence of different chemical pollutants and the need for regular monitoring and assessment of water quality [3], [4]. In this case, nitrite nitrogen (NO₂-N) and nitrate nitrogen (NO₃-N) are widely used as the main indicators for deterioration of water quality.

Nitrite and nitrate compounds have important roles in the nitrogen cycle of aquatic systems. High nitrate concentrations indicate that they are often associated with agricultural fertilizer runoff. Nitrite accumulation indicates incomplete nitrification processes in polluted waters.

Excessive consumption of nitrates can pose a health risk. Exposure to nitrates can cause methemoglobinemia in infants. It may also contribute to long-term health risks through the formation of carcinogenic compounds in drinking water [5], [6]. High nitrate loads can also contribute to eutrophication processes. This process may cause algal blooms and oxygen depletion in aquatic ecosystems. High nitrite concentrations may pose additional risks to human health and aquatic ecosystems due to the toxicity of nitrites and their involvement in nitrogen transformation processes.

Water quality time series have a complex and often nonlinear evolution. These characteristics make it difficult to predict future values. For this reason, are necessary researchers of statistical and computational models for the analysis and forecasting of these series. The AutoRegressive (ARIMA) is one of the most widely used models. This model was introduced by Box and Jenkins. It is one of the most widely used stochastic approaches because [7]. ARIMA models have been applied in river water quality analysis, including nutrient dynamics and water quality forecasting in various hydrological systems [8].

However, environmental time series often show strong seasonal patterns and long-term trends. These characteristics can influence the stationarity needed by traditional ARIMA models. The Seasonal-Trend Decomposition using Loess (STL) method helps manage this limitation. It separates the

time series into trend, seasonal, and residual components. STL decomposition is well suited to environmental data. It captures both regular seasonal patterns and irregular variations. STL decomposition is highly effective for environmental data. It records both typical seasonal trends and unexpected fluctuations [9], [10].

Recent studies have increasingly applied hybrid and decomposition-based approaches to water quality forecasting. ARIMA-based and hybrid models have been used to forecast different water quality variables, including dissolved oxygen, water temperature, and boron, using monthly monitoring data. Other studies have incorporated data decomposition and error-correction procedures to improve water quality forecasting. Recent research has also explored hybrid frameworks that combine temporal decomposition, predictor selection, and machine-learning methods for long-term prediction of water quality indicators [11], [12], [13], [14]. These studies indicate the importance of accounting for temporal structure, seasonality, and nonlinear variation when forecasting environmental variables. However, applications specifically focused on monthly nitrate nitrogen and nitrite nitrogen forecasting in the Dniester River remain limited. The present study addresses this gap by applying a STL-ARIMA approach to monthly observations collected in the Olănești locality during 2019–2024 and by developing forecasts for 2025–2027.

Nitrogen compounds are important environmental and public health pollutants. In order to develop early warning systems and sustainable water management strategies their temporal evolution can be forecasted with sufficient accuracy [5], [6]. For this reason, in the present study was used a hybrid STL-ARIMA approach in order to forecast the content of nitrite nitrogen and nitrate nitrogen in the water samples collected on a monthly basis in the Olănești locality on the Dniester River for the period 2019-2024.

The main objective of this work is to assess the ability of the proposed approach, to capture both seasonal and long-term evolutions of indicators of water pollution.

METHODOLOGY

This study analyzes the temporal behavior and forecasts some water quality indicators in the Olănești locality on the Dniester River, Republic of Moldova. This analysis focuses on two indicators of water quality: nitrite nitrogen ($\text{NO}_2\text{-N}$) and nitrate nitrogen ($\text{NO}_3\text{-N}$). These two indicators are of greater importance as they indicate the nutrient pollution and affect river ecosystems due agricultural use of land, human activity, wastewater discharges on river ecosystems. High levels of nitrogen compounds can cause eutrophication processes, pollute the water, and cause significant harm to aquatic life and human health.

In order to analyze the selected indicators over time, it was chosen a time series forecasting approach. This approach is based on combination of the method of Seasonal-Trend decomposition using Loess (STL) and the Autoregressive Integrated Moving Average (ARIMA) model

To analyze the temporal evolution of these indicators, the study uses the STL-ARIMA approach, structured in three stages: data preparation, STL decomposition, and ARIMA forecasting. The proposed methodology analyzes and forecasts the water quality indicators and consists of three main stages: data preparation, STL decomposition, and ARIMA-based forecasting.

Data

The data from this research paper was provided by the monthly reports on the environmental monitoring activities of the Environmental Agency of the Republic of Moldova [10]. The reports contain environmental data that were recorded and updated on a monthly basis from January 2019 until December 2024. The paper analyzes two water quality indicators that exceeded the maximum admissible concentration (MAC). The MAC values for Class I water quality were considered in accordance with the Regulation on Environmental Quality Requirements for Surface Waters [15], namely:

1. Nitrite nitrogen ($\text{NO}_2\text{-N}$), with a MAC value of 0.01 mg/L;
2. Nitrate nitrogen ($\text{NO}_3\text{-N}$), with a MAC value of 1.0 mg/L.

Nutrient indicators are measured in aquatic ecosystems to assess nutrient pollution. Indicators are influenced by agricultural run-off, wastewater discharges and other human activities in a river basin.

There are monthly time series for each of the pollutants that can be used for building the statistical model and making a forecast. Each pollutant has a set of monthly values for the respective concentration other time. The temporal structure shown is enough to generate time series and to monitor the concentration of polluting substances in water over time.

A summary of the main statistics for the indicators included in this study are shown in Table 1.

Seasonal-Trend Decomposition Using Loess (STL)

Water quality indicators frequently exhibit long-term trends and seasonal fluctuations generated by climatic conditions, hydrological processes, and anthropogenic activities. In order to separate these components and improve forecasting performance, the Seasonal-Trend decomposition using Loess (STL) method proposed by [10].

STL decomposes a time series into three additive components:

$$Y(t) = T(t) + S(t) + R(t) \quad (1)$$

where $Y(t)$ is the observed value at time t , $T(t)$ is the trend component representing the long-term evolution of the series, $S(t)$ is the seasonal component describing periodic variations, and $R(t)$ is the residual component containing random fluctuations and unexplained variability.

STL uses LOESS smoothing to estimate local trends and is suitable for environmental time series that exhibit nonlinear and seasonal behavior.

The decomposition helps to better understand the temporal behavior of nitrite nitrogen (NO_2-N) and nitrate nitrogen (NO_3-N), while improving the quality of subsequent forecast models.

ARIMA Forecasting Model

The time series are used for forecasting within the context of the ARIMA model that has been presented in [6]. The ARIMA model is the most common models for the modeling and forecasting of univariate time series and it has been applied in many environmental and water quality studies.

The ARIMA model is often expressed by the parameters of the model ARIMA (p, d, q) where p is the order of the autoregressive part of the model (AR), d is the number of differences that need to be taken of a time series so that it becomes stationery and q is the order of the moving average part (MA).

The general form of the ARIMA model can be expressed as:

$$Y_t = c + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \dots + \varphi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} \quad (2)$$

where Y_t is the value of the time series at time t; c is a constant term; $\varphi_1, \varphi_2, \dots, \varphi_p$ are autoregressive coefficients; $\theta_1, \theta_2, \dots, \theta_q$ are moving-average coefficients; ε_t represents the random error term.

The STL-ARIMA Approach

It was implemented in two sequential steps, using programming language R. First of all, each monthly time series were decomposed using STL with a periodic seasonal component (s.window = "periodic"). The seasonal component was then removed using the seasadj() procedure, resulting in a seasonally adjusted series. Secondly, an ARIMA model was fitted to the seasonally adjusted series using the auto.arima() procedure. The selected ARIMA model was subsequently used to generate forecasts for a 36-month period, covering the period January 2025 - December 2027. This procedure uses STL to identify and remove the seasonal component before applying ARIMA to the seasonally adjusted series.

ARIMA Model Selection

The ARIMA model orders were selected automatically using the auto.arima() procedure applied to the seasonally adjusted series. The procedure evaluates alternative ARIMA specifications based on information criteria. For the analyzed series, the resulting models were ARIMA(2,1,1) for NO_2-N and ARIMA(2,0,2) for NO_3-N . Model adequacy was further assessed using residual diagnostics, including the Ljung–Box test at a significance level of 0.05.

Model Training and Forecasting

The models were fitted using the complete monthly dataset covering January 2019 to December 2024. No separate testing subset was used. The fitted models were then used to generate forecasts for the 36 months from January 2025 to December 2027. The MAE, RMSE, and MAPE values

reported in Table 3 therefore represent in-sample model errors rather than independent out-of-sample forecast accuracy

Forecast Evaluation

The predictive performance of the STL-ARIMA models was evaluated using three commonly employed forecasting accuracy measures: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE).

The Mean Absolute Error is defined as:

$$MAE = (1/n) \sum |Y(t) - \hat{Y}(t)| \quad (3)$$

where $Y(t)$ is the observed value, $\hat{Y}(t)$ is the predicted value, and n is the number of observations.

The Root Mean Square Error is calculated as:

$$RMSE = \sqrt{(1/n) \sum (Y(t) - \hat{Y}(t))^2} \quad (4)$$

RMSE assigns greater weight to large forecasting errors and is therefore useful for identifying models with substantial deviations.

The Mean Absolute Percentage Error is given by:

$$MAPE = (100/n) \sum |Y(t) - \hat{Y}(t)| / Y(t) \quad (5)$$

MAPE measures forecasting errors in percentage terms and facilitates comparisons between variables.

Lower values of MAE, RMSE and MAPE mean a better forecasting performance. These metrics were used to check how accurate the STL-ARIMA models were in predicting nitrite nitrogen (NO_2-N) and nitrate nitrogen (NO_3-N) levels of concentration in the Dniester River.

The models were estimated using the complete monthly series covering January 2019–December 2024. The reported MAE, RMSE, and MAPE values represent training-set errors of the fitted models and were used to assess model fit. The final selected models were then used to generate 36-month forecasts for 2025–2027.

RESULTS AND DISCUSSION

Descriptive Statistics

Table 1 shows the statistics for the studied pollutants. Nitrate nitrogen had higher variation compared to nitrite nitrogen. Its concentration went from 0.005 mg/L up to 4.08 mg/L. The highest value found was than four times higher than the maximum admissible concentration of 1.0 mg/L. At the time nitrite nitrogen concentrations stayed, between 0 and 0.10 mg/L and even that exceeded the maximum admissible level. These results show substantial differences in the observed concentration ranges and variability of the two indicators and provide a basis for their further time-series analysis.

Table 1. Descriptive statistics of the analyzed water pollution indicators (2019–2024)

Statistic	NO ₃ -N (mg/L)	NO ₂ -N (mg/L)
Minimum	0.005	0.000
1st Quartile (Q1)	0.050	0.005
Median	0.057	0.0125
Mean	0.348	0.025
3rd Quartile (Q3)	0.310	0.043
Maximum	4.080	0.100
Standard Deviation	0.712	0.026
Variance	0.506	0.000686
MAC	1.000	0.010

Forecasting Results

In the figure 1 is shown the result for the nitrite nitrogen (NO₂-N) time series for the period 2025–2027. The historical series varies moderately with some local maxima and minima values. The STL-ARIMA model presents seasonal behavior and forecast for the same period of time 2025–2027.

The forecasted values remain relatively stable throughout the prediction horizon, oscillating around the historical level without displaying a pronounced upward or downward trend. The same fluctuations as seen in the historical data are also shown in the forecast, and the same seasonal effects will be expected.

The width of the confidence intervals increases for longer time horizons. In the same time the central forecast remains within a relatively narrow band around it and it indicates good and stable performance of the model.

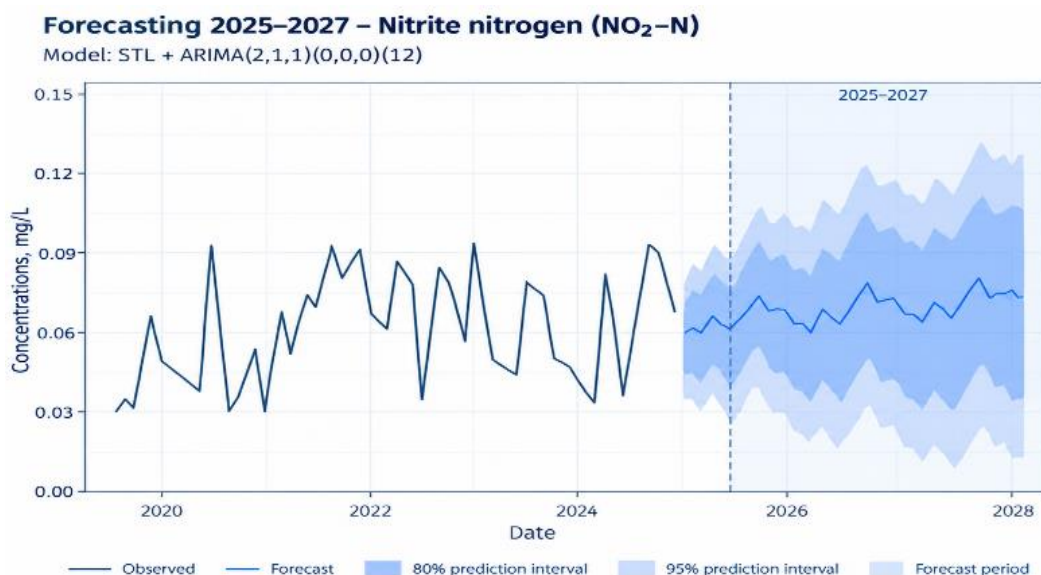


Figure 1. Forecast of Nitrite Nitrogen (NO₂-N) Concentrations for 2025–2027

The sentence is grammatically correct. The only real issues are stylistic: "increasing" is repeated, and "forecast" appears twice. "Nitrite nitrogen" is also usually hyphenated when used as a modifier or measure (nitrite-nitrogen, NO₂-N).

Although long-term forecast will increase the uncertainty and this is why occasional monitoring is required to be done.

Fig. 2 shows the results for Nitrate Nitrogen (NO₃-N):

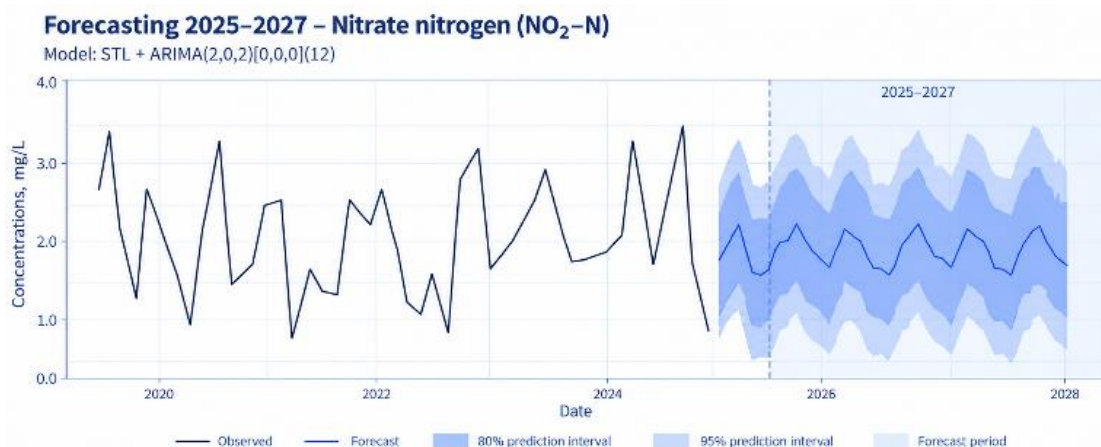


Figure 2. Forecast of Nitrate Nitrogen (NO₃-N) Concentrations for 2025–2027

The forecasted nitrate-nitrogen values fluctuate around a relatively constant level, with no pronounced long-term upward trend. These fluctuations may reflect seasonal and hydrological influences, although this study did not analyze those factors directly.

The STL-ARIMA forecast retains the seasonal component of the given time series and the values of the future time series are concentrated around mean value. This indicates the absence of a persistent trend. Overall, the forecast does not indicate a pronounced long-term increase in nitrate nitrogen concentrations during 2025–2027.

Similar to the above case the uncertainty of the forecast increases with time. The forecast does not indicate any increase of the nitrate nitrogen concentration in the future. The results indicate a relatively stable nutrient regime in the monitored section of the Dniester River. In the same time nutrient concentrations may show seasonal peaks. Environmental conditions and human activities can contribute to these short-term changes. However, fluctuations may continue during the forecast period.

Comparative Analysis of Nitrite and Nitrate Nitrogen Forecasts

The STL-ARIMA forecasts for nitrite-nitrogen (NO₂-N) and nitrate-nitrogen (NO₃-N) show different temporal patterns over 2025–2027. Nitrite-nitrogen fluctuates moderately, whereas nitrate-nitrogen varies more widely. Neither forecast indicates a pronounced long-term upward trend.

The STL-ARIMA forecasts for nitrite nitrogen (NO₂-N) and nitrate nitrogen (NO₃-N) show clear seasonal variations during the 2025–2027 forecasting period. The STL decomposition identified the main temporal characteristics of the historical series. The forecasting model used these characteristics to capture repeated fluctuations and estimate future values with good accuracy.

Both indicators have a seasonal variation, but their evolution over time is different. Nitrite nitrogen (NO₂-N) concentrations oscillate moderately around a relatively stable average value. This behavior indicates a more constant pollution level in the monitored sector of the Dniester River. In the case of nitrate nitrogen (NO₃-N), the values are more fluctuating. The oscillations are more pronounced, and the differences between the minimum and maximum values are greater, both in the analyzed period and in the forecast interval.

Overall, the results obtained do not indicate a clear long-term increasing trend in the concentrations of the two pollutants in the period 2025–2027. Therefore, it is expected that the level of pollution with nitrogen compounds in Olănești will generally remain within relatively stable limits. The results show that the STL-ARIMA framework is suitable for modeling the analyzed water quality time series and can support water quality monitoring and management. This approach can be used as a supporting tool in monitoring water quality and in the decision-making process regarding the management of water resources in the Dniester River basin.

Table 2. Comparative summary of STL-ARIMA forecasting results

Indicator	Seasonality	Variability	Trend (2025 –2027)	Interpretation
NO ₂ -N	Moderate	Moderate	Stable	No significant increase expected.
NO ₃ -N	Pronounced	Moderate–High	Stable	Seasonal fluctuations persist; no long-term increase projected.

Forecast Evaluation

The forecasting accuracy measures obtained for the selected STL-ARIMA models are presented in Table 3:

Table 3. Forecasting performance of the STL-ARIMA approach

Indicator	STL decomposition	ARIMA Model	MAE	RMSE	MAPE (%)
NO ₃ -N	STL, periodic seasonality	ARIMA (2,0,2)	0.4983	0.6938	43.38
NO ₂ -N	STL, periodic seasonality	ARIMA (2,1,1)	0.0186	0.0246	83.84

The ARIMA (2,0,2) model, applied to the nitrate nitrogen (NO₃-N) series, obtained an MAE of 0.4983 mg/L and an RMSE of 0.6938 mg/L. For nitrite nitrogen (NO₂-N), the ARIMA (2,1,1) model recorded lower error values, with an MAE of 0.0186 mg/L and an RMSE of 0.0246 mg/L. These results show a better approximation between the observed values and those estimated by the model for this indicator.

The relatively high MAPE calculated for NO₂-N should be interpreted with caution. The analyzed series includes values very close to zero, and in such situations the error expressed

in percentages can become artificially high. For this reason, MAE and RMSE are more suitable for assessing the model performance in the case of this indicator.

The adequacy of the models was also verified by the Ljung-Box test. The p-values obtained, 0.9009 for NO₃-N and 0.8599 for NO₂-N. These values clearly exceed the significance threshold of 0.05. This result suggests that the residuals no longer show significant autocorrelations and approach the behavior of a white noise. Therefore, the STL-ARIMA models were able to capture the main

characteristics of the temporal evolution of the two pollutants. The results support their use for forecasting future nitrogen concentrations in the analyzed sector of the Dniester River.

CONCLUSION

This paper applied a hybrid STL-ARIMA approach to model and forecast the concentrations of nitrite nitrogen ($\text{NO}_2\text{-N}$) and nitrate nitrogen ($\text{NO}_3\text{-N}$) in Olănești locality on the Dniester River. The STL method was used to decompose the time series into trend, seasonal, and remainder components, after which the seasonally adjusted series were modeled using ARIMA.

The STL decomposition identified seasonal patterns in both nitrogen indicators. Nitrate nitrogen showed greater variability than nitrite nitrogen over the analyzed period. The selected ARIMA models provided an adequate fit to the analyzed series, while the residual diagnostics indicated acceptable model adequacy.

The forecasts for 2025–2027 do not indicate a pronounced long-term increase in the concentrations of either nitrogen indicator. The forecast values remain relatively stable over the prediction horizon, although seasonal fluctuations are still observed. These results indicate that the STL-ARIMA approach can be useful for analyzing and forecasting water-quality indicators and may support environmental monitoring.

REFERENCES

- [1] F. N. Chaudhry, M. F. Malik, "Factors Affecting Water Pollution: A review," *Journal of Ecosystem & Ecography*, vol. 7, no. 1, 1000225, 2017. DOI: <https://doi.org/10.4172/2157-7625.1000225>
- [2] Ludwig W., Dumont E., Meybeck M., Serge Heussner, River discharges of water and nutrients to the Mediterranean and Black Sea: Major drivers for ecosystem changes during past and future decades?, *Progress in Oceanography*, Volume 80, Issues 3–4, 2009, pp. 199–217, ISSN 0079-6611, DOI: <https://doi.org/10.1016/j.pocean.2009.02.001>
- [3] Diamanti, K.S., Alygizakis, N.A., Nika, MC. et al. Assessment of the chemical pollution status of the Dniester River Basin by wide-scope target and suspect screening using mass spectrometric techniques. *Anal Bioanal Chem* 412, 4893–4907 (2020). DOI: <https://doi.org/10.1007/s00216-020-02648-y>.
- [4] R. Branîște, Numerical modeling of hydrodynamics of river-type aquatic systems. *Journal Of Engineering Science*, 33(1), 120–132, 2026. DOI: [https://doi.org/10.52326/jes.utm.2026.33\(1\).08](https://doi.org/10.52326/jes.utm.2026.33(1).08)
- [5] World Health Organization, *Guidelines for Drinking-water Quality: Fourth Edition Incorporating the First and Second Addenda*. Geneva, Switzerland: WHO, 2022.
- [6] World Health Organization, "Nitrate/Nitrite Fact Sheet," WHO Guidelines for Drinking-water Quality, Geneva, Switzerland, 2022.
- [7] Box G. E. P., Jenkins G. M., Reinsel G. C., and Ljung G. M., *Time Series Analysis: Forecasting and Control*, 5th ed. Wiley, 2015. DOI: <https://doi.org/10.1111/jtsa.12194>
- [8] Durdu Ö., F., "A hybrid neural network and ARIMA model for water quality time series prediction," *Engineering Applications of Artificial Intelligence*, vol. 23, no. 4, pp. 586–594, 2010. ISSN 0952-1976, DOI: <https://doi.org/10.1016/j.engappai.2009.09.015>
- [9] Hyndman, R. J., & Athanasopoulos, G. (2014). *Forecasting: principles and practice* (Vol. 2). Melbourne, Australia: OTexts.
- [10] R. B. Cleveland, W. S. Cleveland, J. E. McRae, and I. Terpenning, "STL: A Seasonal-Trend Decomposition Procedure Based on Loess," *Journal of Official Statistics*, vol. 6, no. 1, pp. 3–73, 1990.
- [11] J. Su, Z. Lin, F. Xu, G. Fathi, K. A. Alnowibet, et al., "A hybrid model of ARIMA and MLP with a Grasshopper optimization algorithm for time series forecasting of water quality," *Scientific Reports*, vol. 14, Art. no. 23927, 2024. DOI: [10.1038/s41598-024-74144-7](https://doi.org/10.1038/s41598-024-74144-7).
- [12] J. Kim, J. Yu, C. Kang, G. Ryang, Y. Wei, and X. Wang, "A novel hybrid water quality forecast model based on real-time data decomposition and error correction," *Process Safety and Environmental Protection*, vol. 162, pp. 553–565, 2022. DOI: [10.1016/j.psep.2022.04.020](https://doi.org/10.1016/j.psep.2022.04.020).
- [13] Zhang et al. A new framework for water quality forecasting coupling causal inference, time-frequency analysis and uncertainty quantification," *Journal of Environmental Management*, vol. 350, Art. no. 119613, 2024. DOI: [10.1016/j.jenvman.2023.119613](https://doi.org/10.1016/j.jenvman.2023.119613).
- [14] J. Han and L. Wang, "A novel hybrid model for long-term water quality prediction with the 'decomposition–inputs–prediction' hierarchical optimization framework," *Journal of Hydroinformatics*, vol. 26, no. 11, pp. 3008–3026, 2024. DOI: [10.2166/hydro.2024.244](https://doi.org/10.2166/hydro.2024.244).
- [15] Government of the Republic of Moldova, *Regulation on Environmental Quality Requirements for Surface Waters*, Government Decision No. 890/2013, online, accessed 02.01.2025. <https://www.legis.md>