Generative artificial intelligence applying an adaptive algorithm with real-time dynamics allocation for ecological monitoring

Przemyslaw Hawro
State University of Applied Sciences in Jaroslaw,
Faculty of Technical Engineering
Jaroslaw, Poland *przemyslaw.hawro@pansjar.edu.pl*

Tadeusz Kwater
State University of Applied Sciences in Jaroslaw,
Faculty of Technical Engineering
Jaroslaw, Poland *tadeusz.kwater@pansjar.edu.pl*

Boguslaw Twarog
University of Rzeszow, Institute of Computer Science
Rzeszow, Poland *btwarog@ur.edu.pl*

Jacek Bartman
University of Rzeszow, Institute of Computer Science
Rzeszow, Poland *jbartman@ur.edu.pl*

Abstract

This article presents an innovative approach to monitoring river water quality in real time by generating estimates of difficult-to-measure signals such as biochemical oxygen demand. Laboratory tests take too long for real-time monitoring. Therefore, an adaptive PDALM algorithm (Proportional Differential Algorithm with a Latch Mechanism) was developed, integrating mathematical modelling with measurement data to enable instantaneous estimation of water quality signals using a special latch mechanism. The forced eigenvalue distribution guarantees system dynamics and ensures stability and robustness to disturbances. In the proposed RTMS system, the PDALM algorithm functions as an adaptive soft sensor generating high-quality training data. This data is then used by a generative neural network for anomaly detection and forecasting of atypical scenarios in dynamic environmental systems. The system can function as an intelligent environmental monitoring module capable of learning, predicting, and responding to changing environmental conditions.

Keywords: water quality monitoring, BOD estimation, latch mechanism algorithm, data-driven generative models, eigenvalue-based tuning.

1. Introduction

Water quality is critical for environmental health, biodiversity, and public safety. Key indicators, such as biochemical oxygen demand (BOD) and dissolved oxygen (DO), are commonly used to assess water conditions. However, traditional methods, such as laboratory-based BOD_5 analysis, require several days, limiting their use in real-time monitoring [1]. This delay hinders timely responses to pollution events. To overcome this limitation, researchers have developed estimation techniques that rely on mathematical models and artificial intelligence. Ecological systems can be described by ordinary differential equations that capture dynamic interactions between oxygen levels and organic matter degradation. Efficient signal estimation algorithms are therefore required for real-time monitoring, capable of handling disturbances and incomplete data. In recent years, artificial intelligence, including generative AI, has opened new possibilities for forecasting and anomaly detection in such dynamic systems [2].

1.1 Overview of current solutions in the water quality monitoring process

Due to the challenges of direct BOD measurement, many approaches rely on indirect estimation using soft sensors and surrogate variables (e.g. pH, COD, redox potential) [6], [9], [11]. These methods fall into two main categories: model-based and data-driven.

Model-based approaches use known physicochemical relationships, such as the Streeter-Phelps model, to describe oxygen depletion and recovery in rivers. These models can be enhanced with tools such as Kalman filters that integrate sensor data with state estimators to provide optimal predictions. Extended and unscented Kalman filters, as well as particle filters, are often used for non-linear systems [2], [12]. Hybrid observers, such as the combination of Elman networks with unscented filters, have shown high accuracy in estimating difficult parameters such as BOD and nitrogen [8]. Kalman filters combined with LSTM networks and attention mechanisms have also achieved excellent results DO estimation ($R^2 \approx 0.94$) [1, 2], [15].

Data-driven methods rely on machine learning to infer patterns from data. Artificial neural networks (ANN) are widely used to estimate water quality based on accessible inputs such as TOC, UV absorbance, nitrate levels and temperature [11], [16]. Models such as extreme learning machines (ELMs), echo state networks (ESNs), and stochastic neural networks (SNNRW) have been applied to predict BOD, with improvements in both speed and accuracy [3, 4], [8].

Hybrid solutions, which integrate model- and data-driven approaches, are gaining popularity. For example, Kalman-LSTM models benefit from physical interpretability while improving learning generalisation [4, 5]. These approaches are often implemented on IoT platforms, allowing data collection and estimation in real-time [5, 6]. Such systems mark a shift toward proactive water quality management [1], [7], [12].

1.2 A model of a polluted river described by the Streeter-Phelps equation

The Streeter-Phelps model describes the dynamics of dissolved oxygen (DO) and biochemical oxygen demand (BOD) in rivers [10], [13]. The dynamics of oxygen in a river can be characterised by ordinary differential equations that determine the level of oxygen and the degradation of organic pollutant as a result of the reaeration and deoxygenation processes. By developing a description of these phenomena, we can write the equation in the following form:

$$\frac{d}{dt}DO = K_r(DO_s - DO) - K_d BOD \tag{1}$$

where:
DO - oxygen concentration in water $[mg/l]$,
DO_s - oxygen saturation under equilibrium conditions,
K_r - reaeration factor (restoration of oxygen in water),
K_d - coefficient of oxygen consumption by organic matter degradation,
BOD - biochemical oxygen demand.

The Streeter-Phelps model describes how, after wastewater (e.g., from a wastewater treatment plant) is discharged downstream, DO drops rapidly and reaches a minimum (the so-called critical zone - the greatest oxygen deficit) and then gradually recovers through reaeration (a self-purification process).

The eigenvalues of the system of these equations make it possible to determine the stability of water quality and predict the critical points at which oxygen deficiency may occur. The system of Streeter-Phelps equations, taking into account the influence of stochastic disturbances $w_1,\ w_2$, can be written in the matrix form:

$$\frac{d}{dt}\begin{bmatrix} BOD \\ DO \end{bmatrix} = \begin{bmatrix} -K_d & 0 \\ K_d & -K_r \end{bmatrix}\begin{bmatrix} BOD \\ DO \end{bmatrix} + \begin{bmatrix} 0 \\ K_r DO_s \end{bmatrix} + \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \tag{2}$$

$$A_M = \begin{bmatrix} -K_d & 0 \\ K_d & -K_r \end{bmatrix} \tag{3}$$

Eigenvalues of the object matrix:

$$\lambda_1 = -K_r, \lambda_2 = -K_d \tag{4}$$

Interpretation of eigenvalues:

- If $\lambda_1, \lambda_2 < 0$, the system is stable, and the river has the ability to clean itself.
- If $\lambda_1 > 0$, excessive oxygenation (e.g., eutrophication) may occur.
- If $\lambda_2 > 0$, it means no decomposition of organic matter and ecosystem degradation.

The key objective of the research is to monitor water quality in real time and apply predictive algorithms to predict changes in its parameters, taking into account the analysis of the eigenvalues of the dynamic system. The authors of the article, after reviewing modern methods for estimating BOD and DO in real time, proposed the solution of the author with special attention to a new algorithm based on the location of eigenvalues. This algorithm is a training dataset generator for the RTMS (Real-Time Monitoring System) using generative artificial intelligence. The principles of its operation, the quality of the generated results, and the potential for integration with generative artificial intelligence techniques are discussed.

2 RTMS application of Proportional-Differential Algorithm with a Latch Mechanism (PDALM).

The objective of the study was to develop and test an adaptive real-time system to estimate signals in a river model described by nonlinear differential equations. The RTMS is made up of three main modules:

1. Data Acquisition Module (AM): Collects real-time sensor data (DO, flow, precipitation, anthropogenic factors).
2. Training Data Generation Module (PDALM): Processes current signals and enforces dynamic constraints using a latch mechanism.
3. Prediction & Optimisation Module (LSTM): Applies AI to predict oxygen deficits and classify degradation states.

The PDALM algorithm estimates BOD in real time using DO measurements and a mathematical model subjected to simulated disturbances. Fig. 1 illustrates the structure of RTMS and how PDALM generates learning data for the ANN.

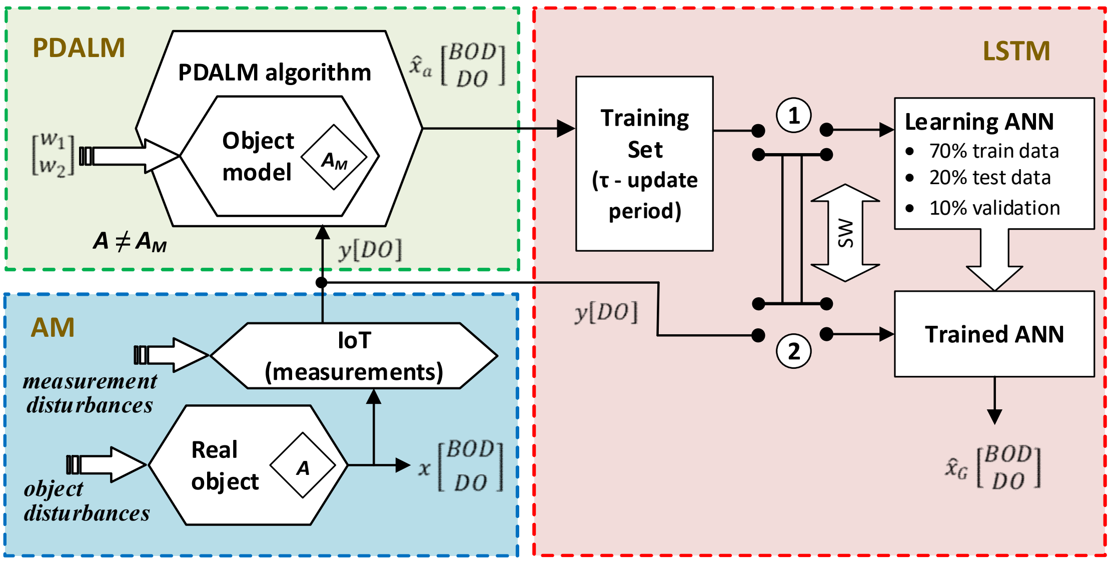

Fig. 1 Schematic of the RTMS system using the PDALM algorithm

The PDALM estimator assumes that the real system dynamics differ from the model ($A \neq A_M$). It generates a time-dependent estimate $\hat{x}_a\begin{bmatrix}BOD\\DO\end{bmatrix}$, updated over interval τ. Once trained (SW=1), the ANN switches to prediction mode (SW=2), estimating the final output $\hat{x}_G\begin{bmatrix}BOD\\DO\end{bmatrix}$ based on the DO inputs $y[DO]$.

At its core, PDALM adapts filter gains to ensure that the system's eigenvalues remain in a predefined stable region (left half-plane), controlled by parameters α (oscillation) and η (response rate). When eigenvalues enter this region for the first time, the algorithm latches on to the gains and uses them as a reference to reject future destabilising updates. Fig. 2 shows the allowed eigenvalue region.

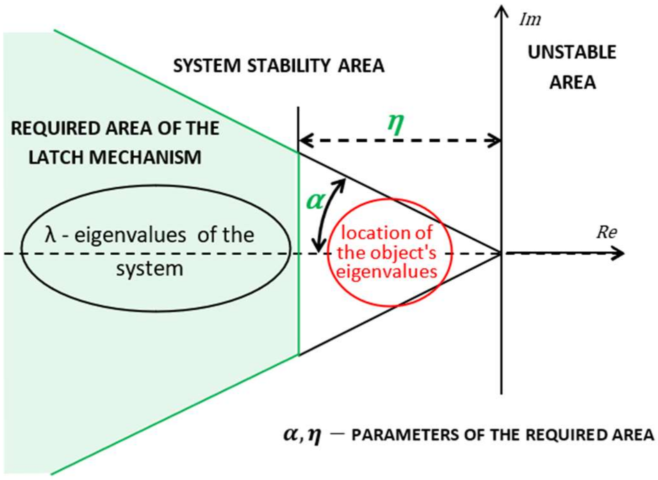

Fig. 2 Area with the required location of the monitoring system eigenvalues

Unlike classic filters, PDALM actively shapes the system dynamics by modifying the gain values ΔK based on the adaptation error and its derivative. This ensures a fast and stable response, even under severe perturbations, without requiring knowledge of noise intensity.

The following is the simplified pseudocode of the algorithm:

1. Initialisation:
   initial values: state estimate $\hat{x}_0$, gain K_0 , and auxiliary flags $first\_hit = false$ and parameters α, η
2. // measurement and calculation loop
for each time step i:
- measurement y_i //defined as $y_i = Cx_i + v_{pi}$
- generation of state estimation $\hat{x}_{i+1}$
- calculation of adaptation error $\varepsilon_i = y_i - C\hat{x}_i$
- calculation of gain correction $\Delta K_i = \xi_i(k_p\varepsilon_i + T_d\frac{d_{\varepsilon_i}}{dt})$ taking into account changes in the direction of error $\xi_i \in \{1, 0, -1\}$
- update reinforcement $K_{i+1} \leftarrow K_i + \Delta K_i$
3. // eigenvalue analysis (latch mechanism)
- calculation of $\lambda_i(real\ and\ imag.parts) \leftarrow eig.\ of\ (A - K_i\ C)$
  - $if\ \lambda_i \in required\_area(\alpha, \eta)$
    - current K_i is saved as "good set" and $first\_hit = true$
  - $if\ (first\_hit)$
    - $\lambda_{i+1}(real\ and\ imag.parts) \leftarrow eig.of\ (A - K_iC)$
    - $if\ (\lambda_{i+1} \notin required\_area(\alpha, \eta))$
      - restoring previous values K_i "good set".

where:
A - coefficient matrix describing dynamics of the river's natural self-purification process;
$C = [0, 1]$ - measurement matrix indicating the measured DO signal;
k_p - proportionality vector;
T_d - vector of time constant.

Once latched, the system tolerates only safe gain adjustments, preserving the desired dynamic behaviour. The algorithm is robust, self-stabilising, and adaptive to disturbances such as pollution surges or changes in flow. It was validated through simulations using a virtual river system as a reference. Importantly, PDALM does not require knowledge of the intensity of any disturbances. The use of a differentiating member allows the algorithm to react more quickly to changes in the signal trend, for example, to anticipate a drop in oxygen before it fully develops, based on the rate of change of the BOD. As a result, even when the system is suddenly perturbed (a spike in pollutant load, a change in lateral inflow conditions), the algorithm maintains stability and forced dynamics.

3 Pre-processing of learning data with the PDALM algorithm

Preprocessing of the learning data for the neural network was carried out with the PDALM algorithm using the mathematical model of the object, in particular, taking into account the variability of the object's excitations and parameters in the algorithm. Fig. 3 shows selected time waveforms of the object and algorithm signals for different values of system and measurement disturbances and the distribution of eigenvalues.

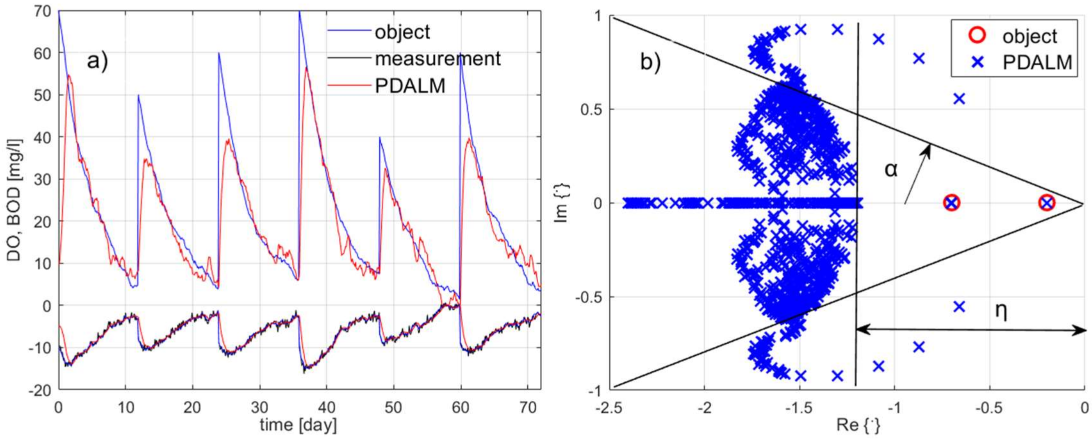

Fig. 3 Time waveforms of the BOD and DO signals (a), distributions of eigenvalues of the object, and obtained by the PDALM algorithm (b)

The waveforms in Fig. 3a confirm the correct approximation of the BOD signals, for which real-mode measurement is impossible, and the measured DO signal. Despite the lack of access to measurement and process noise models or any information about them, and despite sudden changes in the signal, the PDALM algorithm achieved high estimation precision (Tab 1). Fig. 3b shows the successive locations of the eigenvalues of the RTMS system during the operation of the algorithm. A small number of eigenvalues outside the required area occur only during the initial phase of the algorithm's operation.

The BOD and DO signals generated by the PDALM algorithm were evaluated with the quality indicators of the RMSE and MAE monitoring adopted, and the results are summarised in Table 1.

Tab. 1 RMSE and MAE quality indicators of the PDALM algorithm with parameter changes k_p and T_d

	RMSE		MAE	
PDALM parameters	BOD	DO	BOD	DO
$k_p = [-0.4; 0.06]$ $T_d = [0.055; -0.006]$	9.98	1.36	4.26	0.65
$k_p = [-0.6; 0.09]$ $T_d = [0.065; -0.0085]$	9.92	1.33	4.4	0.64
$k_p = [-0.8; 0.12]$ $T_d = [0.095; -0.01]$	9.28	1.22	4.04	0.58
Parameters: $\alpha = 22^o, \eta = 1.4, W = [3, -2; -2, 1], V = [0.1]$				

For BOD, both adopted quality indicators take higher values than for the measured DO signal. Adopting larger absolute values of the parameters k_p and T_d from among those presented in Table 1 results in the best results confirmed by the RMSE and MAE indices.

In the simulations, it was noted that the algorithm compensated for changes in noise by correcting the gains, maintaining stability, and the required speed of response. Thus, it can be concluded that the algorithm with eigenvalue localisation has a certain robustness - it can adapt to changing conditions, providing good quality estimation even where the classical filter becomes suboptimal. In conclusion, the new PDALM algorithm proved to be effective and accurate, especially when monitoring an unmeasured parameter such as BOD.

4 Integration of the PDALM algorithm with generative artificial intelligence

The increasing availability of generative AI techniques, such as GANs, diffusion models, or transformers, opens new opportunities to monitor dynamic systems such as water environments [12], [14]. These models can learn the distribution of normal system behaviour (e.g., daily DO/BOD patterns) and identify anomalies by comparing real data with synthetic references [4].

For example, the GAN-based system described in [8] successfully detected pollution incidents in water networks by comparing sensor readings with generated baselines, improving detection accuracy and minimising false alarms. Similarly, [14] used GANs to model pressure profiles and detect leaks in distribution systems with around 70% effectiveness. In water quality applications, such generative models can simulate pollution scenarios (e.g., rainfall, industrial discharge), which may serve as test cases for adaptive algorithms like PDALM. Coupling PDALM with digital twins enhanced by generative AI allows robust, scenario-based training even for events not yet observed [15, 16]. In our implementation, an artificial neural network (ANN) was trained to detect adverse water quality events using DO and BOD as inputs. The network architecture includes the following:

- An input layer (DO, BOD),
- A hidden layer with 10 ReLU neurones,
- A softmax layer for multiclass classification,
- A binary output (normal vs. anomaly) using a sigmoid function,
- The Adam optimiser for training.

The network classifies the data into five classes: normal condition (1), moderate pollution (2), discharge of municipal or industrial wastewater (3), intensive inflow of pollution (4), and failure of the aeration system (5) (Fig. 4).

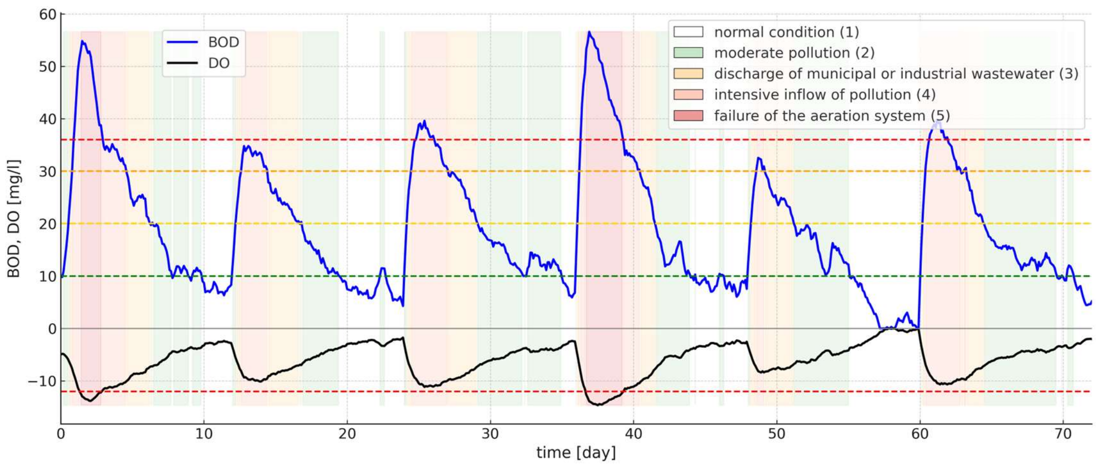

Fig. 4 Classified pollution anomaly by their areas in RTMS system

A known issue in such classification tasks is data imbalance, and some classes (e.g., rare events) are under-represented. Without class weighting, the model tends to favour dominant classes. To address this, training was repeated using class weights, which improved recognition of less frequent but critical anomalies (Fig. 5).

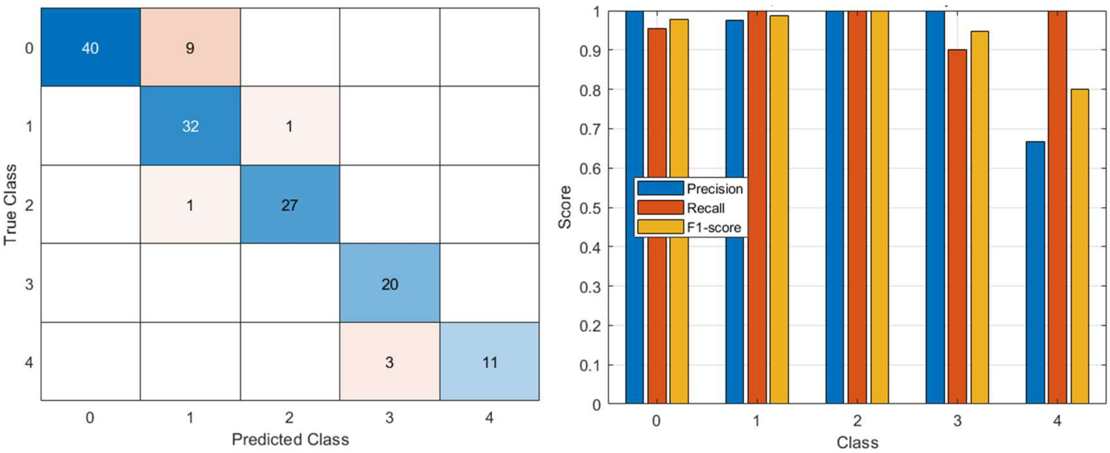

Fig. 5 Confusion matrix for object state recognition by the RTMS system for configurations with class weights (a - left) and Precision, Recall and F1-score per Class (b - right)

To evaluate the effectiveness of the anomaly detection module in the RTMS system, a neural network classifier was evaluated using standard performance metrics: precision, recall and F1-score, calculated for each class. The classification task involved five environmental conditions: from normal state to severe pollution or aeration system failure. The model was trained with class weights to compensate for the imbalance in the dataset, particularly the small number of examples for rare but critical classes. Figure 5a presents the confusion matrix, showing that the classifier achieved high accuracy in all classes. Notably:

- Normal conditions (Class 0) were recognised with high precision (1.0) and recall (0.95),
- Moderate pollution (Class 1) was detected with an F1-score of 0.99,
- Wastewater discharge (Class 2) achieved perfect classification (F1 = 1.0),
- Intensive pollutant runoff (Class 3) and aeration failure (Class 4) were also effectively identified, with F1-scores of 0.95 and 0.8, respectively.

The corresponding F1-scores per class are visualized in Fig. 5b. While the rarest class (Class 4) showed lower precision (0.67), the model maintained full recall (1.0), indicating that critical anomalies were not overlooked even if a small number of false positives occurred. The overall weighted F1-score reached 0.98, confirming that the class-weighted approach significantly improves model performance in realistic monitoring scenarios. This is particularly important for early detection of dangerous or rare environmental events, such as infrastructure failures or unreported pollution discharges, where missed detections could have serious consequences.

Beyond classification, generative models can support visual interpretation by producing synthetic images or risk maps for decision making [2]. For example, they can simulate the effects of industrial discharge on DO/BOD dynamics, aiding scenario analysis, and operational planning. The PDALM algorithm, through eigenvalue-driven adaptation and stability control, acts as a soft sensor that generates high-fidelity synthetic training data (e.g., estimated BOD signal), enabling the learning process even in the absence of full measurement coverage. This generative mechanism ensures data diversity, supports model generalisation, and mimics real-world signal variability.

In summary, integrating generative AI with adaptive algorithms like PDALM enables the development of intelligent monitoring systems that not only estimate but also simulate, predict, and adapt to complex environmental events, thus acting as self-learning, proactive surveillance tools.

5 Summary and Conclusions

This paper presents an innovative approach to real-time water quality monitoring (RTMS) that combines mathematical modelling of dynamic systems with the capabilities of generative artificial intelligence. Central to the solution is the proprietary PDALM algorithm, which enables immediate estimation of hard-to-measure parameters, such as

BOD, eliminating the need for time-consuming laboratory analysis. A key feature of PDALM is its latch mechanism, which adaptively adjusts estimator parameters while ensuring system stability through enforced eigenvalue localisation. This allows the system to actively shape its dynamic response and to remain robust under disturbances.

The integrated RTMS, supported by generative AI, includes modules for prediction, anomaly detection, and generation of realistic simulation scenarios. The use of an artificial neural network allows classification of environmental conditions, including rare but critical events such as industrial discharges or aeration failures. Introducing class weights during training significantly improved the recognition of these under-represented cases. Simulations demonstrated high estimation accuracy and robustness of the system under varying measurement and model conditions. PDALM effectively compensated for disturbances by adapting dynamic parameters, making it suitable for real-world scenarios with high uncertainty.

The results confirm that the combination of model-based estimation with adaptive generative AI offers a promising path toward intelligent environmental monitoring. This hybrid approach improves system accuracy, resilience, and adaptability, laying the foundation for self-learning architectures capable of predicting complex, previously unobserved phenomena. Future work includes integrating digital twin concepts and advanced predictive models, such as diffusion models and temporal transformers, to further enhance monitoring and decision support. The convergence of AI and adaptive algorithms marks a key direction in environmental technology that enables systems that are not only reactive but proactive and robust in the face of real-world variability. Therefore, the generative aspect of our system lies in the hybrid synergy between model-based synthetic data generation and planned integration of GenAI models, enabling accurate classification and proactive, simulation-based reasoning.

References

1. Bagheri, M., Bagheri, K., Farshforoush, N., Velazquez, A., Liu, Y.: A novel hybrid deep learning model for real-time monitoring of water pollution using sensor data. Journal of Water Process Engineering. 68 106595 (2024)
2. Cai, H., Zhang, C., Xu, J., Wang, F., Xiao, L., Huang, S., Zhang, Y.: Water Quality Prediction Based on the KF-LSTM Encoder-Decoder Network: A Case Study with Missing Data Collection. Water (Basel). 15 (14), 2542 (2023)
3. Chhipi-Shrestha, G., Mian, H.R., Mohammadiun, S., Rodriguez, M., Hewage, K., Sadiq, R.: Digital water: artificial intelligence and soft computing applications for drinking water quality assessment. Clean Technol Environ Policy. 25 (5), 1409–1438 (2023)
4. El-Shafeiy, E., Alsabaan, M., Ibrahem, M.I., Elwahsh, H.: Real-Time Anomaly Detection for Water Quality Sensor Monitoring Based on Multivariate Deep Learning Technique. Sensors. 23 (20), 8613 (2023)
5. Essamlali, I., Nhaila, H., El Khaili, M.: Advances in machine learning and IoT for water quality monitoring: A comprehensive review. Heliyon. 10 (6), e27920 (2024)
6. Islam, M.B.E., Faheem, Y., Ahmad, A., Fraz, M.M.: From Measured pH to Hidden BOD: Quasi Real-Time Estimation of Key Indirect Water Quality Parameters Through Direct Sensor Measurements. (2024)
7. Kwater, T., Hawro, P., Krutys, P., Gołębiowski, M., Dralus, G.: Comprehensive online estimation of object signals for a control system with an adaptive approach and incomplete measurements. Bulletin of the Polish Academy of Sciences Technical Sciences. 150111–150111 (2024)
8. Li, Z., Liu, H., Zhang, C., Fu, G.: Generative adversarial networks for detecting contamination events in water distribution systems using multi-parameter, multi-site water quality monitoring. Environmental Science and Ecotechnology. 14 100231 (2023)
9. Liu, Y., Yuan, L., Li, D., Li, Y., Huang, D.: Process Monitoring of Quality-Related Variables in Wastewater Treatment Using Kalman-Elman Neural Network-Based

Soft-Sensor Modeling. Water (Basel). 13 (24), 3659 (2021)
10. Lung, W.: Progression of river BOD/DO modeling for water quality management. Water Environment Research. 95 (4), (2023)
11. Pattnaik, B.S., Pattanayak, A.S., Udgata, S.K., Panda, A.K.: Machine learning based soft sensor model for BOD estimation using intelligence at edge. Complex & Intelligent Systems. 7 (2), 961–976 (2021)
12. Quaghebeur, W., Torfs, E., De Baets, B., Nopens, I.: Hybrid differential equations: Integrating mechanistic and data-driven techniques for modelling of water systems. Water Res. 213 118166 (2022)
13. Streeter, H.W., Phelps, E.B.: A Study of the Pollution and Natural Purification of the Ohio River. US Public Health Service Bulletin. (No.146), (1925)
14. Taiwo, R., Yussif, A.-M., Zayed, T.: Making waves: Generative artificial intelligence in water distribution networks: Opportunities and challenges. Water Res X. 28 100316 (2025)
15. Yu, P., Cao, J., Jegatheesan, V., Du, X.: A Real-time BOD Estimation Method in Wastewater Treatment Process Based on an Optimized Extreme Learning Machine. Applied Sciences. 9 (3), 523 (2019)
16. Zhu, M., Wang, J., Yang, X., Zhang, Y., Zhang, L., Ren, H., Wu, B., Ye, L.: A review of the application of machine learning in water quality evaluation. Eco-Environment & Health. 1 (2), 107–116 (2022)