DOI: 10.2478/auoc-2025-0015

Ovidius University Annals of Chemistry Volume 36, Issue 2, pp. 132 - 136, 2025

Time series analysis and forecasting of river pollution using statistical techniques and ARIMA: A case study of the Nistru River at Olanesti

Galina MARUSIC,[1] Nina SEVCENCO,[2] Victor ABABII[*],[1] Mihaela RUSANOVSCHI,[1] and Viorel CARBUNE[1]

[1]*Technical University of Moldova, Stefan cel Mare Street, 168, Chisinau, Republic of Moldova*
[2]*N. Testemitanu State University of Medicine and Pharmacy, Stefan cel Mare Street, 192, Chisinau, Republic of Moldova*

Abstract. This paper presents a statistical approach to water pollution analysis and forecasting based on time series data collected from a section of the Nistru River near Olănești, Republic of Moldova. The study focuses on three key pollutants: ammonia nitrogen, total phosphorus, and mineral phosphorus, using the values of concentrations recorded in 2019-2023. Descriptive statistics and exploratory visualizations were used to assess the variability, central trends and potential exceedance of maximum allowable concentrations (MACs). The forecasting methodology is based on the AutoRegressive Integrated Moving Average (ARIMA) model, widely recognized for its efficiency in modeling univariate time series with time dependence. The individual ARIMA models have been fitted and validated for each pollutant and the forecasts have been extended to the year 2026. The proposed methodology supports both environmental decision-making and early warning systems by integrating robust statistical models with domain-specific knowledge.

Keywords**:** time series analysis; ARIMA; statistical methods; pollution forecasting; Nistru River; ammonium nitrogen; phosphorus; water quality.

1. Introduction

The impact of pollutants on the human body is a complex issue, as most often the source of pollution is non-point source pollution. Non-point sources are difficult to identify because they are indirect sources of emissions and usually originate from multiple locations at a time. Nitrogen and phosphorus are naturally present in aquatic ecosystems, but human activities such as fertilizer use, wastewater management, fossil fuel combustion, soap and detergent runoff are polluting ecosystems with excess nutrients faster than they can adapt [1].

To this day, nitrogen and phosphorus compounds are widely used in various industries, such as chemical and metalworking. Their use in agriculture, including total and mineral phosphorus and ammoniacal nitrogen, is also possible as fertilizers. Fertilizers are a necessary tool for improving soil health. However, it should also be remembered that their thoughtless use can lead to soil, air, and water pollution. This pollution can occur both from the fertilizer components themselves and from their contaminants, including toxic micronutrients from inorganic fertilizers and pathogens from organic fertilizers. Since humans began using inorganic fertilizers, levels of nitrogen and phosphorus compounds in water, air, and soil have doubled over the past 100 years. This, of course, has a negative impact on plants, animals, and humans [2].

Today, the use of fertilizers in agriculture is increasing, especially in recent years, and information about their impact on human health and the environment is very limited. If we examine the report of the third session of the UN Environment Assembly, we find a lack of accessible information on the impact of fertilizers on both human health and the environment [3].

There is compelling evidence that exposure to inorganic fertilizers is associated with an increased risk of cancer, and a small number of studies even have methodological limitations and poor reproducibility. There are also scientific studies linking organic fertilizers to various infectious diseases and diarrhea. Agricultural runoff, which contains many toxic chemicals and pesticides, is particularly dangerous. These are typically intended to kill harmful insects and rodents. The most hazardous compounds found in this wastewater are nitrogen and phosphorus. People are typically exposed to these chemicals in several ways, for example, by eating fish, swimming, drinking contaminated water, or even inhaling polluted air. This causes a variety of health problems. These can include skin rashes, liver and kidney disease, neurological problems, or respiratory issues. Nitrates, a form of nitrogen most commonly found in fertilizers, can leach into drinking water in high concentrations, especially in agricultural areas, causing serious health problems. Infants are particularly susceptible to these negative impacts. Recent studies examining the effects of inorganic fertilizers in water and their impact on the human body have found a link between their exposure and the development of tumors in organs such as the lungs, breasts, and brain. Hematological malignancies have also been reported, accompanied by general

[*] Corresponding author. *E-mail address:* victor.ababii@calc.utm.md (Victor Ababii)

© 2025 Victor Ababii et al.
This is an open access article licensed under the Creative Commons Attribution-NonCommercial-NoDerivsLicence (https://creativecommons.org/licenses/by-nc-nd/3.0/).

symptoms such as skin lesions, neurological, respiratory, and a number of other diseases [4].

The current medical literature contains data on congenital disorders of various organs, developmental diseases, hematological diseases, infections, rheumatic diseases, diabetes, neurological, and vascular diseases. Data on the development of psychiatric diseases can also be found. If we consider a 2024 systematic review, we will find data indicating that exposure to inorganic fertilizers has been associated—in a limited number of studies—with an increased risk of oncological outcomes, including tumors of solid organs and hematological malignancies [5]. There is a study that long-term consumption of nitrates (usually obtained from fertilizers in drinking water) was associated with an increased risk of prostate cancer, and the effect was stronger for aggressive tumors subtypes [6]. Scientific studies on soil and water pollution show that chemicals from agricultural runoff, such as nitrates, heavy metals and other agropollutants, significantly contribute to the development of cardiovascular diseases. This occurs through mechanisms such as oxidative stress, endothelial damage and chronic inflammation [7].

Finally, a scientific study on the effects of heavy metals present in widely used nitrogen, phosphorus and potassium (NPK) fertilizers showed that after the application of fertilizers, the levels of metals such as Pb, Cu, Cd and Cr in water and sediments increase. These studies cannot but raise concerns among experts about the impact of these harmful substances through the food chain on aquatic flora and fauna and especially on human health in general [8].

Based on the above, it can be concluded that epidemiological data indicate possible links between inorganic fertilizers such as nitrogen and phosphorus compounds and various diseases, including malignant tumors. However, the available evidence is limited and heterogeneity prevails. Further research is needed to expand the evidence base and improve the reliability of replication and robustness of the results.

The assessment and prediction of water pollution are essential components of sustainable environmental management, especially in regions where water resources are vulnerable to anthropogenic pressures. River ecosystems such as the Nistru are exposed to various sources of contamination, including agricultural runoff, domestic effluents, and industrial discharge. Water pollution assessment and prediction are essential components of sustainable environmental management, especially in regions where water resources are vulnerable to anthropogenic pressures. In this context, statistical analysis plays a central role in understanding historical trends and in supporting informed decision-making for pollution control and mitigation [9].

Exploratory data analysis (EDA) provides a structured approach to identifying patterns, anomalies, and distributional features in environmental data sets. Techniques such as descriptive statistics, histograms, and skewness coefficients are commonly used to uncover the underlying structure of time-series data prior to formal modeling. EDA not only improves data quality assessment, but also guides the selection of appropriate modeling techniques, revealing seasonality, trends, and outliers [10, 11]. It is particularly important in environmental time-series studies, where issues such as missing values, non-stationarity or auto-correlation may bias forecasts if not first explored and addressed. Recent research underlines how EDA enables the detection of structural changes and hidden temporal dependencies before applying predictive models, improving both interpretability and forecast reliability [12, 13, 14].

The ARIMA (AutoRegressive Integrated Moving Average) model is widely used for pollution prediction. By incorporating autoregressive components and moving averages, ARIMA models can produce accurate short-term forecasts, which are essential for environmental monitoring and early warning systems. The use of ARIMA in water quality prediction has been validated in several studies, strengthening its relevance in environmental statistics. One of the key advantages of the ARIMA model lies in its flexibility to capture both trend and seasonality components in time-dependent data. Moreover, by optimizing parameters through techniques such as the Akaike Information Criterion (AIC), the model can achieve a balance between accuracy and complexity, ensuring robust and interpretable forecasts [15, 16, 17].

Applications show that ARIMA and its seasonal extensions (*e.g.*, SARIMA) remain competitive even when compared to more complex machine-learning approaches in the water-quality domain. Furthermore, the interpretability of ARIMA models makes them particularly suitable for regulatory and management contexts, where understanding model structure and residual behaviour is critical [18, 19].

2. Methodology

This study investigates the temporal evolution and forecasting of key pollutants in a monitored sector of the Nistru River, using time series data collected over a five-year period. The study of pollutant dynamics in river ecosystems requires both a descriptive understanding of the data and the ability to model and predict future concentrations. For this purpose, the methodological approach used in this study includes Exploratory Data Analysis (EDA) and time series forecasting with the ARIMA model.

2.1. Data

The data for the analysis was taken from monthly reports on environmental quality in the Republic of Moldova for the period 2019-2023, provided by the Environmental Agency [20]. Data was collected with a total of seven measurements for mineral phosphorus, six for total phosphorus, and five for ammonium nitrogen. Only the values of concentrations exceeding the maximum allowable concentrations (MACs) were recorded. The MAC values for Class I water quality were considered in accordance with the Regulation on Environmental Quality Requirements for Surface Waters [21], namely 0.05 mg/L for mineral phosphorus, 0.1 mg/L for total phosphorus, and 0.2 mg/L for ammonium nitrogen. For each year in the mentioned period, between 1 and 2 samples each for the three key pollutants (mineral phosphorus, total phosphorus and ammonium nitrogen) were available in the dataset.

Mineral phosphorus concentrations ranged from 0.11 to 0.44 mg/L, with the highest level observed in December 2019 and the lowest in January 2023. Except for the 2019 maximum, most values clustered between 0.11–0.16 mg/L, suggesting relative stability and a slight decreasing trend in the later years. Total phosphorus showed a wider interval, between 0.19 and 0.59 mg/L, peaking in December 2019. After 2020, concentrations stabilized in a narrower range (0.20–0.24 mg/L), pointing to more balanced conditions. Ammonium nitrogen values were between 0.08 and 0.70 mg/L, with the exception of an extreme of 8.3 mg/L recorded in September 2022.

2.2. Analysis methods for the pollutants

For the analysis of ammonium nitrogen, total phosphorus, and inorganic phosphorus, Exploratory Data Analysis (EDA) techniques were applied, followed by time series modeling using the ARIMA (AutoRegressive Integrated Moving Average) approach to forecast future pollutant concentrations.

Exploratory Data Analysis (EDA). Exploratory Data Analysis serves as a preliminary step in any statistical investigation and is crucial in environmental science, especially when assessing water quality over time. The main goal of EDA is to summarize the main characteristics of the data before applying more complex modeling techniques. It helps to identify patterns, detect outliers, highlight trends or seasonality, and understand the variability within the data set [10, 13].

In this study, basic descriptive statistics were used to analyze the temporal variation in the concentrations of pollutants such as ammonium nitrogen, total phosphorus, and mineral phosphorus. The mean and standard deviation provide insight into the central tendency and dispersion of the data, while the skewness coefficient indicates the asymmetry of the distribution. These indicators are essential for assessing the level of fluctuations and the potential risk of exceeding regulatory thresholds.

Graphical techniques were also employed, such as histograms and time plots, which allow for visual identification of changes over time. Time series plots in particular are valuable for detecting non-random structures in the data, such as trends, cycles, or abrupt changes due to natural or anthropogenic factors. These representations support the detection of pollution events and help inform the selection of appropriate statistical models for forecasting.

Time Series Forecasting with ARIMA. To forecast future pollutant concentrations, this study used the ARIMA (AutoRegressive Integrated Moving Average) model, which is widely recognized for its robustness in modeling and predicting univariate time series [12, 13, 14]. The ARIMA model integrates three key components:

- The autoregressive part models the dependency between an observation and a specified number of its own previous values.
- The integrated part refers to the process of differencing the data to achieve stationarity — a property necessary for consistent forecasting.
- The moving average part models the relationship between an observation and past forecast errors.

The modeling process follows the Box–Jenkins methodology, which involves identifying an appropriate model structure based on autocorrelation patterns, estimating model parameters, and performing diagnostic checks to validate model assumptions.

The ARIMA model is particularly suitable in environmental studies where historical data is available, but external variables are limited or not incorporated. In this case, it allows for medium-term projections of pollutant levels and the evaluation of whether forecasted concentrations are likely to exceed the maximum admissible concentration (MAC) set by environmental standards.

3. Results and discussion

Data on pollution of the Nistru River sector in Olănești, Republic of Moldova were analyzed. It was found that in the sector of the mentioned frequent branches, during the period 2019-2023, the maximum permissible concentrations for the pollutants: mineral phosphorus, total phosphorus and ammonium nitrogen were exceeded. For this reason, the data on the values of the listed pollutant concentrations were subjected to analysis using statistical methods and the ARIMA algorithm.

First, descriptive statistics were applied for all three pollutants. The results obtained using the R language are presented in Table 1.

Table 1. Descriptive statistics.

Pollutant	Sample count	Mean (mg/L)	Median (mg/L)	Std. deviation	Min (mg/L)	Max (mg/L)
Mineral phosphorus	7	0.174	0.129	0.119	0.11	0.44
Total phosphorus	6	0.331	0.240	0.175	0.19	0.59
Ammonium nitrogen	5	2.070	0.683	3.490	0.08	8.30

We can conclude that *Mineral and total phosphorus* show moderate and steady exceedances. *Ammonium nitrogen* displays critical and irregular peaks, possibly indicating occasional pollution events or failures in waste treatment systems.

Next, the data under examination were presented graphically using a histogram (see Figure 1).

From Figure 1 we observe the frequency distribution of concentrations (mg/L) for three types of pollutants measured in a river sector:

1. Mineral phosphorus (red bars)
   - The majority of observations (3 occurrences) fall within the 0.1–0.2 mg/L range.
   - A smaller number (1 observation) falls into the 0.3–0.4 mg/L range.
   - Mineral phosphorus levels are generally low, clustering near the lower end of the scale.

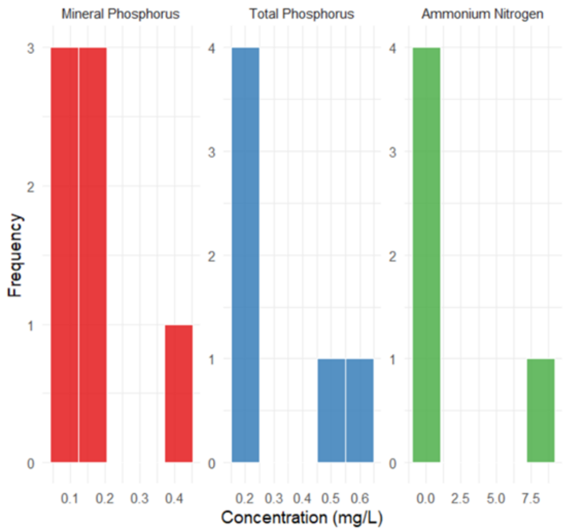

Figure 1. Distribution of exceeding concentrations by pollutant.

2. Total phosphorus (blue bars)
   • The highest frequency (4 values) is observed in the 0.2–0.3 mg/L interval.
   • Two other values appear in the 0.4–0.5 and 0.5–0.6 mg/L intervals.
   • Although total phosphorus concentrations are generally moderate, some higher values are also present, suggesting variability in overall phosphorus input.

3. Ammonium nitrogen (green bars)
   • Most values (4 observations) fall below 2.5 mg/L, indicating a generally low concentration.
   • However, a noticeable outlier (1 observation) is in the 7.5–10 mg/L range.
   • While ammonium nitrogen is usually at a low level, the presence of an outlier suggests a potential pollution event or spike that should be further investigated.

To make prediction scenarios for the analyzed pollutants for the years 2024-2026, the ARIMA algorithm was applied. The results for mineral phosphorus are presented in Figure 2.

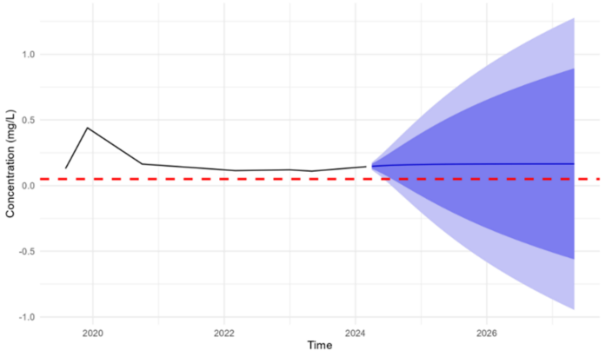

Figure 2. Forecast for mineral phosphorus.

The detailed interpretation of the results in Figure 2 is as follows:

1. Black line: Represents the historical values of the pollutant concentration for the period 2019–2024. A slight decrease and stabilization over time can be observed.

2. Blue line: Shows the forecasted average concentration for the years 2025 and 2026, based on the ARIMA model.

3. Blue bands (confidence intervals):
   • The darker blue area indicates the 80% confidence interval.
   • The lighter blue area corresponds to the 95% confidence interval.

These bands reflect the uncertainty of the estimation — actual values may fluctuate within this range.

4. Red dashed line: Marks the regulatory threshold (maximum admissible concentration) for the pollutant. If the forecast exceeds this line, it indicates a potential risk of pollution above the legal limit.

For the years 2025–2026, the model estimates that average values will remain close to, but still below, the admissible limit. However, uncertainty increases over time — after 2025, the confidence intervals widen significantly. Although the mean forecast remains safe, there is a real probability that actual values may temporarily exceed the threshold if they fall within the upper bound of the blue confidence band.

The results of applying ARIMA for *total phosphorus* are shown in Figure 3.

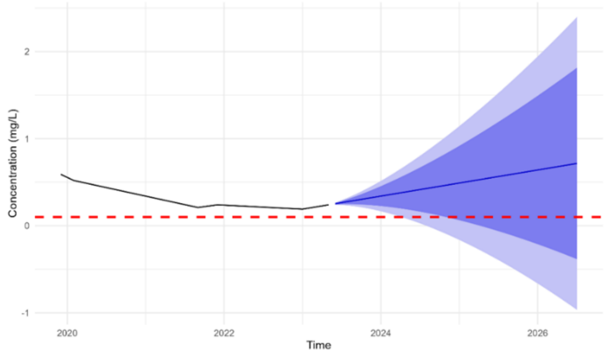

Figure 3. Forecast for total phosphorus.

The detailed interpretation of the results in Figure 3 is as follows:

1. Black line (historical data):
   • This line shows the observed concentrations of Total Phosphorus from 2019 to 2023.
   • The trend appears to be slightly decreasing and relatively stable, with values close to or slightly above the MAC (Maximum Admissible Concentration).

2. Blue line (forecast mean):
   • Represents the ARIMA-predicted average concentrations for the years 2024 to 2026.
   • The line shows a slight upward trend, indicating a potential increase in concentration over time.

3. Blue shaded areas (confidence intervals):
   • The inner and outer shaded regions represent 80% and 95% confidence intervals, respectively.
   • The widening funnel shape reflects increasing uncertainty in long-term forecasts.
   • By 2026, values could range from well below to significantly above the MAC.

4. Red dashed line (MAC – Maximum Admissible Concentration):
   • Set at 0.1 mg/L, this is the reference threshold for safe phosphorus levels.

• Historical values are slightly above, and predicted values have a high probability of remaining above the MAC, especially the upper bound of the forecast.

The results of applying ARIMA for *ammonium nitrogen* are shown in Figure 4.

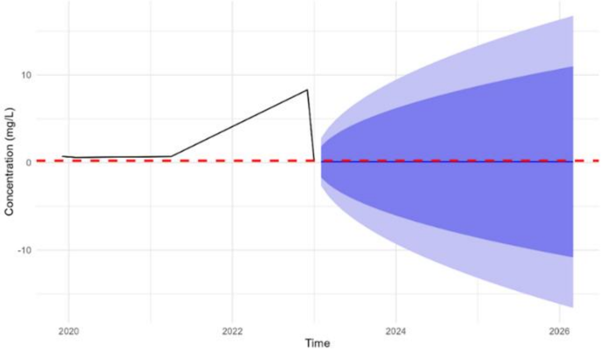

Figure 4. Forecast for ammonium nitrogen.

The detailed interpretation of the results in Figure 4 is as follows:

1. Black line: Represents the historical values of the pollutant concentration from 2019 to 2024. The graph shows a slight increase between 2021 and 2023, followed by a sudden drop before entering the forecast period.

2. Blue line: Depicts the forecasted average concentration for the years 2025 and 2026 based on the ARIMA model.

3. Blue bands (confidence intervals):
   • The darker blue area represents the 80% confidence interval.
   • The lighter blue area shows the 95% confidence interval.

These intervals express the range of possible future values, considering model uncertainty. A wider band indicates greater uncertainty in the prediction.

4. Red dashed line: Indicates the *maximum admissible concentration* according to environmental regulations.

Although the historical data shows values fluctuating around the admissible limit, the forecast suggests that the mean predicted concentration for 2025–2026 will stabilize close to the normative limit. However, the uncertainty grows significantly into the future — as seen from the widening blue bands. This means there's a growing chance that actual future values could fall well above or below the average estimate.

4. Conclusions

This study demonstrates the utility of statistical methods – in particular Exploratory Data Analysis (EDA) and ARIMA time series modeling – for assessing and forecasting the temporal evolution of pollutants in the Nistru River near Olănești. Forecasts were generated for the period up to 2026, supporting the use of data-driven approaches to monitor environmental conditions and inform local water management strategies.

The three most frequently encountered pollutants were assessed: mineral phosphorus, total phosphorus and ammonium nitrogen. The following findings were made:

• Mineral phosphorus: Projections indicate that concentrations are likely to remain below the regulatory limit until 2026, but occasional exceedances remain possible. Continued monitoring is recommended.
• Total phosphorus: Levels are expected to increase moderately until 2026, with a high risk of exceeding the MAC. Preventive or mitigation actions are needed due to significant uncertainty in the forecast.
• Ammonium nitrogen: Forecasted concentrations remain close to the safety threshold through 2026, but wide confidence intervals highlight the importance of sustained monitoring.

Overall, statistical forecasting provides significant insights into the behavior of pollutants, enabling early warning signals and supporting ecological protection efforts through informed decision-making.

Acknowledgments

This research was carried out under the Contract No. 01-23p-018/02-48-2024 *"Development and elaboration of computational solutions for obtaining numerical models of the behavior and prediction of pollutant dispersion in time and space in river-type aquatic ecosystems,"* which is part of the National Environment Fund Program , of PI National Office for the Implementation of Projects in the Environmental Domain, with the technical and technological support provided by the Department of Informatics and Systems Engineering, Technical University of Moldova.

Conflict of interest

The authors declare that they have no conflict of interest.

References

[1]. A. Smith Haghighi, What does the water pollution have on human health, Medical News Today (2024), online, accessed 03.01.2025. https://www.medicalnewstoday.com/articles/water-pollution-and-human-health

[2]. Atlas Scientific, 7 effects of water pollution, (2025), online, accessed 02.01.2025. https://atlas-scientific.com/blog/effects-of-water-pollution

[3]. A. Mazeikeine, Removal of nitrogen and phosphorus from wastewater using layered filter media, Sustainability 14 (2022) 10713. DOI: 10.3390/su141710713

[4]. European Environment Agency, Health impacts of contaminants from fertilisers, March 2025, online, accessed 09.03.2025. https://www.eea.europa.eu/en/european-zero-pollution-dashboards/indicators/health-impacts-of-contaminants-from-fertilisers-signal-1

[1]. Ch.F. Tagkas, E.C. Rizos, G. Markozannes, E.E. Ntzani, Fertilizers and human health—A systematic review of the epidemiological evidence, Toxics 12 (2024) 694. DOI: 10.3390/toxics12100694.

[5]. C. Donat-Vargas, M. Kogevinas, G. Castaño-Vinyals, N. Aragonés, A. Pérez-Gómez, R. Pollán, M. Garcia-Closas, J.L. Llorca, J.M. Huerta, M. Dorronsoro, Long-term exposure to nitrate and

trihalomethanes in drinking water and prostate cancer: A multicase–control study in Spain (MCC-Spain), Environmental Health Perspectives 131 (2023) 037004. DOI: 10.1289/EHP11391
[6]. T. Münzel, O. Hahad, J. Lelieveld, S. Daiber, A. Münzel, Soil and water pollution and cardiovascular disease, Nature Reviews Cardiology 22 (2025) 71–89. DOI: 10.1038/s41569-024-01068-0
[7]. S. Naz, M. Arshad, S. Majeed, M. Hussain, M.H. Rashid, M. Arif, Assessing heavy metal contamination in commonly used fertilizers for polyculture fish ponds and its implications for human health: A comprehensive investigation, Biological Trace Elements Research 203 (2025) 3318–3334. DOI: 10.1007/s12011-024-04409-9
[8]. D.R. Helsel, R.M. Hirsch, Statistical methods in water resources, Techniques of Water Resources Investigations, Book 4, Chapter A3, U.S. Geological Survey, 2002.
[9]. W. Tukey, Exploratory data analysis, Addison-Wesley, 1977.
[10]. G. Marusic, R. Braniște, N. Sava, V. Ababii, V. Cărbune, T. Tiholaz, Statistical evaluation of data on the pollution of the Prut River with ammonium nitrogen, Journal of Engineering Science 31 (2024) 166–174. DOI: 10.52326/jes.utm.2024.31(4).12
[11]. D. Ries, K. Goode, K. McClernon, B. Hillman, Using feature importance as an exploratory data analysis tool on Earth system models, Geoscientific Model Development 18 (2025) 1041–1065. DOI: 10.5194/gmd-18-1041-2025
[12]. U. Javed, K. Ijaz, M. Jawad, S. Khalid, A. Ahmed, Exploratory data analysis based short-term electrical load forecasting: A comprehensive analysis, Energies 14 (2021) 5510. DOI: 10.3390/en14175510
[13]. H. Lange, M. Hauhs, Complexity analysis of environmental time series, Entropy 27 (2025) 381. DOI: 10.3390/e27040381
[14]. G.E.P. Box, G.M. Jenkins, G.C. Reinsel, G.M. Ljung, Time series analysis: Forecasting and control, 5th ed., Wiley, 2015.
[15]. R.J. Hyndman, G. Athanasopoulos, Forecasting: Principles and practice, 3rd ed., OTexts, 2021.
[16]. C. Chatfield, The analysis of time series: An introduction, 6th ed., Chapman & Hall/CRC, 2003.
[17]. T. Gowthaman, K. Sathees, V.S. Adarsh, B. Bhattacharyya, Trend analysis and ARIMA models for water quality parameters of Brahmani River, Odisha, India, International Journal of Environment and Climat Change 12 (2022) 219–228. DOI: 10.9734/ijecc/2022/v12i121457
[18]. D.Y. Hardiyanti, H. Novianty, D. Lestarini, Forecasting river water quality using autoregressive integrated moving average (ARIMA), Proc. CONRIST 2019, pp. 158–163.
Environment Agency, Monthly bulletin on environmental quality in the Republic of Moldova, 2025, online, accessed 04.01.2025. https://am.gov.md/ro/node/215
[19]. 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

Received: 13.07.2025
Received in revised form: 19.10.2025
Accepted: 30.10.2025