Electronics and Information Technologies, 2026, 35, 97–108
http://publications.lnu.edu.ua/collections/index.php/electronics/index

UDC: 519.246.8:504.064.2

ANALYSIS AND FORECASTING OF RIVER WATER POLLUTION DYNAMICS USING LONG SHORT-TERM MEMORY NEURAL NETWORKS AND MULTIFRACTAL ANALYSIS

***Mykhailo Bordun*[1] & *Yaroslav Sokolovskyy*[2]**
[1] *Information Systems Department, Ivan Franko National University of Lviv, 1 Universytetska Street, Lviv, 79007, UKRAINE*
[2] *Department of Computer-Aided Design, Lviv Polytechnic National University, 5 Metropolyta Andreia Street, Lviv, 79016, UKRAINE*

Bordun, M.I., Sokolovskyy, Y.I. (2026). Analysis and Forecasting of River Water Pollution Dynamics Using Long Short-Term Memory Neural Networks and Multifractal Analysis. *Electronics and Information Technologies*, 35, 97–108. https://doi.org/10.30970/eli.35.7

ABSTRACT

Background. Organic water pollution has complex, nonlinear, nonstationary, and self-similar dynamics that conventional forecasting models often fail to represent. This study develops an automated hybrid framework combining multifractal analysis and deep learning to improve Biochemical Oxygen Demand (BOD) forecasting and support early risk identification.

Materials and Methods. A retrospective BOD time series from a river monitoring station was processed through an automated data and modeling pipeline. Sliding windows extracted localized features and captured long-range correlations. Multifractal Detrended Fluctuation Analysis (MFDFA) estimated generalized Hurst exponents and singularity spectra for each window. A fractional-exponential approximation converted the continuous spectra into discrete coefficients while preserving their structural characteristics. These multifractal descriptors were combined with statistical measures to form the feature space. A Long Short-Term Memory (LSTM) network was trained using time-series cross-validation and hyperparameter search. Results were compared with Autoregressive Integrated Moving Average (ARIMA) and Autoregressive Fractionally Integrated Moving Average (ARFIMA) benchmarks.

Results and Discussion. The hybrid MFDFA-LSTM framework produced more accurate and stable forecasts than the benchmark models across short-term and extended horizons. Fractional-exponential descriptors supplied information about multi-scale fluctuations, persistence, spectrum width, and asymmetry, enabling the LSTM to represent nonlinear pollution dynamics more effectively. Hyperparameter optimization balanced temporal memory, model complexity, and generalization, reducing root mean square error (RMSE) and maintaining stable predictive performance.

Conclusion. The study validates an automated system for water-quality forecasting. Its innovation is transforming continuous multifractal spectra into informative discrete features through fractional-exponential approximation. Combined with LSTM modeling, this framework improves forecast reliability and provides environmental authorities with an early-warning and decision-support tool for proactive water-quality management.

Keywords: LSTM, BOD, multifractal analysis, MFDFA, ARFIMA, singularity spectrum.

INTRODUCTION

The aquatic environment is currently undergoing significant transformation driven by the dual pressures of global climate change and the rapid expansion of industrial and agricultural

© 2026 Mykhailo Bordun & Yaroslav Sokolovskyy. Published by the Ivan Franko National University of Lviv on behalf of Електроніка та інформаційні технології / Electronics and Information Technologies. This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 License which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

sectors [1, 2]. Statistical data indicates a staggering global discharge of approximately 400 billion cubic meters of wastewater into natural water bodies, leading to the contamination of nearly 5.5 trillion cubic meters of water worldwide [3]. Critically, over 80% of this wastewater enters the environment without adequate treatment. This escalating anthropogenic burden systematically erodes the self-restoration capacity of aquatic ecosystems [4].

The Dnipro River is fundamental to Ukraine's water security, supplying more than 70% of the population and driving the country’s industrial and agricultural productivity. However, the basin is currently under immense anthropogenic stress. Among the pollutants recorded by the monitoring network, organic substances pose a significant threat to the ecosystem’s health [5]. The resulting depletion of dissolved oxygen, combined with nutrient loading from urban and agricultural sources, fosters massive water 'blooming' through the intensive reproduction of blue-green algae.

The diversity of contaminants infiltrating river systems underscores the absence of a singular, universal monitoring methodology. Effective ecosystem management necessitates continuous, real-time tracking of various physicochemical parameters to ensure compliance with environmental standards. Among these, Biochemical Oxygen Demand (BOD) serves as a critical integrative indicator of water quality. It quantifies the concentration of dissolved oxygen consumed by aerobic microorganisms during the biochemical oxidation of organic matter into stable inorganic compounds [6]. Consequently, BOD provides a comprehensive assessment of the biological load on the aquatic environment. Elevated BOD levels signify excessive contamination from municipal or agricultural discharges, posing a severe risk of hypoxia and significantly impeding the river's natural self-purification mechanisms.

Notwithstanding the rapid advancements in electronic and sensor technologies, the direct real-time measurement of BOD via physicochemical sensors remains unattainable. Consequently, traditional laboratory methods are inherently incapable of providing the early-warning functions necessary for environmental disaster prevention. To address this limitation, environmental monitoring practices are increasingly shifting toward predictive strategies. By leveraging easily measurable parameters — such as temperature, pH, electrical conductivity, dissolved oxygen, and turbidity-mathematical models can 'calculate' probable BOD concentrations. This approach underpins the 'soft sensor' concept: a software-based algorithm that acts as a virtual replacement for complex laboratory analyses [7, 8]. This paradigm necessitates the application of neural networks, which are uniquely suited to identifying the intricate, nonlinear correlations between these proxy indicators. Alternatively, predictive modeling can be driven by longitudinal datasets of historical BOD observations [9-12]. The precision of such forecasts is directly correlated with the scale and quality of the historical database.

While standard methods, such as Support Vector Regression (SVR), provide reliable point forecasts, they often fail to capture the underlying complexity of dynamic systems. This limitation stems from two primary factors: first, the 'inertial memory' inherent in the Dnipro River system, where current water quality is intrinsically linked to upstream historical events — a dependency that classical algorithms often treat as a series of independent occurrences. Second, the physicochemical parameters within the Dnipro basin exhibit intricate nonlinear correlations; for instance, the relationship between dissolved oxygen and BOD is highly non-stationary and fluctuates with seasonal temperature shifts. Building upon the findings of recent research [13], which highlights the potential of hybrid modeling for such complex systems, this article aims to develop a synergistic framework for BOD prediction. This approach integrates a Long Short-Term Memory (LSTM) recurrent architecture, capable of retaining systemic temporal dependencies, with multifractal analysis for advanced data characterization and preprocessing.

MATERIALS AND METHODS

Software tools analysis

The computational framework for this study integrates both Python and R to leverage their respective analytical strengths. Data preprocessing and the implementation of the LSTM architecture were performed in Python, utilizing TensorFlow for deep learning and NumPy and Pandas for efficient numerical operations. However, the multifractal characterization and fractal modeling using the AutoRegressive Fractionally Integrated Moving Average (ARFIMA) were executed in the R environment. This dual-language approach was adopted because R offers a more mature and comprehensive ecosystem for fractal analysis, specifically through the mfdfa, forecast, and arfima packages, which currently lack equivalent robust implementations in Python's standard libraries.

Dataset of river water pollution

Data utilized in this research originates from the official environmental monitoring program of Ukraine, accessed via data.gov.ua [14]. The dataset encompasses a wide array of water quality indicators collected from 2003 onwards across major river basins. To ensure the high-frequency and long-term data needed for multifractal modeling, this study focuses on the Dnipro-Donbas Canal station. This site was selected because it provides the largest dataset, containing 410 observations. While the monitoring station records various indicators such as nitrogen, ammonium ions and dissolved oxygen levels, BOD was specifically isolated as the critical variable for assessing organic pollution due to its comprehensive ability to reflect the oxygen-consuming potential of organic matter.

Raw data were retrieved using automated HTTP GET protocols, yielding numerous CSV-formatted files covering various time intervals. The data extraction focused exclusively on monitoring records from the Dnipro-Donbas Canal station. Following retrieval, the dataset underwent preprocessing to address non-recorded BOD measurements, thereby validating its suitability for deep learning applications. The final, refined time series was structured to include only chronological date records and their corresponding BOD values.

Investigating seasonality, trend, stationarity, and long-memory presence

To ensure the best model selection, we first analyzed the characteristics of the Dnipro River BOD time series. Following data cleaning and preparation, we studied the measurements' statistical properties to build a reliable foundation for the forecasting process.

Fig. 1 depicts the time series decomposition, highlighting well-defined seasonal fluctuations and an overall trend. The remaining residual series captures the noise component, providing a clear visualization of data outliers that deviate from the established patterns. The residual analysis reveals a high degree of model convergence over the stable period. However, distinct spikes in the residual plot at the temporal boundaries suggest transient non-linear shocks in the BOD levels that exceed the predictive capacity of the trained model. These anomalies highlight the stochastic nature of the river ecosystem.

An important characteristic of time series analysis is examining evidence of long-term memory. **Fig. 2** presents the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) graphs. The ACF measures the linear relationship between the current value of the series and its past values at specific lags, capturing both direct and indirect effects. In contrast, the PACF isolates the direct correlation between an observation and its lag by removing the influence of intermediate observations. In the analyzed dataset, several time lags in both the ACF and PACF plots extend beyond the 95% confidence interval (depicted by the shaded blue area). This observation confirms that the series possesses a structured memory, which carries critical implications for model selection.

To validate the stationarity of the Dnipro River BOD data, we applied the ADF and KPSS statistical tests. The ADF test rejected the null hypothesis of non-stationarity

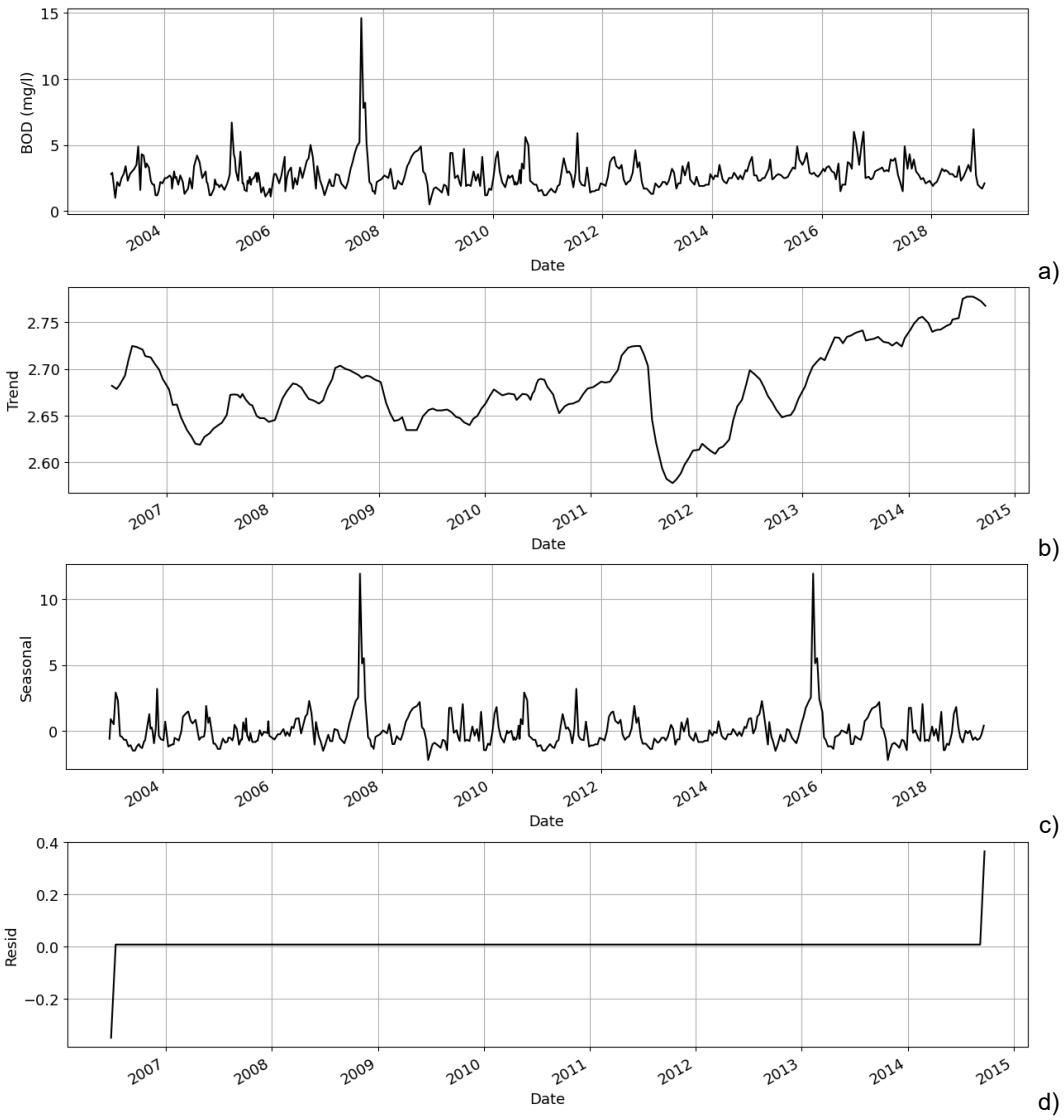


Fig. 1. BOD time series of the Dnipro River (a), decomposed trend (b), seasonality (c) and residuals (d)

($p < 0.05$), while the KPSS test failed to reject the null hypothesis of stationarity ($p > 0.05$) [15]. The consistency between these two complementary tests confirms that the time series does not possess a unit root and is characterized by stationary dynamics, facilitating more accurate modeling.

Mathematical description of the MFDFA method

While sharing core conceptual mechanics with conventional Detrended Fluctuation Analysis (DFA), Multifractal Detrended Fluctuation Analysis (MFDFA) [16] introduces expanded evaluation phases designed to analyze a diverse spectrum of scaling exponents. Consequently, this technique has become a standard analytical instrument for uncovering hidden non-linear structures and determining the degree of multifractality in temporal data.

The execution of the MFDFA algorithm yields a distinct set of mathematical outputs that characterize the multi-scale structure of the time series. The key mathematical vectors output by this pipeline are the generalized Hurst exponent curve (h(q)), the scaling exponent function (τ(q)), and the singularity spectrum (f(α)). This spectrum plots the

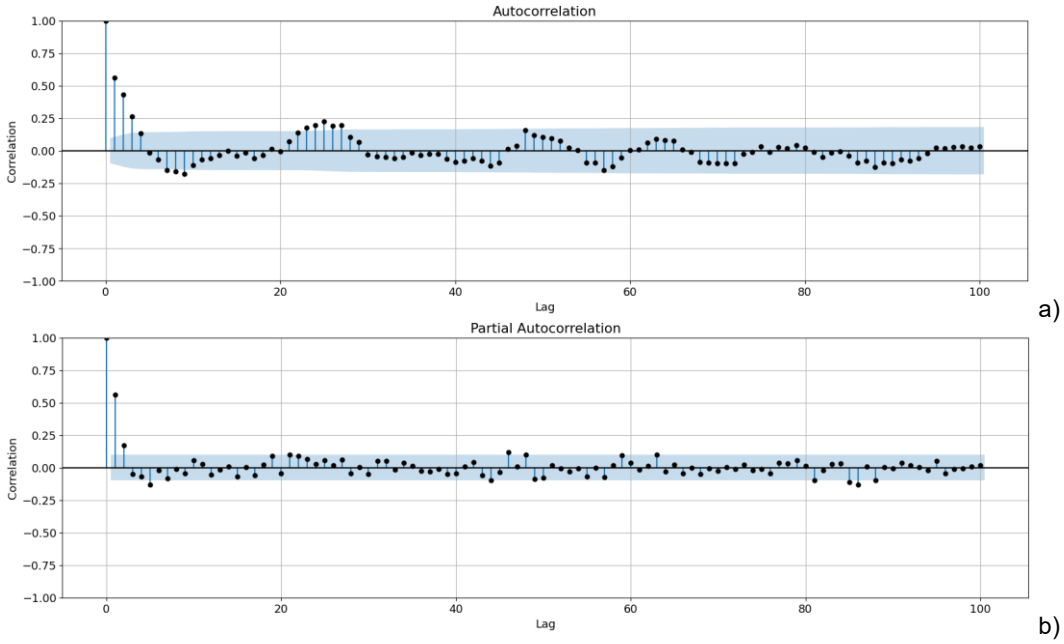


Fig. 2. ACF (a) and PACF (b) functions for the BOD time series of the Dnipro River

singularity dimension against the singularity strength α, providing the foundational curves necessary for calculating crucial structural parameters such as the spectrum width, asymmetry, and peak location.

The mathematical apparatus for the MFDFA processing and the subsequent singularity spectrum f(α)) derivation follows the exact theoretical framework detailed in [12, 16].

To derive quantitative descriptors, the multifractal spectrum is approximated by a second-order polynomial centered at its maximum α_0 by the following function:

$$f(\alpha) = A(\alpha - \alpha_0)^2 - B(\alpha - \alpha_0) + C, \quad (1)$$

where the coefficients A, B, and C are obtained by the least squares method [16].

To provide a more versatile and adaptable fit for the multifractal spectrum, an exponential function was proposed as follows:

$$f(\alpha) = A(\alpha - \alpha_{min})^{\beta_1}(\alpha_{max} - \alpha)^{\beta_2} + \varepsilon, \quad (2)$$

where the coefficients A, β_1, β_2 and ε are obtained by the nonlinear least squares method.

RESULTS AND DISCUSSION

The raw time series was organized into sliding windows of 150 observations each, advancing by one point at a time. This window size provided enough observations for reliable multifractal and statistical analysis while preserving sufficient temporal resolution. This approach allowed us to calculate multifractal and statistical descriptors for every step. We chose this specific window size to balance two needs: providing enough data for an accurate multifractal analysis and creating a long enough sequence to effectively train the LSTM model.

To quantify the multifractal properties of each data window, we employed the MFDFA algorithm, utilizing the specialized MFDFA R package. Key parameters included a defined scale range for segment-wise detrending and a comprehensive set of q-values.

The scaling range for the MFDFA procedure was optimized based on the temporal span of the dataset to ensure a robust and statistically reliable detrending process. We defined the minimum scale at 5 data points ($s_{min} = 5$) to maintain accuracy in the local fluctuation calculation. The maximum scale (s_{max}) was dynamically set at 25% of the window length (N/4), following standard recommendations to avoid artifacts caused by insufficient segment counts at large scales. To achieve a uniform distribution across all time scales, the values were calculated using an exponential sequence of 25 unique points, determined by:

$$s = \exp\left[\ln(s_{min}) + i * \frac{\ln(s_{max}) - \ln(s_{min})}{n - 1}\right], \quad (3)$$

where i ranges from 0 to n – 1, and n = 25.

The q parameter, ranging from –10 to 10, served as a multi-resolution filter: negative values were used to capture the scaling behaviour of small fluctuations, while positive values were dedicated to assessing the scaling of large-magnitude events, thereby providing a complete multifractal profile of the river's organic load.

The MFDFA procedure generated a comprehensive set of parameters, including the Hurst exponent, scaling exponent, singularity spectrum, and fluctuation function. To characterize each sliding window with discrete multifractal properties, the following values were extracted: the Hurst exponent at q = 2 (H(2)), which describes the persistence and long-range correlations of the signal; ΔH, which reflects the variation of the Hurst exponent across different (q)-orders; the spectrum width $\Delta\alpha$, which quantifies the degree of multifractality; α_0, the singularity exponent corresponding to the maximum of the spectrum; the asymmetry parameter B, and the estimated coefficients A, β_1, β_2 derived from the exponential expression in Eq. (2). Here, A controls the overall amplitude of the approximated spectrum, while β_1 and β_2 determine its shape on the left and right sides, respectively. Their relative values describe how rapidly the spectrum rises and falls and therefore provide additional information about its asymmetry.

A comparison of fitting residuals confirmed that the fractional-exponential model provided a more precise representation of the empirical data. This is evidenced by its lower root mean square error (RMSE) of 0.012, which represents a substantial improvement over the quadratic fit's error of 0.034. The superior performance of this approach in representing the empirical $f(\alpha)$ values were statistically proven in our analysis and visually demonstrated in the comparative plots in **Fig. 3**.

To construct the high-dimensional feature space for the 100-step-ahead forecast, we integrated discrete multifractal properties with fundamental statistical descriptors. This feature vector included volatility, slope, mean, entropy, skewness, kurtosis, and data-window width. The computational extraction was performed using the R-language, with the resulting multidimensional dataset exported as a CSV file to a local data repository. This intermediate storage facilitated a seamless transition to the Python-based environment for the subsequent modeling and forecasting phases.

The original time series contained 410 BOD observations. Applying a 150-observation sliding window produced 261 possible feature windows. However, the last window could not be paired with a subsequent BOD observation; therefore, 260 valid feature–target pairs were available for forecasting. The final 100 pairs were reserved as an independent chronological test set, leaving (N = 160) observations for model development. Thus, (N = 160) refers to the number of labeled features — target samples used for training and validation, rather than the number of raw measurements. The 100-step-ahead forecast should also be understood as a 100-observation forecast horizon composed of consecutive one-step-ahead predictions.

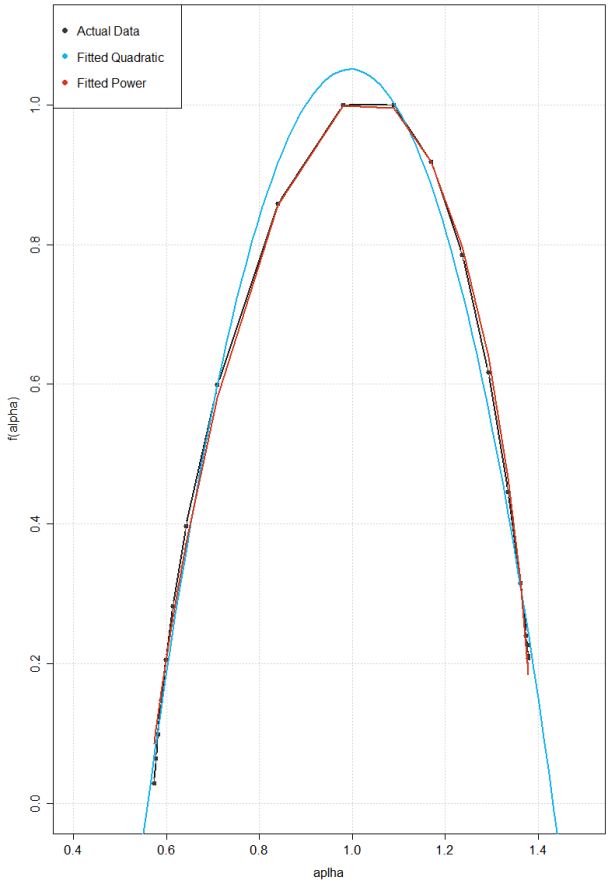


Fig. 3. Singularity spectrum and its quadratic (blue line) and fractional-exponential (red line) approximations

The predictive modeling phase was initiated by aligning each high-dimensional feature vector with its corresponding target BOD observation. We structured the data for supervised learning by aligning the multifractal and statistical features of each 150-observation window with the immediate following BOD record. In other words, the first 150 data points were used to predict the 151st observation, the next shifted window predicted the following observation, and so forth. This one-step-ahead mapping generated a continuous sequence of training pairs. The data were divided chronologically, without random shuffling, to prevent information from future observations from entering the training process. Within the first 160 observations, three expanding time-series folds were used for validation: earlier observations formed the training subset, while the subsequent contiguous observations formed the validation subset. The final 100 observations remained untouched until the final performance evaluation.

The LSTM was selected for its ability to utilize internal memory cells and gating mechanisms, which allow the model to retain long-term dependencies and capture the “inertial memory” inherent in the Dnipro River’s pollution dynamics [17]. The model used a many-to-one structure: several consecutive feature vectors were provided as input, and the network generated one prediction for the immediately following BOD observation. In the final configuration, the input consisted of consecutive time steps and the selected feature variables. A single LSTM layer with hidden units and a hyperbolic-tangent activation function transformed this sequence into a temporal representation. Each LSTM unit maintained both a cell state and a hidden state, while the input, forget, and output gates controlled which information was stored, discarded, or passed to the next time step. The

resulting hidden representation was passed through a dropout layer to reduce overfitting and then through a fully connected output layer with one neuron for the BOD prediction. The network was trained using the Adam optimizer and mean squared error loss.

To ensure the LSTM architecture remained robust despite a relatively small sample size (N = 160), a feature selection process was conducted prior to training. By limiting the input space to an optimal range of 5–7 features, we prevented the model from capturing noise, thereby addressing potential overfitting constraints. Feature and target values were scaled using parameters calculated from the corresponding training data. This dimensionality reduction, together with chronological validation, dropout regularization, and early stopping, was critical for maintaining the reliability of the multifractal-based forecasting.

We employed a tree-based Random Forest Regressor algorithm to perform feature importance estimation. This allowed for a streamlined selection of the most influential predictors, ensuring a more efficient and robust forecasting process by utilizing only the highest-ranking features. The forecast was executed using a refined feature space consisting of six key predictors: width, mean, Hurst exponent, entropy, α_0, and slope. Together, these variables represent the signal's multifractal scaling, complexity, local variability, and temporal trend, providing the LSTM network with a compact but informative description of the BOD dynamics.

A critical challenge in this stage involved optimizing the model's hyperparameters, which define the network's learning capacity and temporal sensitivity [18]. Specifically, we tuned the lookback window, which determines the 'temporal memory' or the number of previous time steps the model analyses to capture the inertial dynamics of the river system. We also optimized the number of model units, representing the hidden neurons that facilitate the learning of complex, non-linear mappings between multifractal descriptors and pollution levels. To control the learning process, the batch size (the number of samples processed before internal weight updates) and the number of epochs (full passes through the training set) were carefully calibrated to ensure convergence without overfitting.

To determine the most robust model configuration, we employed a hyperparameter grid search integrated with time-series cross-validation (a specialized k-fold method for sequential data). We implemented an exhaustive grid search to optimize the neural network, testing combinations of model units (16, 24, 32), batch sizes (8, 16, 24), epochs (100, 150), and lookback windows (2, 3, 4). This rigorous evaluation of 54 distinct configurations allowed for the selection of an optimal model structure. The process was designed to balance high-precision forecasting with robust performance across different stages of the water quality time series.

By utilizing the RMSE metric to minimize error across these chronological splits, the optimal hyperparameter set was selected to produce the final forecasting output. The optimal LSTM configuration for the 100-point forecasting horizon of the Dnipro River BOD time series was identified through a minimized RMSE of 0.8617. This top-performing model utilized a lookback window of 2, a batch size of 8, 100 training epochs, and 32 hidden units. Accordingly, the network processed two consecutive feature vectors and generated a single prediction for the following BOD observation. The LSTM layer maintained internal cell and hidden states and used input, forget, and output gates to regulate the retention and updating of temporal information. Its 32 hidden units, together with the hyperbolic-tangent activation, enabled the model to learn nonlinear relationships between the selected multifractal descriptors and BOD levels. The resulting predictive performance is depicted in **Fig. 4**, where the orange forecast line is plotted against the observed data.

To evaluate the predictive efficiency of the hybrid MFA-LSTM model against traditional Auto Regressive Integrated Moving Average (ARIMA) and Auto Regressive Fractionally Integrated Moving Average (ARFIMA) benchmarks, we utilized independent test datasets comprising 10-point (short-range) and 100-point (long-range) horizons. These data points were strictly excluded from the training phase to ensure an unbiased assessment of model generalization. The comparative performance was quantified using the RMSE metric,

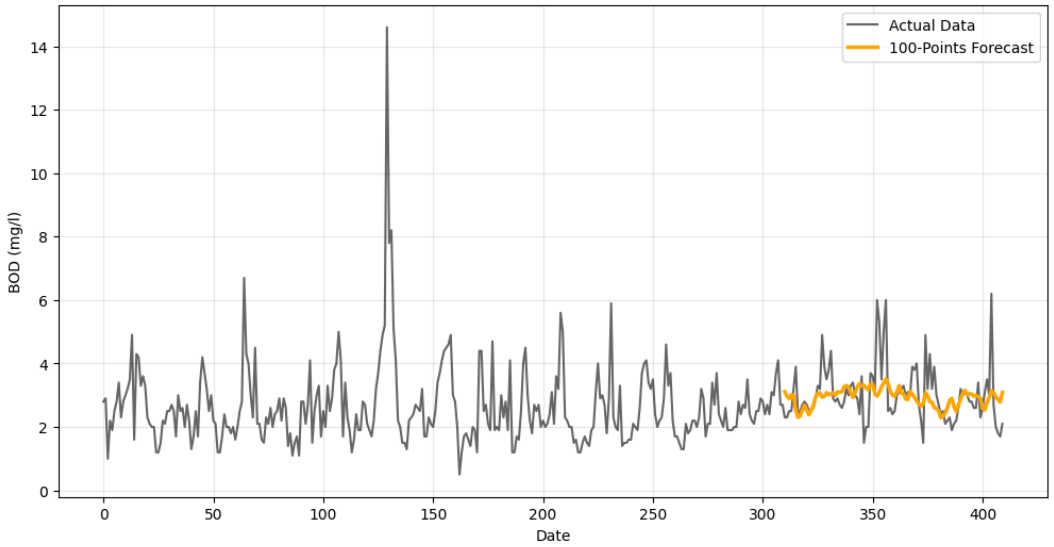


Fig. 4. LSTM forecasting (orange line) for the next 100 observations of BOD time series

allowing for a rigorous evaluation of each model's accuracy across different forecasting scales. The results of this analysis is presented in **Table 1**.

Table 1. **Evaluation of the prediction accuracy using RMSE metric**

Model	10 test data	100 test data
ARFIMA	1.2786	0.9875
ARIMA	1.2782	0.9878
MFA-LSTM	1.1808	0.8617

Based on the comparative RMSE results, the hybrid MFA-LSTM model demonstrates the highest predictive efficiency among the evaluated architectures for both short-term and long-term forecasting horizons. However, this superior accuracy is associated with increased computational complexity, requiring significant time and CPU resources for high-dimensional feature engineering and extensive hyperparameter optimization. Computational efficiency was evaluated by measuring training times on an Intel(R) Core(TM) i7-8550U processor. While the optimization of the MFA-LSTM architecture necessitated a 13-minute execution window to evaluate the complete hyperparameter grid, the ARIMA and ARFIMA benchmarks were processed nearly immediately. These results underscore the substantial overhead associated with high-dimensional feature engineering and neural network tuning compared to traditional statistical modeling. Furthermore, given that different training runs can yield models with varying internal characteristics, the 'black-box' nature of the LSTM must be acknowledged, necessitating careful interpretation of the model's decision-making process.

CONCLUSION

This research presents a predictive framework for BOD dynamics that integrates LSTM neural networks with multifractal analysis.

The methodology encompasses an assessment of stationarity, long-memory, and multifractal properties using the MFDFA algorithm, followed by a custom fractional-exponential approximation of the resulting singularity spectrum. By combining these

multifractal measures with statistical features within a sliding window approach, a high-dimensional feature vector was constructed for the LSTM model. Feature selection was optimized using a Random Forest Regressor, while model hyperparameters were refined through grid-search cross-validation.

The hybrid MFA-LSTM framework demonstrated superior predictive performance compared to all tested models, consistently achieving the lowest RMSE for the Dnipro River station. Specifically, the model yielded an RMSE of 1.1808 for a 10-step-ahead forecast and improved to 0.8617 for the 100-step-ahead horizon. While this performance advantage required greater computational time and CPU resources due to the intensive hyperparameter optimization process, the results validate the model's capacity to capture the complex, multi-scale dynamics of organic water pollution. Consequently, this methodology provides a robust analytical tool for the early detection and proactive management of river pollution events.

ACKNOWLEDGMENTS AND FUNDING SOURCES

The authors received no financial support for the research, writing, and/or publication of this article.

COMPLIANCE WITH ETHICAL STANDARDS

The authors declare that the research was conducted in the absence of any conflict of interest.

AUTHOR CONTRIBUTIONS

Conceptualization, [M.B.]; methodology, [M.B., Y.S.]; validation, [Y.S.]; writing – original draft preparation, [M.B.]; writing – review and editing [M.B.]; supervision, [Y.S.].

All authors have read and agreed to the published version of the manuscript.

REFERENCES

[1] Donat-P. Häder, Anastazia T. Banaszak, Virginia E. Villafañe, Maite A. Narvarte, Raúl A. González, E. Walter Helbling. (2020) Anthropogenic pollution of aquatic ecosystems: Emerging problems with global implications. *Science of The Total Environment*, *713*, 136586. https://doi.org/10.1016/j.scitotenv.2020.136586

[2] Hu, J., Han, G., & Zhang, Q. (2025). Impacts of Environmental Change and Human Activities on Aquatic Ecosystems. *Water*, *17*(11), 1669. https://doi.org/10.3390/w17111669

[3] Zakharova, S.O, Arkhipova, K.K. (2018) Problemy Zabrudnennia Vodnykh Resursiv. Ekolohichnyi Aspekt. *World Science*, *8*(36), Vol.2, 58-62. (In Ukrainian) https://doi.org/10.31435/rsglobal_ws/30082018/6068

[4] Oberdorff, T. (2022) Time for decisive actions to protect freshwater ecosystems from global changes. *Knowl. Manag. Aquat. Ecosyst*., *423*, 19. https://doi.org/10.1051/kmae/2022017

[5] Horoshkova L., Menshov, O. I., Korniichuk, Y. D., Horoshkov, S. V., & Maslov, D. V. (2025). Ecological assessment and forecasting of surface water conditions in the Dnipro river in Zaporizhzhia. *Man and Environment. Issues of Neoecology*, *43*, 47-62. https://doi.org/10.26565/1992-4224-2025-43-04

[6] Jouanneau, Sulivan & Recoules, Loïc & Durand, Marie & Boukabache, A & Picot, V & Primault, Y & Lakel, Abdel & Sengelin, M & Barillon, Bruno & Thouand, Gerald. (2013). Methods for assessing biochemical oxygen demand (BOD): A review. *Water research,* 49C, 62-82. https://doi.org/10.1016/j.watres.2013.10.066

[7] Pattnaik, B. S., Pattanayak, A. S., Udgata, S. K., & Panda, A. K. (2021). Machine learning based soft sensor model for BOD estimation using intelligence at edge. *Complex & Intelligent Systems*, *7*(2), 961-976. https://doi.org/10.1007/s40747-020-00259-9
[8] Ching, P. M., So, R. H., & Morck, T. (2021). Advances in soft sensors for wastewater treatment plants: A systematic review. *Journal of Water Process Engineering*, *44*, 102367. https://doi.org/10.1016/j.jwpe.2021.102367
[9] Liu, Y., Li, K., & Huang, D. (2021, December). Prediction of Biochemical Oxygen Demand Based on VIP-PSO-Elman Model in Wastewater Treatment. In *Proceedings of the 2021 ACM International Conference on Intelligent Computing and its Emerging Applications* (pp. 182-187). https://doi.org/10.1145/3491396.3506502
[10] Liang, Y., Ding, F., Liu, L., Yin, F., Hao, M., Kang, T., Zhao, C., Wang, Z., & Jiang, D. (2025). Monitoring water quality parameters in urban rivers using multi-source data and machine learning approach. *Journal of Hydrology*, *648*, 132394. https://doi.org/10.1016/j.jhydrol.2024.132394
[11] Bordun, M., & Mokrytska, O. (2024). Development of software and algorithmic equipment for prediction of river water pollution using fractal analysis methods. *Comput. Des. Syst. Theory Pract*, *6*(1), 178-188. https://doi.org/10.23939/cds2024.01.178
[12] Bordun, M., & Salapak, O. (2025). Mathematical models for the analysis and forecasting of river water pollution using the multifractal method. *Comput. Des. Syst. Theory Pract.*, *7*(1), 310-318. https://doi.org/10.23939/cds2025.01.310
[13] Yan, X., Cheng, H., Chi, S., Liu, S., & Zhu, Z. (2025). Spatiotemporal modeling of dissolved oxygen in a semi-enclosed water body with a LSTM-GRU hybrid approach. *Processes*, *13*(10), 3325. https://doi.org/10.3390/pr13103325
[14] State surface water monitoring data (n.d.) *State Agency of Water Resources of Ukraine.* (In Ukrainian). https://data.gov.ua/dataset/surface-water-monitoring
[15] Safitri, D., Mustafid, Ispriyanti, D., & Sugito. (2019, May). Gold price modeling in Indonesia using ARFIMA method. In *Journal of Physics: Conference Series* (Vol. 1217, No. 1, p. 012087). IOP Publishing. https://doi.org/10.1088/1742-6596/1217/1/012087
[16] Zhang, X., Zhang, G., Qiu, L., Zhang, B., Sun, Y., Gui, Z., & Zhang, Q. (2019). A modified multifractal detrended fluctuation analysis (MFDFA) approach for multifractal analysis of precipitation in Dongting Lake basin, China. *Water*, *11*(5), 891. https://doi.org/10.3390/w11050891
[17] He, M., Wu, S., Huang, B., Kang, C., & Gui, F. (2022). Prediction of total nitrogen and phosphorus in surface water by deep learning methods based on multi-scale feature extraction. *Water*, *14*(10), 1643. https://doi.org/10.3390/w14101643
[18] Chen, F., Sha, Y., Ji, H., Peng, K., & Liang, X. (2025). Integrating multifractal features into machine learning for improved prediction. *Fractal and Fractional*, *9*(4), 205. https://doi.org/10.3390/fractalfract9040205

АНАЛІЗ ТА ПРОГНОЗУВАННЯ ДИНАМІКИ ЗАБРУДНЕННЯ РІЧКОВИХ ВОД ЗА ДОПОМОГОЮ НЕЙРОННИХ МЕРЕЖ ДОВГОЇ КОРОТКОСТРОКОВОЇ ПАМ'ЯТІ ТА МУЛЬТИФРАКТАЛЬНОГО АНАЛІЗУ

***Михайло Бордун*[1], *Ярослав Соколовський*[2]**
[1] *Кафедра інформаційних систем, Національний університет ім. Івана Франка, вул. Університетська, 1, Львів, 79007, УКРАЇНА*
[2] *Кафедра систем автоматизованого проектування, Національний університет «Львівська політехніка», вул. Степана Бандери, 12, м. Львів, 79013, УКРАЇНА*

АНОТАЦІЯ

Вступ. Органічне забруднення води характеризується складною, нелінійною, нестаціонарною та самоподібною динамікою, яку традиційні моделі прогнозування часто не відтворюють. Метою дослідження є розроблення автоматизованої гібридної системи, що поєднує мультифрактальний аналіз і нейронне навчання для підвищення точності прогнозування біохімічного споживання кисню (БСК) та раннього виявлення екологічних ризиків.

Матеріали та методи. Ретроспективний часовий ряд БСК із річкового пункту моніторингу обробляли за допомогою автоматизованого конвеєра підготовки даних і моделювання. Ковзні вікна використовували для виділення локальних ознак і врахування довготривалих кореляцій. Метод мультифрактального детрендованого флуктуаційного аналізу (МФДФА, англ. MFDFA) застосовували для оцінювання узагальнених показників Херста та спектрів сингулярності в кожному вікні. За допомогою дробово-експоненційної апроксимації неперервні спектри перетворювали на дискретні коефіцієнти зі збереженням їхніх структурних характеристик. Мультифрактальні дескриптори поєднували зі статистичними показниками. Мережу довготривалої короткочасної пам’яті (ДКЧП, англ. LSTM) навчали за допомогою перехресної перевірки часових рядів та пошуку гіперпараметрів. Результати порівнювали з еталонними показниками авторегресивного інтегрованого ковзного середнього (АРІКС, англ. ARIMA – AutoRegressive Integrated Moving Average) та авторегресивного дробово інтегрованого ковзного середнього (АРДІКС, англ. ARFIMA – AutoRegressive Fractionally Integrated Moving Average).

Результати. Гібридна система МФДФА-ДКЧП забезпечила точніші й стабільніші прогнози, ніж базові моделі, для коротких і тривалих горизонтів. Дробово-експоненційні дескриптори надали інформацію про багатомасштабні флуктуації, персистентність, ширину спектра та асиметрію, що дало змогу ДКЧП ефективніше описувати нелінійну динаміку забруднення. Оптимізація гіперпараметрів збалансувала часову пам’ять, складність моделі та узагальнювальну здатність, зменшивши середньоквадратичну похибку.

Висновки. Дослідження підтверджує ефективність автоматизованої системи прогнозування якості води. Її інноваційність полягає у перетворенні неперервних мультифрактальних спектрів на інформативні дискретні ознаки за допомогою дробово-експоненційної апроксимації. У поєднанні з ДКЧП цей підхід підвищує надійність прогнозів і забезпечує природоохоронні органи інструментом раннього попередження та підтримки рішень для проактивного управління якістю води.

Ключові слова: ДКЧП, БСК, мультифрактальний аналіз, МФДФА, АРДІКС, спектр сингулярності.

Received / Одержано	Revised / Доопрацьовано	Accepted / Прийнято	Published / Опубліковано
04 September, 2026	15 September, 2026	21 September, 2026	28 September, 2026