



UWL REPOSITORY

repository.uwl.ac.uk

Towards sustainable WWTPs: explainable modeling and scenario analysis of
N₂O emissions

Sadaghat, Behnam, Behzadian, Kourosh ORCID logo ORCID: <https://orcid.org/0000-0002-1459-8408>
and Najjar-Ghabel, Saeid (2025) Towards sustainable WWTPs: explainable modeling and scenario
analysis of N₂O emissions. Journal of Water Process Engineering, 79. ISSN 2214-7144

<https://doi.org/10.1016/j.jwpe.2025.108818>

This is the Published Version of the final output.

UWL repository link: <https://repository.uwl.ac.uk/id/eprint/15453/>

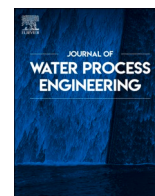
Alternative formats: If you require this document in an alternative format, please contact:
open.research@uwl.ac.uk

Copyright: Creative Commons: Attribution 4.0

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy: If you believe that this document breaches copyright, please contact us at open.research@uwl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Rights Retention Statement:



Towards sustainable WWTPs: Explainable modeling and scenario analysis of N₂O emissions

Behnam Sadaghat^a, Kourosh Behzadian^b, Saeid Najjar-Ghabel^{b,*}

^a Department of Civil and Water Engineering, University of Tabriz, Tabriz, Iran

^b School of Computing and Engineering, University of West London, St Mary's Rd, London, W5 5RF, United Kingdom

ARTICLE INFO

Editor: Soroush Abolfathi

Keywords:

Nitrous oxide
Water treatment plants
Regression methods
Sensitivity techniques
Prediction

ABSTRACT

Accurate prediction of nitrous oxide (N₂O) emissions from wastewater treatment plants (WWTPs) has been hindered by the dynamic, non-linear nature of the processes involved and by a lack of uncertainty treatment and scenario analysis in past studies. While machine learning (ML) has been shown to improve accuracy, limitations have been found in standardizing preprocessing, dual optimizer usage, post-hoc interpretability, and decision-focused validation. In this study, N₂O emission rates were predicted using Elastic Net (EN), Quantile Regression (QR), Gaussian Process Regression (GPR), and Gradient Boosting Regression (GBR). Data preprocessing was conducted by daily averaging of WWTP N₂O emission rate, Z-score filtering, VIF-based multicollinearity check and feature selection, and normalization. The models were evaluated in a 5-fold cross-validation with five iterations. Hyperparameter optimization was performed using the Puma Optimizer (PO) and Bayesian Optimization (BO). Post-hoc sensitivity (as here specifically defined as feature-attribution interpretability) was achieved using SHAP (global effects) and LIME (local effects). Uncertainty was approximated using a Monte Carlo procedure (noisy inputs, 1000 iterations).

Numerically, baseline GPR reached R^2 (test) = 0.958; RMSE (test) = 0.020; MAE (test) = 0.012. With optimization, GPR + BO (GPBO) achieved the highest performance with R^2 (test) = 0.986, RMSE (test) = 0.012, and MAE (test) = 0.007, while GPR + PUMA (GPPO) performed as high as R^2 (test) = 0.980; RMSE (test) = 0.014; MAE (test) = 0.009. SHAP and LIME analysis further established dissolved N₂O as the most significant predictor, with conditional contributions from ammonium (NH₄⁺), dissolved oxygen (DO), nitrate (NO₃⁻), and temperature. Scenario analysis with CGAN (N₂O + 20 %; NH₄⁺ – 15 % with +5 °C; and multi-variate perturbations) was conducted with the top model to demonstrate decision-oriented responsiveness.

The study contribution is an end-to-end, decision-ready pipeline that combines standardized preprocessing, dual optimizers, SHAP and LIME interpretability, Monte Carlo uncertainty, and CGAN-based what-if testing, capabilities that have been limited or scattered in prior work. Relevance is demonstrated by actionable environmental and economic co-benefits: peak N₂O abatement potential (minimized GHG footprint, alignment with IPCC-style reporting) and aeration energy and chemical dosing cost minimization via predictive control, thereby facilitating sustainable and low-cost WWTP operation.

Abbreviations: N₂O, nitrous oxide; NH₄⁺, ammonium; DO, dissolved oxygen; NO₃⁻, nitrate; WWTPs, wastewater treatment plants; ML, machine learning; EN, Elastic Net; QR, Quantile Regression; GPR, Gaussian Process Regression; GBR, Gradient Boosting Regression; PO, Puma Optimizer; BO, Bayesian Optimization; XAI, Explainable Artificial Intelligence; SHAP, Shapley Additive explanation; LIME, Local Interpretable Model-Agnostic Explanations; CGAN, Conditional Generative Adversarial Network; R², coefficient of determination (RMSE), (MAE), (RSE), (MAPE), and (SMAPE); RMSE, root mean square error; RSE, relative squared error; MAE, mean absolute error; SMAPE, symmetric mean absolute percentage error; MAPE, mean absolute percentage error.

* Corresponding author.

E-mail address: saeid.najjarghabel@uwl.ac.uk (S. Najjar-Ghabel).

<https://doi.org/10.1016/j.jwpe.2025.108818>

Received 13 May 2025; Received in revised form 15 September 2025; Accepted 22 September 2025

Available online 9 October 2025

2214-7144/© 2025 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

1.1. Study background

N₂O is one of the most powerful greenhouse gases with severe impacts on global warming and atmospheric chemistry [1]. While CO₂ and CH₄ have been intensely recognized as the major forces of global change, N₂O possesses nearly 300-fold global warming potential of CO₂ in comparison on average in a 100-year time frame [2]. N₂O is also the single most important ozone-depleting substance that is currently emitted by anthropogenic activities, and therefore is one of the control measures of priority. Due to the long atmospheric lifetime and increasing anthropogenic emissions, efforts towards monitoring and limiting N₂O are crucial both from the viewpoint of global change mitigation and ozone layer preservation [3].

In Europe, urban water management supplies 1–3 % of total electricity consumption and 3–10 % of global warming potential [4]. Wastewater treatment plants (WWTPs) emit fossil CO₂ from energy usage and biogenic CO₂, CH₄, and N₂O resulting from biological processes, the latter two having global warming potentials 30 and 300 times higher than CO₂, respectively [5]. As the global shift towards more sustainable wastewater practices, N₂O emissions from WWTPs are increasingly recognized as a major aspect that needs to be understood and addressed. The Intergovernmental Panel on Climate Change (IPCC) and the U.S. Environmental Protection Agency (EPA) recognize the need for effective monitoring and mitigation strategies to limit the environmental impact of such emissions [6].

In WWTPs, N₂O is produced primarily during nitrification and denitrification, two of the most important microbial processes involved in nitrogen removal. Nitrification is the aerobic process by which ammonium (NH₄⁺) is converted to nitrate (NO₃⁻) with nitrite (NO₂⁻) as an intermediate product [7]. It is catalyzed by two groups of microbes, namely, ammonia-oxidizing bacteria (AOB) and nitrite-oxidizing bacteria (NOB) [8]. Under good oxygen conditions, nitrification is conducted with minimal production of N₂O. AOB, however, emit N₂O as a byproduct under low or interrupted oxygen environments due to incomplete ammonia oxidation. Accumulation of nitrite and competition with oxygen also lead to N₂O production, and hence N₂O is a major issue in wastewater treatment operations [9].

Denitrification is an anaerobic process during which NO₃⁻ is reduced by facultative bacteria to nitrogen gas (N₂) [9]. Stepwise reduction reduces NO₃⁻ in sequence to NO₂⁻, nitric oxide (NO), N₂O, and N₂ [10]. Under ideal conditions, the process is operated to yield harmless nitrogen gas, but under low availability of carbon or shifting environmental conditions, denitrification is incomplete and N₂O accumulates and is vented [11]. Dissolved oxygen (DO) concentration, carbon: nitrogen ratio, pH, and temperature are significant parameters that are involved in determining the extent of production of N₂O. Since WWTPs operate with maximum nitrogen removal efficiency, balancing these parameters is essential in order to avoid undesired N₂O emissions [7].

One of the greatest challenges in mitigating N₂O emission from WWTPs is the variability in emission rates that is inherent in the process. N₂O production is dynamic and varies with plant-specific factors such as influent characteristics, treatment process configuration, and operating conditions [12]. Unlike the relative homogeneity and ease of quantifiability of CO₂ emissions, N₂O emissions are highly variable in space and time. These preclude the formulation of general emission factors or prediction models with general applicability across treatment facilities. Common approach involves mechanistic models like the Activated Sludge Model No. 1 (ASM1) and its extensions that attempt to simulate nitrogen transformation in accordance with biochemical reactions [13]. These require heavy model calibration, are highly sensitive to site-specific parameters, and are unable to account for N₂O emissions variability [14,15].

Machine learning (ML) is one of the promising methods to improve N₂O prediction modeling in WWTPs. Compared with mechanistic

models that use prespecified functions to describe nitrogen removal processes, ML algorithms can capture nonlinear and complex relationships from data [16]. The use of historical process data allows ML-based models to discover hidden patterns and make more accurate predictions of N₂O emissions under different conditions [17]. ML methods such as feature importance analysis, Shapley Additive Explanations (SHAP), and permutation importance also allow researchers to determine the most important parameters driving emissions. With this information, operation strategies are improved to reduce nitrogen pollution and greenhouse gas emissions.

Unlike process-based models, which require exact kinetic parameters and thorough calibration, regression-based ML models can be trained using conventionally gathered operational data. This approach avoids reducing intricate biological and chemical interactions to oversimplified assumptions [18]. With further improvements in data acquisition capacity and computational power, ML-based methods will play increasingly crucial roles in the sustainable operation of wastewater treatment [19,20].

1.2. Literature review

The advent of ML has significantly advanced environmental modeling, particularly in the realm of wastewater treatment, where complex, dynamic, and non-linear processes often defy traditional mechanistic modeling. ML techniques offer flexibility and adaptability in modeling diverse biological and chemical phenomena such as N₂O emissions, which are influenced by fluctuating operational and environmental parameters. These approaches allow for the integration of real-time data, historical trends, and variable interactions to produce accurate, interpretable, and scalable predictive models. Recent research reflects a growing interest in leveraging ML not only for enhanced prediction but also for optimization, anomaly detection, and process control in WWTPs.

Khalil et al. [21] investigated the predictive modeling of N₂O emissions using ensemble learning methods, including Random Forest and XGB, optimized via Bayesian techniques. Their study emphasized the trade-off between model complexity and interpretability, using SHAP analysis to elucidate the influence of input features. The models achieved high predictive performance while maintaining transparency, making them suitable for regulatory and operational applications. More recently, Khalil et al. [22] suggested a multi-objective hybrid feature selection and hyperparameter optimization using NSGA-II for N₂O prediction in WWTPs, trading off model accuracy and complexity. Applied to AdaBoost, it reduced estimators and tree depth without sacrificing high accuracy ($R^2 = 0.94$) while maintaining the RMSE rise at bay, improving generalizability and online emission monitoring efficiency.

In a related study, Xu et al. [23] applied long short-term memory (LSTM) networks to forecast N₂O emissions in a municipal WWTP. The temporal capabilities of LSTM allowed the model to capture intricate time-dependent patterns, yielding a coefficient of determination (R^2) as high as 0.92, thus proving its efficacy in dynamic emission modeling. Seshan et al. [24] applied LSTM-based encoder–decoder models to a full-scale year of WWTP data to forecast N₂O emissions 6 h ahead. The optimum 256–256 LSTM model obtained R^2 values between 0.59 and 0.98, accuracy decreasing with increasing horizons. Dominant predictors were lagged N₂O emissions, influent flow, NH₄⁺, NO_x, and aerobic tank dissolved oxygen. Short-term forecasting was emphasized, and hybrid deep learning–biokinetic models were suggested for robust predictions over longer horizons. Xu et al. [25] employed LSTM models in a Swiss municipal WWTP and contrasted them with the common RNNs. The best LSTM model outperformed RNNs for MAE, RMSE, and R^2 and exhibited better robustness to extreme N₂O fluctuations, which confirms LSTM's excellent performance for precise N₂O prediction across heterogeneous WWTPs. Moreover, Huang et al. [26] employed symmetry-informed deep learning models, including a PLO-CNN-BiLSTM-Attention model, to predict Altenrhein WWTP N₂O emissions. The model yielded an R^2 of

0.99 and demonstrated good generalization performance across 48 validation sets. SHAP-based feature analysis identified water temperature, dissolved oxygen, and inlet flow rate as the dominant drivers with intriguing seasonal symmetry patterns, indicating the promise of explainable and stable deep learning models for operational N_2O emission management. Yun et al. [27] utilized an explainable XGB model for optimizing TN removal in a full-scale municipal WWTP with ecological ponds. The model demonstrated outstanding predictive accuracy, and SHAP analyses revealed key operating factors for optimal TN minimization, energy consumption, and COD dosage. A GUI-accessible implementation enabled dynamic control, resulting in a substantial reduction in effluent and carbon emissions, reflecting the potential of ML for the sustainable operation of WWTP.

Beyond emissions prediction, ML techniques were also successfully used for other key water quality parameters, e.g., total nitrogen (TN) and phosphorus removal, dissolved oxygen (DO). Ensemble models, e.g., Gradient Boosting, and Random Forest have predicted DO up to R^2 of 0.987 [28], whereas deep learning models GRU and LSTM have predicted DO in WWTPs up to R^2 of 0.7 [29]. PSO-based optimization-driven ML approaches combining PSO with predictive models have been revealed to increase accuracy for DO dynamic control [30], while Bayesian optimization has enhanced TN and energy consumption prediction [31]. Explainers like XGB and CatBoost have been employed successfully for multi-parameter saline wastewater treatment processes with high accuracy [32].

Moreover, ML has also been employed for optimizing resource use and understanding chemical interactions. Feng et al. [33] utilized a two-layer feed-forward neural network (TLFFNN) and a voting ensemble to predict N_2O capture in designable ionic liquids. The TLFFNN model demonstrated exceptional performance with an MSE of 0.00002 and R^2 of 0.9981, indicating its robustness in modeling molecular interactions. In Greece, Karadimos and Anthopoulos [34] developed an ANN model using FANN tools and WEKA to estimate energy consumption in WWTPs. Their model, driven by input variables such as influent load and treatment capacity, achieved high accuracy (MSE = $8.99\text{e-}5$), proving its utility in energy planning and cost reduction.

Energy optimization was also explored by Alali et al. [35], who compared 23 ML models to predict energy consumption in WWTPs. They incorporated time-lagged variables to enhance model performance and identified the dynamic K-nearest neighbors model as superior to state-of-the-art alternatives. Piłat-Rożek et al. [36] applied ML in conjunction with sensor technologies to develop an electronic nose system for process classification in WWTPs. By using t-SNE for dimensionality reduction and Random Forest for classification, the model demonstrated perfect accuracy, validating the feasibility of sensor-integrated ML applications.

On the system design front, Singh et al. [37] implemented ML models for monitoring and optimization of biological nitrogen removal processes. Their results indicated significant improvements in system reliability and nutrient removal efficiency through adaptive ML-based control mechanisms. El Fels et al. [38] conducted a comprehensive text mining analysis of ML applications in wastewater treatment and found that ANN, SVM, RF, and gradient boosting were the most frequently adopted models. Genetic algorithms were identified as the leading optimization technique, especially for parameter calibration.

In terms of hybrid and ensemble techniques, Mustafa et al. [39] developed a phytoremediation model using SVM-M1 and ensemble learning for predicting total dissolved solids in a hydroponic system. Their model reached an R^2 of 0.9986 during validation, emphasizing the strength of ensemble approaches in environmental prediction. Adibimanesh et al. [40] applied ML for incineration monitoring in WWTPs, achieving improved control over sludge treatment and reducing energy waste. Lastly, Sahu et al. [41] explored ML applications in enhancing microalgae-bacteria systems for eco-friendly wastewater treatment, offering novel avenues for biological nutrient recovery and N_2O emission mitigation.

Collectively, these studies demonstrate the breadth and depth of ML's potential in wastewater treatment, from accurate emissions forecasting and energy efficiency to real-time monitoring and intelligent system design. The convergence of domain-specific knowledge with data-driven insights continues to redefine best practices in sustainable water management.

1.3. Limitations and research novelties

Existing studies on N_2O emission modeling from wastewater treatment have typically employed process-based models or a single machine learning regressor without paying attention to hyperparameter tuning or uncertainty management. The gap in existing studies has been the lack of multicollinearity filtering, which may result in biased parameter estimates. Similarly, interpretability has typically been restricted to single-method feature importance without offering an overall description of variable interactions. Besides, a few applications of XAI methods, such as SHAP and LIME, have been used to guide the development of what-if scenarios, which limits the scientific underpinning of operational perturbations. As a result, the development of scenario analyses in previous research was rare, and the impact of particular alterations in critical variables (e.g., DO, NH_4^+ , NO_3^- , temperature) on N_2O emission was not examined. Besides, past studies have scarcely incorporated uncertainty quantification or scenario-based analysis, making their applicability poor for operational decision-making.

The present study introduces some novelties. Firstly, a comprehensive preprocessing pipeline was employed that involved merging Z-score removal of outliers, VIF-based feature selection, and normalization for stability in the model. Secondly, two different hyperparameter optimization methods (PO and BO) were employed and compared systematically, with convergence traces being reported, a part largely lacking in the literature. Third, dual interpretability techniques based on sensitivity analysis for determining imperative features (SHAP for global understanding and LIME for local explanation) were blended to provide a more expert analysis of emissions. Fourth, Monte Carlo-based uncertainty analysis was used to estimate the confidence of the prediction, considering variability in the real world. Lastly, scenario analysis using CGAN was undertaken to explore likely perturbations of influent and process conditions, extending the horizon from prediction quality to decision-focused environmental and operating planning.

2. N_2O emission rate prediction dataset

2.1. Dataset compilation and characteristics

The data used to construct and calibrate prediction models for N_2O emission rates were compiled from open-source operational records of the Avedøre Wastewater Treatment Plant (Copenhagen, Denmark) [42]. The facility treats approximately $70,000 \text{ m}^3/\text{day}$ in dry-weather flow conditions and serves approximately 350,000 PE. The mainstream process includes screening, grit removal, primary settling, secondary clarification, and four parallel Carrousel reactors with two compartments having alternate aerobic/anoxic conditions under STAR Control® based upon in-situ NH_4^+ and NO_3^- levels. Periodic measurements of the important process parameters were made in both compartments of all the reactors at predetermined hydraulic locations, and liquid-phase N_2O was measured only in Reactors 1 and 3 [43]. For the purposes of this study, daily averages were derived for the two compartments of Reactor 1 and Reactor 3 between March 2018 and February 2019, providing a total of 1234 samples.

Seven process-level parameters representing the main physico-chemical drivers of nitrogen transformation and N_2O production in wastewater treatment plants are included in every record:

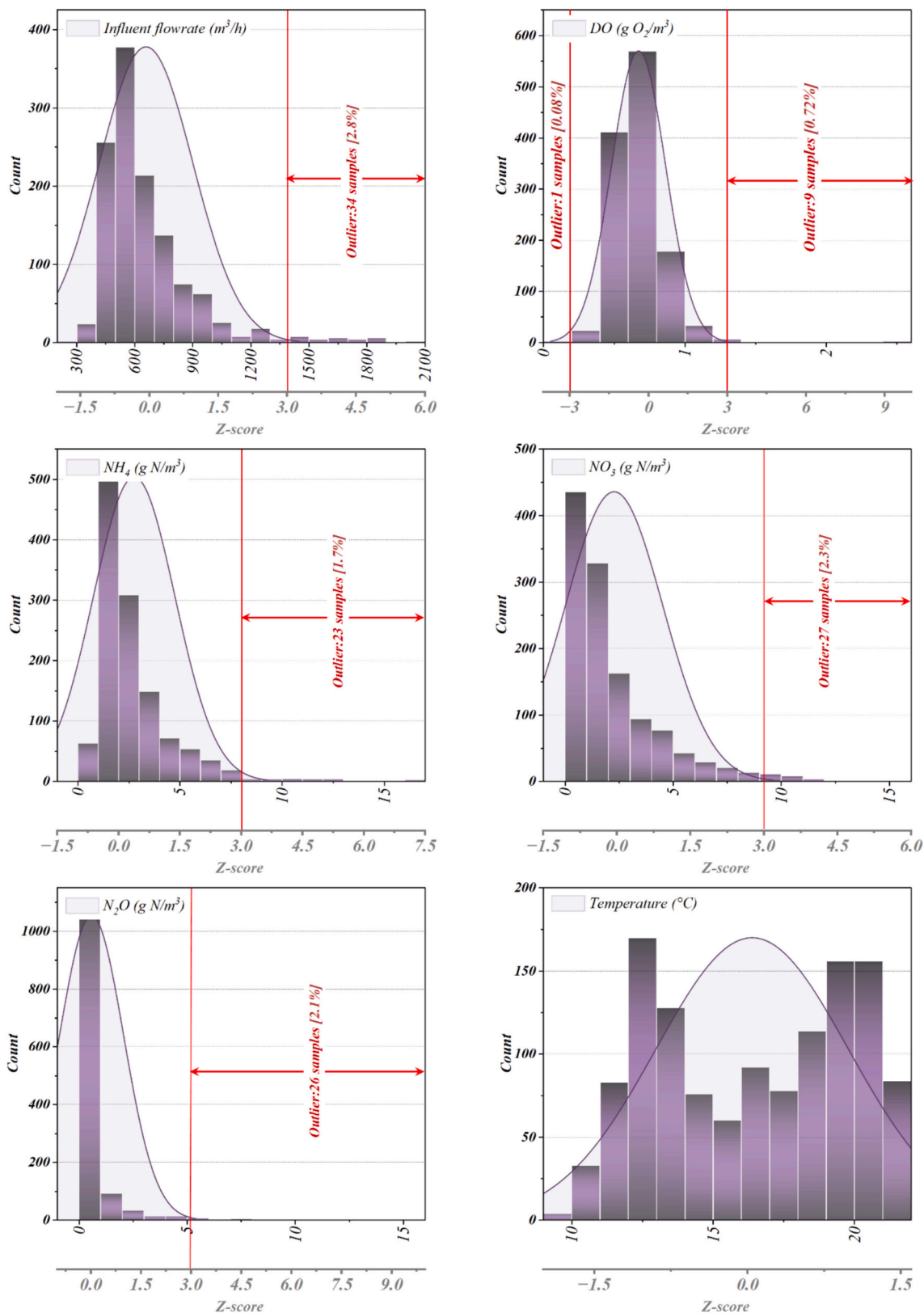


Fig. 1. Histogram and Z-score distribution for variables in the dataset.

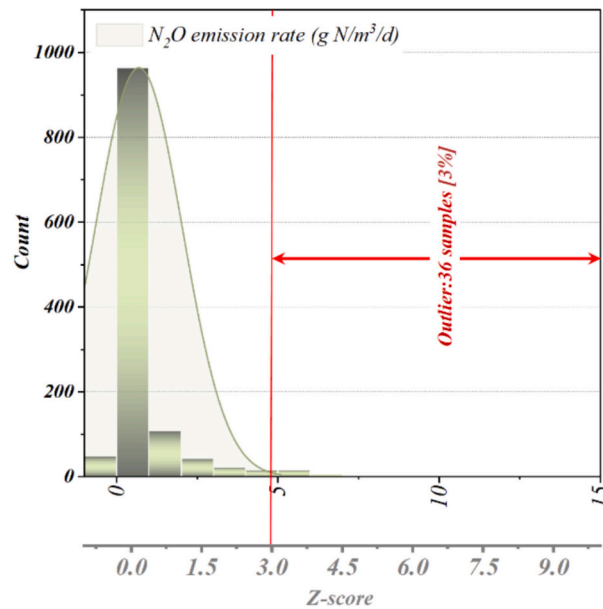


Fig. 1. (continued).

- Influent flowrate (m^3/h): The hydraulic loading rate moving into the reactor(s) under observation, which directly affects retention time and biological reaction kinetics.
- Dissolved oxygen (DO , $\text{g O}_2/\text{m}^3$): The concentration of oxygen dissolved in mixed liquor, necessary for aerobic nitrification and indirectly influencing denitrification rates when aeration is interrupted.
- Ammonium (NH_4^+ , $\text{g N}/\text{m}^3$): The principal nitrogen compound of influent wastewater, the substrate for nitrifying microorganisms.
- Nitrate (NO_3^- , $\text{g N}/\text{m}^3$): Main product of nitrification and electron acceptor for denitrifying microorganisms.
- Dissolved N_2O ($\text{g N}/\text{m}^3$): Liquid-phase intermediate nitrous oxide before emission, reflecting incomplete nitrification–denitrification reactions.
- Temperature ($^\circ\text{C}$): Regulates microbial metabolism, enzyme kinetics, and nitrogen conversion rates, with seasonal fluctuations influencing process efficiency.
- N_2O emission rate ($\text{g N}/\text{m}^3/\text{d}$): The research's focus variable, the rate at which nitrous oxide is emitted per unit time under various operational and environmental conditions.

2.2. Data preprocessing

In the development of prediction models, data preprocessing, including the treatment of missing values and outliers, is crucial for ensuring the high quality of data for model training. No missing values were detected in the compiled dataset for this study.

To treat outliers and clean a dataset, one popular approach is the use of the Z-score method, which is a measure of how far apart each observation is from the series mean in standard deviation units. The outliers in a real-world dataset can skew the results of observations, making the mean no longer representative of the observations. The Z-score of each sample in the dataset (input and target variables) is calculated as:

$$Z - \text{Score} = \frac{x - \mu}{\sigma} \quad (1)$$

Here, μ and σ are the mean and standard deviation of the series, respectively [44]. A Z-score equal to 0 indicates that the observation is the same as the mean, while positive or negative values signal deviations above or below the mean. Observations with Z-scores higher than a chosen threshold, here ± 3 , are considered outliers. According to the Z-

score presentation in Fig.1, 134 samples were removed from the dataset with outliers in variable values, and the cleaned dataset with 1100 samples was prepared for the following steps of preprocessing and feature selection.

In this study, input features were scaled to a common range (normalized) to prevent features with large ranges from overpowering the model training. It improves the stability of algorithms, accelerates optimization, and enables models to better learn the connection between features and outputs. Min-max scaling is a widely used method that preserves original data relationships when normalizing feature ranges [45].

2.3. Multicollinearity check and feature selection

Multicollinearity refers to the correlation among independent variables in a regression model that can influence the accuracy of estimated coefficients. Among the popular techniques for detecting multicollinearity is the Variance Inflation Factor (VIF), which calculates the magnitude of variance inflation of the estimated regression coefficients. The VIF for the i – th predictor is:

$$\text{VIF} = \frac{1}{1 - R_i^2} \quad (2)$$

Here R^2 is the coefficient of determination. A VIF greater than the commonly used threshold of 10 indicates an objectionable degree of multicollinearity that must be corrected by removing the features [46]. According to the VIF score reports in Table 1, all values are lower than 10 for variables in the utilized dataset, so all variables should be selected as predictors for models.

2.4. Correlation matrix extraction

Fig. 2 illustrates the correlation matrix among key physicochemical

Table 1
VIF scores report for multicollinearity check.

Variable	Values	Variable	Values
Influent flowrate (m^3/h)	1.622	NO_3 ($\text{g N}/\text{m}^3$)	2.873
DO ($\text{g O}_2/\text{m}^3$)	1.590	N_2O ($\text{g N}/\text{m}^3$)	1.031
NH_4 ($\text{g N}/\text{m}^3$)	2.437	Temperature ($^\circ\text{C}$)	1.466

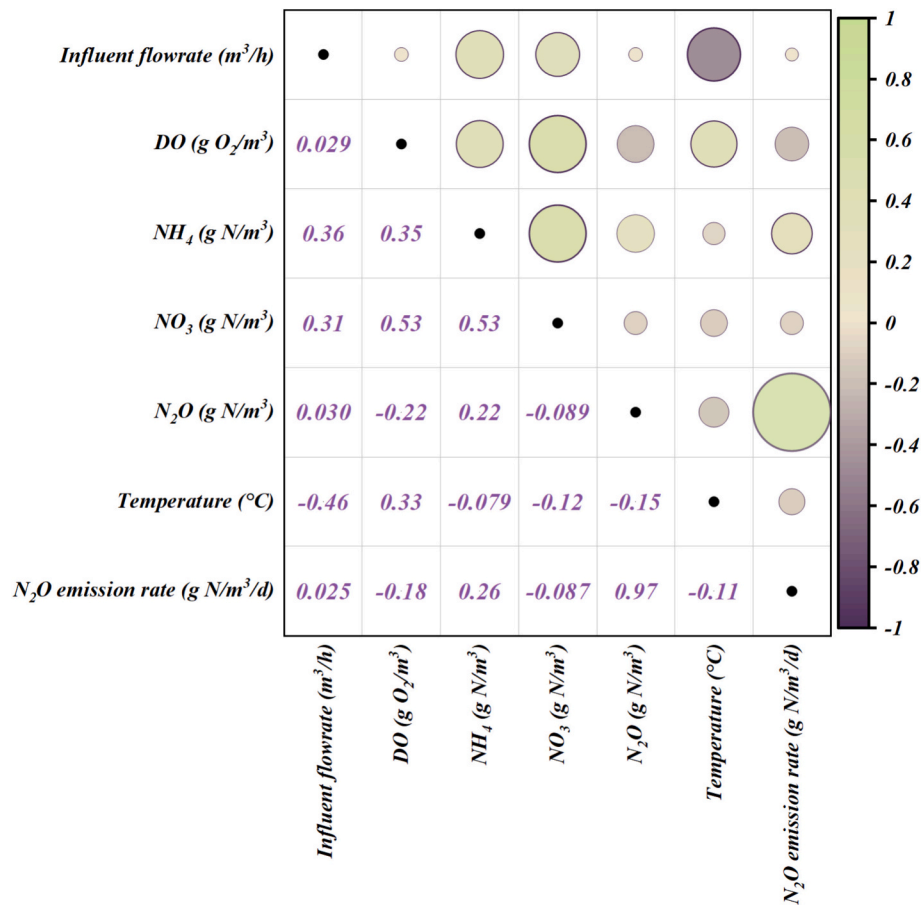


Fig. 2. Correlation plot of variables depicted based on normalized data.

parameters influencing N_2O emissions in the wastewater treatment process. The analysis is based on normalized data, and the strength of linear relationships is depicted using Pearson correlation coefficients.

DO exhibits moderate positive correlations with both NH_4 ($r = 0.35$) and NO_3 , $r = 0.53$, indicating that elevated oxygen levels may be associated with higher concentrations of nitrogen species, potentially reflecting operational or environmental conditions that favor aerobic nitrification stages. A strong linear relationship is observed between N_2O and N_2O emission rate ($r = 0.97$), which is expected due to their intrinsic linkage. However, all other correlations with N_2O , whether concentration or emission rate, are relatively weak. For example, the emission rate correlates modestly with NH_4 ($r = 0.26$) and weakly with NO_3 ($r = -0.087$) and temperature ($r = -0.11$), suggesting that N_2O dynamics may be driven more by complex interactions and microbial kinetics than by direct linear associations with single variables. Temperature, overall, shows weak correlations with the other variables, with the most notable being a moderate negative correlation with influent flowrate ($r = -0.46$).

3. Methodology

3.1. Machine learning base prediction models

Among the prediction models utilized in this study, Elastic Net is the combination of L1 and L2 penalties. It is thus appropriate for feature selection in high-dimensional data sets with multicollinearity. In contrast, Quantile Regression extends mean prediction to estimate different conditional quantiles with the advantage of robustness to outliers and distributional skewness. On the other hand, advanced machine learning methods such as Gaussian Process Regression (GPR) and

Gradient Boosting Regression (GBR) are precisely designed for modeling complicated, nonlinear relationships. GPR is a Bayesian non-parametric model that provides predictive distributions and estimates of uncertainty, which are highly useful in small-sample domains but computationally costly for large datasets. GBR is a strong ensemble learning algorithm that constructs additive models with sequential decision trees, which have good accuracy and scalability but incur a loss of interpretability.

3.1.1. Quantile regression (QR)

QR, introduced by Koenker and Bassett in 1978, extends the concept of estimating conditional means in linear regression to estimating conditional quantiles of a response variable. Instead of focusing solely on the average outcome, quantile regression allows the analysis of how predictors influence different points (quantiles) of the outcome distribution, such as the median or the 10th and 90th percentiles.

This method is particularly useful when the impact of explanatory variables is not uniform across the distribution of the dependent variable. For example, an independent variable might have a stronger effect on low-income individuals than on high-income ones, a distinction that would be obscured by traditional mean regression.

Quantile regression provides a more comprehensive view of the relationship between variables by modeling how covariates influence the shape, scale, and location of the outcome distribution. It is especially valuable when the data contain outliers or are heteroscedastic, as quantile regression is more robust to outliers in the dependent variable compared to ordinary least squares regression.

In this framework, each quantile is associated with its own set of regression coefficients, allowing a more nuanced analysis of the conditional distribution. However, interpreting these coefficients requires

caution, as a change in a covariate may not imply that individuals remain in the same quantile before and after the change [47].

3.1.2. Gradient boosting regression (GBR)

GB is an ML technique commonly used for both regression and classification tasks. It builds predictive models by combining multiple weak learners, typically decision trees, into a single, stronger model. The method has been successfully applied in areas such as regression analysis and energy theft detection, gaining prominence in both statistical modeling and artificial intelligence.

GBR specifically focuses on regression problems and follows an adaptive boosting approach. It works by sequentially adding weak regression models, each of which tries to correct the errors of the previous ones. The model optimization is guided by a loss function, which measures how well the model fits the data. Gradient descent is often used to minimize this loss, ensuring that each new learner moves the overall model closer to the optimal prediction.

Like other boosting techniques, GBR builds the final model in a greedy, additive fashion—adding new weak learners one at a time, each improving the model's performance by focusing on the instances that were hardest to predict in previous iterations [48,49].

3.1.3. Gaussian process regression (GPR)

GPR is a non-parametric, probabilistic Bayesian method employed for both regression and classification tasks. One of its key advantages lies in its capacity to detect nuanced, non-linear relationships within data through the use of kernel (covariance) functions that gauge the similarity between points in the input space. It operates under the assumption that the data originates from a multivariate Gaussian distribution, and it leverages Bayesian inference to forecast outputs from given inputs.

Rather than delivering a single-point prediction, GPR yields an entire predictive distribution, including both the mean prediction and the uncertainty (variance), for any specified input. This feature is particularly beneficial in applications where understanding the confidence level of predictions is crucial. A widely used kernel in GPR is the squared exponential, also known as the RBF kernel, which facilitates smooth and adaptable function modeling.

After training, the model estimates the mean and variance of outputs at test points by considering the covariance between training and test inputs. During the training process, kernel hyperparameters such as the length scale and signal variance are typically optimized. However, due to the computational intensity of the method, GPR is most effective with small to medium-sized datasets [50].

3.1.4. Elastic net regression (EN)

EN is a regularization technique specifically engineered for high-dimensional data scenarios, particularly where predictors exhibit high levels of correlation. This method proves invaluable when conventional regression approaches falter due to multicollinearity issues among variables. A significant advantage of EN is its capacity to simultaneously select groups of correlated predictors, instead of arbitrarily choosing a single variable from such groups, which ultimately contributes to constructing models that are both more interpretable and reliable amid strong inter-feature dependencies. Furthermore, EN is capable of handling situations where the number of predictors exceeds the number of observations, making it particularly adept for applications involving ultra-high-dimensional datasets [51].

3.2. Cross-validation

To ensure validation of model performance, five-fold cross-validation with five random runs was employed. That is, the data was split into five equal-sized subsets. For each run, one of the five subsets was employed in the role of the test set (20 %), and the remaining four were used to train the model (80 %). This was repeated five times to enhance

the reliability of the model and reduce variations in outcome measures. Averaging across runs of the performance score gave a better estimate of the generalizability of the model to new data. Not only did the procedure avoid the danger of overfitting the training data, but it also allowed an unbiased evaluation of the model's prediction capacity [52].

3.3. Optimization methods

Bayesian optimization and metaheuristics are widely used to solve complex optimization problems, particularly when objective functions are non-smooth, non-convex, or discontinuous. Bayesian optimization efficiently guides the search for improved parameters using probabilistic surrogate models and acquisition functions, while metaheuristic algorithms evolve inspiration from nature or biological phenomena to explore and exploit the solution space [53,54]. One of the most well-known subclasses of metaheuristics is population-based optimization algorithms, including evolutionary algorithms, genetic algorithms, particle swarm optimization, and evolutionary strategies. The process of initiation and update of a population marks the beginning of each generation in population-based optimization algorithms, followed by the selection of the individuals to ascertain the global optimum. The technique used for the generation and selection of the population varies between different population-based optimization algorithms, and their nature makes parallelization straightforward, as a population of N individuals can be evaluated on N processors simultaneously. In this study, the Bayesian optimization and Puma Optimization, a recent population-based optimization algorithm, were utilized for hyperparameter tuning of base prediction models.

3.3.1. Bayesian optimization

Bayesian optimization is an efficient computational method of hyperparameter tuning that has the ability to discover the best hyperparameters with a small number of evaluations. In model calibration, the significant parameters, i.e., the RBF kernel length scales and the White Kernel noise level, are tuned to optimize the model's log marginal likelihood in order to optimally fit training data. Model performance is quantified by measures such as Root Mean Squared Error (RMSE) and R^2 on training and test sets. Mathematically, the selection of the next evaluation point, x_{next} is represented as [55]:

$$x_{next} = \underset{x}{\operatorname{argmax}} \alpha(x) \cdot \rho(x; \min, \max), \quad (3)$$

Here, $\alpha(x)$ is the acquisition function, controlling the exploration-exploitation trade-off, and $\rho(x; \min, \max)$ is a probability distribution over the input space, typically uniform within bounds. This form ensures that the next input is chosen to make the most informative contribution to the optimization of the objective function, striking a balance between exploration of promising areas and exploitation of uncharted areas.

3.3.2. Puma optimization algorithm

The Puma Optimizer (PO) is a recent addition to the family of metaheuristic techniques, taking its cue from the hunting instincts and social dynamics of pumas. It tackles challenging optimization tasks by employing a flexible "phase-change" strategy, drawn from how pumas switch between roaming for prey and closing in on a target, to alternate adaptively between exploring new regions of the solution space and exploiting the most promising areas it has already discovered. Rather than predefining a fixed ratio of exploration versus exploitation, PO continuously assesses its progress and uses that experience to decide when to search broadly and when to concentrate its efforts, thereby reducing the risk of getting stuck in suboptimal solutions.

In its exploration mode, PO reproduces the puma's tendency to patrol large territories: individual solutions wander widely or exchange information with neighbors to uncover diverse candidate solutions. When it switches to exploitation, it imitates the predator's precision tactics, such as laying an ambush or sprinting after prey, to intensify the

search for the most attractive solutions found so far. Within the population of solutions, each member behaves like a puma, and the current best solution assumes the role of the dominant male, steering the overall search direction. Historical success rates guide the algorithm's adaptive rules, determining which phase to emphasize and how aggressively to refine solutions. PO also incorporates specialized operators tailored to each phase. These include randomized jumps and cooperative moves during exploration to preserve variety, as well as directed adjustments in exploitation to sharpen solution quality. For further details, refer to [56].

PO's success in optimizing ML model performance for future N₂O emissions prediction lies in its adaptive exploration-exploitation trade-off. N₂O emissions data are nonlinear, noisy, and high-dimensional, and model hyperparameters require careful tuning to achieve accurate predictions. PO's phase-change strategy enables thorough exploration of the hyperparameter space while focusing the search effort in the vicinity of the most promising candidates, reducing the risk of local optima. Its population-based method ensures a diverse set of candidate solutions is taken into consideration, and feedback-driven operators optimize hyperparameters according to model performance metrics such as RMSE and R². Together, these characteristics allow PO to effectively discover the optimal set of parameters, thereby enhancing the predictive accuracy and robustness of ML models for N₂O emission prediction.

3.4. Evaluation metrics

To facilitate the comparison of predictive power as well as the potential for generalization of the models developed, a system of statistical measures was employed. The measures are the coefficient of determination (R²), root mean square error (RMSE), mean absolute error (MAE), relative squared error (RSE), mean absolute percentage error (MAPE), and symmetric mean absolute percentage error (SMAPE). In concert, these measures communicate different kinds of information about different aspects of model accuracy, from explained as well as relative and percentage error values. In combination, they allow for a full suite of model quality assessments, enabling more insight into the models' behavior across different sets of circumstances as well as different dimensions of performance.

$$R^2 = \left(\frac{\sum_{j=1}^n (A_j - \bar{A})(P_j - \bar{P})}{\sqrt{\left[\sum_{j=1}^n (A_j - \bar{A})^2 \right] \left[\sum_{j=1}^n (P_j - \bar{P})^2 \right]}} \right)^2 \quad (4)$$

$$RMSE = \sqrt{\frac{\sum_{j=1}^n e_j^2}{n}} \quad (5)$$

$$MAE = \frac{1}{n} \sum_{j=1}^n |e_j| \quad (6)$$

$$RSE = \sum_{j=1}^n \frac{e_j^2}{(A_j - \bar{A})^2} \quad (7)$$

$$MAPE = \frac{1}{n} \sum_{j=1}^n \frac{e_j}{A_j} \quad (8)$$

$$SMAPE = \frac{100}{n} \sum_{j=1}^n \frac{2 \times |e_j|}{|A_j| + |P_j|} \quad (9)$$

Here, A_j represents the actual measured value in example j , and its predicted model value is denoted by P_j . Prediction error e_j (residual, $e_j = A_j - P_j$) tells us how far predicted and observed values are apart. $\bar{A}$ and $\bar{P}$ represent the averages of all observed data ($\bar{A} = \frac{1}{n} \sum_{j=1}^n A_j$ and predicted outputs $\bar{P} = \frac{1}{n} \sum_{j=1}^n P_j$, respectively, providing central references for comparison. The sample size n denotes the total number of

data instances analyzed.

3.5. Sensitivity analysis and explainable AI tools

Sensitivity analysis examines how variations in model inputs or controlling variables dictate outputs and offers a rigorous means of making model behavior interpretable [57–59]. Two methods are used in this work for sensitivity analysis: Local Interpretable Model-agnostic Explanations (LIME) and Shapley Additive explanations (SHAP). SHAP and LIME are popular XAI methods but differ in technique and scope. SHAP, having a game theory approach, provides both global and local explanations by calculating each contribution of a feature to every combination possible, and thus is more complete but resource-intensive. It has the capability to deal with nonlinearities as per the underlying model, though it is poor with collinearities, as well as assuming feature independence. On the other hand, LIME locally approximates a complex model with a simple linear surrogate with efficient, interpretable instance-level explanations. However, it cannot capture nonlinear relationships, is limited by local explanation, and is sensitive to feature correlation [60].

The opacity of most ML models necessitates explainable Artificial Intelligence (XAI) methods to deliver trust and actionable data [61,62]. Counterfactual explanation generation, an XAI technique, enables what-if analysis by tracing how minor variations in input conditions would restore a system to nominal behavior [63]. Contemporary techniques, including the use of Generative Adversarial Networks [64,65], have extended the realism and usability of counterfactuals in multivariate data and temporal domains, providing a practical toolkit for real-time exploration of scenario-based changes in N₂O emission rate of WWTPs and evidence-based decision-making in complex systems.

3.5.1. LIME

LIME explains predictions by locally approximating a complex model with a simple and comprehensible model around a prediction's local neighborhood. Two primary properties of LIME are interpretability, which describes input–output relationships, and local fidelity, which encourages local faithfulness around the instance to the model's behavior. As model-agnostic, LIME can be employed for explaining text, image, and tabular data classifiers and regressors with output presented in textual, numeric, or visual formats [66].

The LIME algorithm provides interpretable explanations of black-box models $f: \mathbb{R}^d \rightarrow \mathbb{R}$ by locally approximating them with an interpretable model $g \in G$. The aim of LIME is to minimize jointly the loss of fidelity between the interpretable model and the black-box model in the local neighborhood around an instance x , and the complexity of the interpretable model [67]:

$$\epsilon(x) = \underset{g \in G}{\operatorname{argmin}} L(f, g, \pi_x) + \Omega(g), \quad (10)$$

Here, $L(f, g, \pi_x)$ estimates the unfaithfulness of g in approximating f locally, $\Omega(g)$ is employed to estimate the complexity of the model, and $\pi_x(z)$ defines the locality around x as:

$$\pi_x(z) = \exp\left(-\frac{D(x, z)^2}{\sigma^2}\right), \quad (11)$$

With $D(x, z)$ as a distance function and σ for the regulation of the width of locality.

The algorithm perturbs the instance x to generate samples $Z' \in \{0, 1\}^d$ representing the presence or absence of interpretable components, which are projected back to the original feature space $z \in \mathbb{R}^d$. The predictions $f(z)$ from the black-box model are computed for these perturbed samples. Each sample is weighted by its proximity to x using $\pi_x(z)$, with greater stress on proximate samples:

$$g(z) = \omega_g \cdot z', \quad (12)$$

Here, ω_g are the interpretable model's learned feature weights. Using these weighted samples, LIME fits a simple model, often a sparse linear model, a local explanation where the feature weights represent their importance to the prediction. In this way, interpretability and local fidelity are assured, and LIME can interpret predictions in text, numeric, or visual forms for tabular, image, or text data.

3.5.2. SHAP

SHAP is an XAI technique that leverages a mathematical foundation to assign each feature a weighting, or Shapley value, of its contribution to a model's prediction [68]. Founded in game theory, SHAP provides a model-agnostic and intrinsically explainable solution capable of explaining any black-box model's predictions by identifying key features and approximating their effect. SHAP is presented here with a simple linear model to forecast a disease. Let F be a set of M features $\{1, 2, \dots, M\}$, S be a subset of F ($S \subseteq F$), and ϕ be an empty set which represents an empty coalition with no features. There are 2^M possible coalitions. A function v gives every coalition a real number for its marginal contribution, with that of the empty coalition being $v(\phi) = 0$.

For every permutation P , the marginal contribution of a coalition S that comes before feature i is calculated first and then the contribution of $S \cup \{i\}$. The Shapley value of feature i can be obtained as:

$$\phi(i) = \frac{1}{|F|!} \sum_P (v(S \cup \{i\}) - v(S)) \quad (13)$$

Here, $|F|$ is the total number of features, $|F|!$ the number of permutations, and $v(S \cup \{i\}) - v(S)$ the marginal contribution of $\{i\}$. For a single coalition S , this is expressed as:

$$|S|! \cdot (|F| - |S| - 1)! (v(S \cup \{i\}) - v(S)) \quad (14)$$

Repeating across all coalitions of $F - \{i\}$ gives the total contribution of feature i across permutations and normalizing over the $|F|!$ possible permutations of F give the Shapley value:

$$\phi(i) = \sum_{S \subseteq F - \{i\}} \frac{|S|! (|F| - |S| - 1)!}{|F|!} (v(S \cup \{i\}) - v(S)) \quad (15)$$

Here, $\phi(i)$ is the average marginal contribution of the feature i to model output over all instances. Doing it for every feature provides an equal distribution of contributions, which can be understood as each feature's effect on a specific prediction.

3.6. Uncertainty analysis

Monte Carlo simulation is a broad computational method applied to estimate system uncertainty by recursively executing a model with randomized inputs. The process begins with the creation of a mathematical model with predetermined input and output variables. They may range from simple analytical equations to sophisticated applications such as HSPF. The uncertain parameters are then specified and given appropriate probability distributions (e.g., normal, uniform). The model is then run repeatedly, sampling the input values from the specified distributions in each iteration. By generating a large ensemble of potential outputs, the Monte Carlo method provides an empirical basis for estimating the variability and confidence interval of the results of the model [69].

4. Results

4.1. Cross-validation results

Cross-validation performance of each model was assessed through 5-fold cross-validation. R^2 and RMSE measures are presented in Tables 2 and 3, respectively, providing information with respect to fold-by-fold variability and predictivity. In addition to particular fold outcomes, mean and standard deviation across folds are measures of the overall

Table 2

5-fold cross-validation results for train/test folds selection concerning R^2 values.

Models	Iteration	Test fold				
		K1	K2	K3	K4	K5
(EN)	1	0.716	0.743	0.703	0.685	0.647
	2	0.612	0.632	0.744	0.701	0.746
	3	0.719	0.711	0.623	0.638	0.660
	4	0.694	0.745	0.696	0.736	0.611
	5	0.717	0.619	0.735	0.862	0.723
	Avg	0.691	0.690	0.700	0.724	0.677
	St. dev	0.041	0.054	0.043	0.076	0.050
(QR)	1	0.712	0.727	0.709	0.735	0.703
	2	0.744	0.761	0.820	0.703	0.730
	3	0.700	0.761	0.713	0.730	0.756
	4	0.706	0.738	0.745	0.708	0.717
	5	0.715	0.713	0.730	0.707	0.736
	Avg	0.715	0.740	0.743	0.717	0.728
	St. dev	0.015	0.019	0.040	0.013	0.018
GPR	1	0.708	0.715	0.643	0.798	0.858
	2	0.879	0.600	0.837	0.698	0.786
	3	0.683	0.864	0.650	0.837	0.862
	4	0.836	0.859	0.948	0.627	0.748
	5	0.712	0.820	0.818	0.736	0.802
	Avg	0.764	0.772	0.779	0.739	0.811
	St. dev	0.078	0.101	0.117	0.074	0.044
GBR	1	0.712	0.515	0.697	0.603	0.662
	2	0.699	0.571	0.674	0.644	0.680
	3	0.477	0.698	0.693	0.590	0.670
	4	0.673	0.690	0.652	0.519	0.672
	5	0.622	0.435	0.647	0.691	0.700
	Avg	0.637	0.582	0.673	0.609	0.677
	St. dev	0.086	0.101	0.021	0.057	0.013

Table 3

5-fold cross-validation results for train/test folds selection concerning RMSE values.

Models	Iteration	Test fold				
		K1	K2	K3	K4	K5
(EN)	1	0.040	0.041	0.040	0.033	0.041
	2	0.033	0.051	0.034	0.040	0.035
	3	0.051	0.030	0.045	0.028	0.036
	4	0.043	0.036	0.027	0.045	0.042
	5	0.038	0.042	0.038	0.039	0.038
	Avg	0.041	0.040	0.037	0.037	0.038
	St. dev	0.006	0.007	0.006	0.006	0.003
(QR)	1	0.042	0.044	0.043	0.034	0.044
	2	0.035	0.053	0.046	0.043	0.037
	3	0.054	0.030	0.048	0.029	0.040
	4	0.045	0.039	0.028	0.047	0.045
	5	0.041	0.045	0.039	0.041	0.040
	Avg	0.043	0.042	0.041	0.039	0.041
	St. dev	0.006	0.007	0.007	0.006	0.003
GPR	1	0.047	0.048	0.049	0.037	0.043
	2	0.034	0.064	0.037	0.048	0.039
	3	0.061	0.031	0.055	0.031	0.038
	4	0.047	0.038	0.027	0.056	0.048
	5	0.045	0.046	0.041	0.045	0.042
	Avg	0.047	0.045	0.042	0.043	0.042
	St. dev	0.009	0.011	0.010	0.009	0.004
GBR	1	0.052	0.072	0.056	0.051	0.060
	2	0.045	0.084	0.049	0.060	0.050
	3	0.094	0.040	0.063	0.045	0.052
	4	0.061	0.046	0.041	0.080	0.051
	5	0.051	0.080	0.048	0.055	0.047
	Avg	0.061	0.064	0.052	0.058	0.052
	St. dev	0.018	0.018	0.008	0.012	0.004

consistency and data partitioning resistance of a model. Models with higher averages and smaller standard deviations can be considered more stable and reliable across different subsets of the data set.

For EN, Fold 4 was the optimum fold that achieved the highest average R^2 of 0.724 and one of the lowest mean RMSE (0.037). This

indicates that EN generalized very well to new data with maintaining the underlying patterns and having a steady error margin. The relatively small standard deviation in R^2 (0.076) and RMSE (0.006) also speaks to its stability.

QR was best in Fold 3, with highest average R^2 (0.743) and lowest average RMSE (0.041). Surprisingly, Fold 3 was most consistent among iterations, with little variation (R^2 standard deviation = 0.040; RMSE standard deviation = 0.007). This indicates the robustness of QR in minimizing prediction errors across partitions when dealing with complex relationships.

GPR worked best in Fold 5 with the highest mean R^2 (0.811) and smallest RMSE value (0.042). Significantly, Fold 5 also had the lowest variation of R^2 values (standard deviation = 0.044), reflecting high stability. The consistently low RMSE for all folds signifies that GPR has the capacity to maintain nonlinearities and yet remain predictively stable.

GBR worked best during Fold 3 with the highest average R^2 (0.673) and the lowest average RMSE (0.052). However, GBR had higher variability of R^2 across folds (standard deviation = 0.021–0.101), which means that it is data partitioning sensitive. However, its low RMSE during Fold 3 indicates good error correction ability through boosting process iteratively.

The results in Tables 2 and 3 are used as a generalizability check for the models and also as the basis for selecting the train/test folds. The folds chosen for each model are presented in Tables 1A, 2A, 3A, and 4A in Appendix I for EN, QR, GPR, and GBR, respectively, such that future analyses are built on the most representative and stable fold of each model.

4.2. Convergence behavior of the developed models

Systematic exploration of the convergence of hybrid models (EN, QR, GPR, and GBR optimized by PO and BO) was conducted by tracking RMSE values across 200 iterations of hyperparameter search, presented in Fig. 3. Convergence patterns in each of the eight models were distinct, reflecting each of their respective algorithmic natures and optimization behaviors. Optimization behavior by Bo related to all four models has a similar pattern, which shows a long drop in steps in a specific number of iterations, but for PO-based hybrid models, especially GPPO, the drop in RMSE values is gradual.

Fig. 4 illustrates the variance in hyperparameter settings over all training iterations of the hybrid model. For instance, the results reveal that GBPO liked relatively small learning rates (optimal ≈ 0.014) and moderate tree depths (optimal ≈ 18), showing a conservative learning policy with controlled model capacity. On the other hand, GBBO trended

towards much higher learning rates (optimal ≈ 0.790) and more profound tree hierarchies (optimal ≈ 45), suggesting an aggressive fitting approach. For GPPO, the hyperparameter ranges were near balanced values of ν (optimal ≈ 1.15) and optimizer restarts (optimal ≈ 5), suggesting balancing flexibility in function approximation and stability in optimization.

4.3. Predictive performance of models

Comparison of performance enhancement between baseline models and hybrid models during training, validation, and test phases is shown in Table 4. The result clearly shows the improvement that has been achieved with the help of optimization, with noticeable enhancement in R^2 , RMSE, MAE, and distance metrics. Average and standard deviation tests by fold (Tables 2 and 3) earlier identified the most generalizable folds to train on, and the greater performance evaluation is tabulated in Table 4. These collectively reflect the value of hybrid optimization in enhancing the accuracy and stability of N_2O emission predictive models.

4.3.1. Base models vs. hybrid counterparts

During the test process, the hybrid models also outperformed their base counterparts. For example, QR yielded an R^2 of 0.921 with an RMSE of 0.031, while QRPO and QRBO exhibited improved test R^2 of 0.968 and 0.979, respectively, with corresponding RMSE reduction to 0.020 and 0.016. Similarly, EN registered a test R^2 of 0.748, whereas ENPO and ENBO registered improved performances of 0.893 and 0.917, respectively, with reduced RMSEs of 0.028 and 0.024. GPR, which was already doing an excellent job as a base learner ($R^2 = 0.958$, RMSE = 0.020), was further improved by GPPO ($R^2 = 0.980$, RMSE = 0.014) and GPBO ($R^2 = 0.986$, RMSE = 0.012). Finally, GBR performed lowest among base learners ($R^2 = 0.706$, RMSE = 0.048), but its optimized versions (GBPO, GBBO) took R^2 to 0.881 and 0.901, and lowered RMSE to 0.030 and 0.028. All these results confirm that optimization enhanced all the models, with the biggest enhancements noticed in the less intense base models like GBR and EN.

4.3.2. Performance comparison among base models

Of the base learners, GPR performed the best, with an R^2 of 0.958 in the test phase, followed by QR (0.921), EN (0.748), and GBR (0.706). GPR also had high generalizability, where R^2 fluctuated by less than 2 % in going from training to test phases, highlighting its stability. QR had good accuracy but was less stable between validation and test phases. EN surpassed GBR in measures of error (RMSE and MAE), as would be anticipated of the strength of its regularization, though both models fell far behind GPR and QR in sensitivity forecasting.

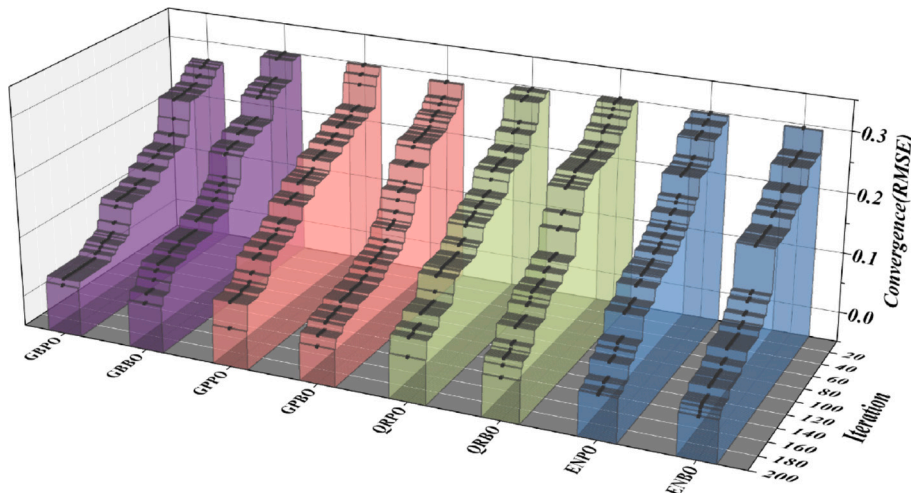


Fig. 3. Convergence behavior presentation in all iterations of hybrid models training.

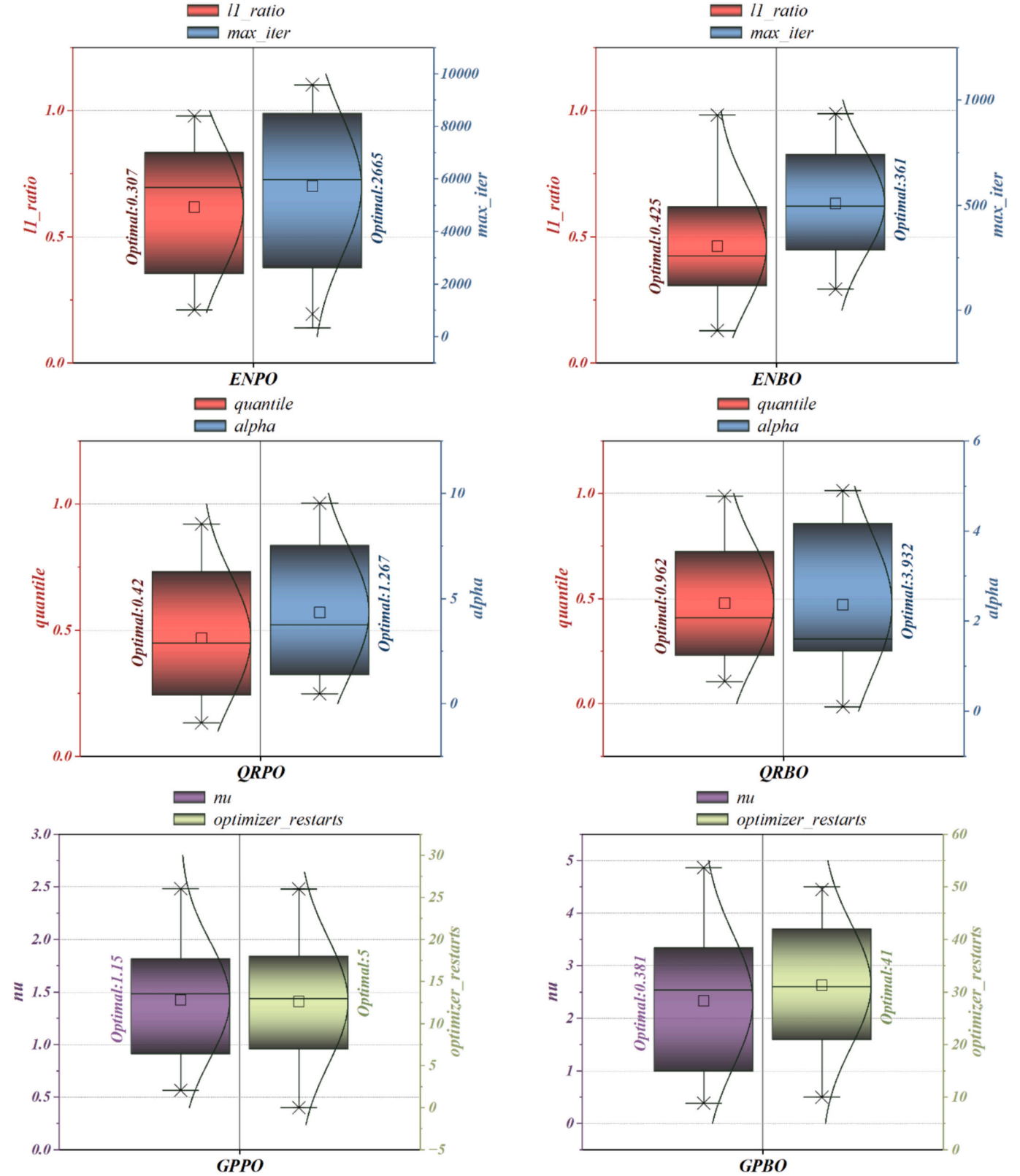


Fig. 4. Box plot for hyperparameter configuration presentation in all iterations of hybrid models training.

4.3.3. Hybrid model comparisons

Of the best of the optimized models, GPBO was the best overall with the highest test R^2 (0.986) and lowest RMSE (0.012). QRBO also fared superbly ($R^2 = 0.979$, RMSE = 0.016), while ENBO and ENPO did much better than EN, with test R^2 values of 0.917 and 0.893, respectively.

GBPO and GBBO, though not outperforming GPR-based models, nevertheless achieved remarkable gains compared to their base analogue GBR, increasing R^2 by ~20 % and cutting RMSE by half. These shifts affirm that optimization did more than raise better learners (like GPR) but also greatly awakened inferior models (like GBR).

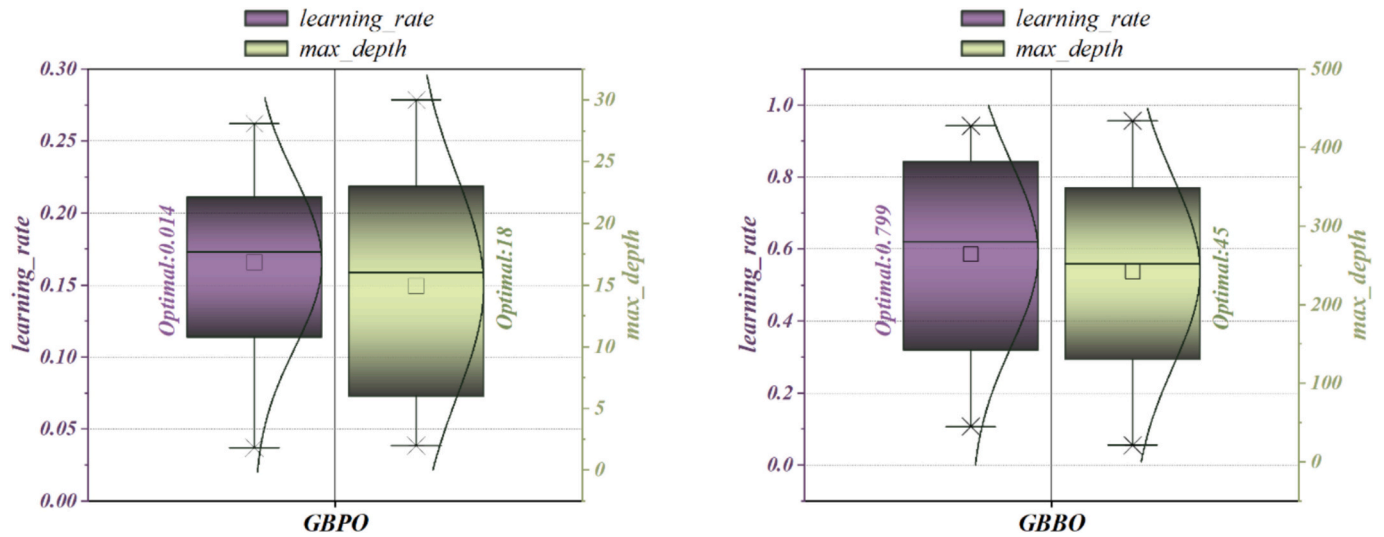


Fig. 4. (continued).

Table 4
Training time and prediction performance reports of the models.

Model and Training Time	Phase	Metrics							
		R ²	RMSE	MAE	RSE	MAPE	SMAPE	Euclidean Distance	Normalized Distance
QR (250 s)	Train	0.822	0.041	0.021	0.178	96.161	0.0004	0.0461	1.321
	Validation	0.691	0.046	0.023	0.309	nan	0.0041	0.0514	1.474
	Test	0.921	0.031	0.016	0.079	94.527	0.0028	0.0349	1
EN (1.5 s)	Train	0.877	0.034	0.018	0.123	nan	0.0004	0.0385	1.103
	Validation	0.872	0.039	0.021	0.128	99.209	0.0032	0.0443	1.27
	Test	0.748	0.042	0.020	0.252	90.076	0.0038	0.0466	1.335
GPR (80 s)	Train	0.956	0.021	0.012	0.044	nan	0.0003	0.0242	0.694
	Validation	0.936	0.019	0.012	0.064	88.785	0.0023	0.0229	0.657
	Test	0.958	0.020	0.012	0.042	86.062	0.0018	0.0233	0.667
GBR (6 s)	Train	0.713	0.054	0.025	0.287	95.272	0.0005	0.0596	1.708
	Validation	0.709	0.045	0.019	0.291	nan	0.0035	0.0489	1.401
	Test	0.706	0.048	0.024	0.294	88.740	0.0039	0.0535	1.535
QRPO (1200 s)	Train	0.943	0.023	0.014	0.057	90.657	0.0003	0.0269	0.771
	Validation	0.913	0.024	0.014	0.087	nan	0.0024	0.0279	0.799
	Test	0.968	0.020	0.012	0.032	89.278	0.0020	0.0233	0.667
ENPO (10 s)	Train	0.947	0.022	0.013	0.053	nan	0.0003	0.0256	0.734
	Validation	0.953	0.024	0.015	0.047	89.406	0.0023	0.0283	0.811
	Test	0.893	0.028	0.015	0.107	90.046	0.0030	0.0318	0.911
GPPO (400 s)	Train	0.981	0.014	0.009	0.019	nan	0.0002	0.0166	0.476
	Validation	0.974	0.012	0.008	0.026	75.673	0.0016	0.0144	0.414
	Test	0.980	0.014	0.009	0.020	74.602	0.0014	0.0166	0.475
GBPO (40 s)	Train	0.896	0.032	0.018	0.104	86.262	0.0003	0.0367	1.052
	Validation	0.929	0.022	0.012	0.071	nan	0.0022	0.0255	0.731
	Test	0.881	0.030	0.017	0.119	84.196	0.0028	0.0343	0.984
QRBO	Train	0.960	0.020	0.011	0.040	70.538	0.0002	0.0228	0.653
	Validation	0.935	0.021	0.011	0.065	nan	0.0019	0.0238	0.681
	Test	0.979	0.016	0.010	0.021	80.724	0.0016	0.0189	0.542
ENBO	Train	0.963	0.019	0.011	0.037	nan	0.0002	0.022	0.63
	Validation	0.966	0.020	0.012	0.034	68.824	0.0019	0.0233	0.667
	Test	0.917	0.024	0.013	0.083	76.053	0.0025	0.0274	0.786
GPBO	Train	0.985	0.012	0.008	0.015	nan	0.0002	0.0144	0.414
	Validation	0.979	0.011	0.007	0.021	64.942	0.0014	0.013	0.374
	Test	0.986	0.012	0.007	0.014	65.145	0.0011	0.0141	0.404
GBBO	Train	0.923	0.028	0.015	0.077	74.169	0.0003	0.0319	0.915
	Validation	0.955	0.018	0.009	0.045	nan	0.0018	0.0201	0.576
	Test	0.901	0.028	0.015	0.099	81.489	0.0025	0.0319	0.914

4.3.4. Euclidean distance and normalized distance

Distance-based metrics also supported these findings. Among base models, GPR posted the lowest test Euclidean Distance (0.0233) and Normalized Distance (0.667) compared to QR (0.0349, 1.000) and EN (0.0466, 1.335). The worst was GBR (0.0535, 1.535). Optimized models, however, significantly reduced distance measurements: GPBO posted the smallest Euclidean Distance (0.0141) and Normalized Distance

(0.404), closely followed by QRBO (0.0189, 0.542). Most significantly, ENBO achieved a 41 % reduction in Normalized Distance to base EN (0.786 vs. 1.335), with GBBO lowering the distance values of GBR by ~40 %. These successive reductions are the strongest confirmation that optimization did not just enhance the accuracy of point predictions but also their overall geometric closeness to actual emission values.

A clear progression from base to hybrid models, illustrated by in-

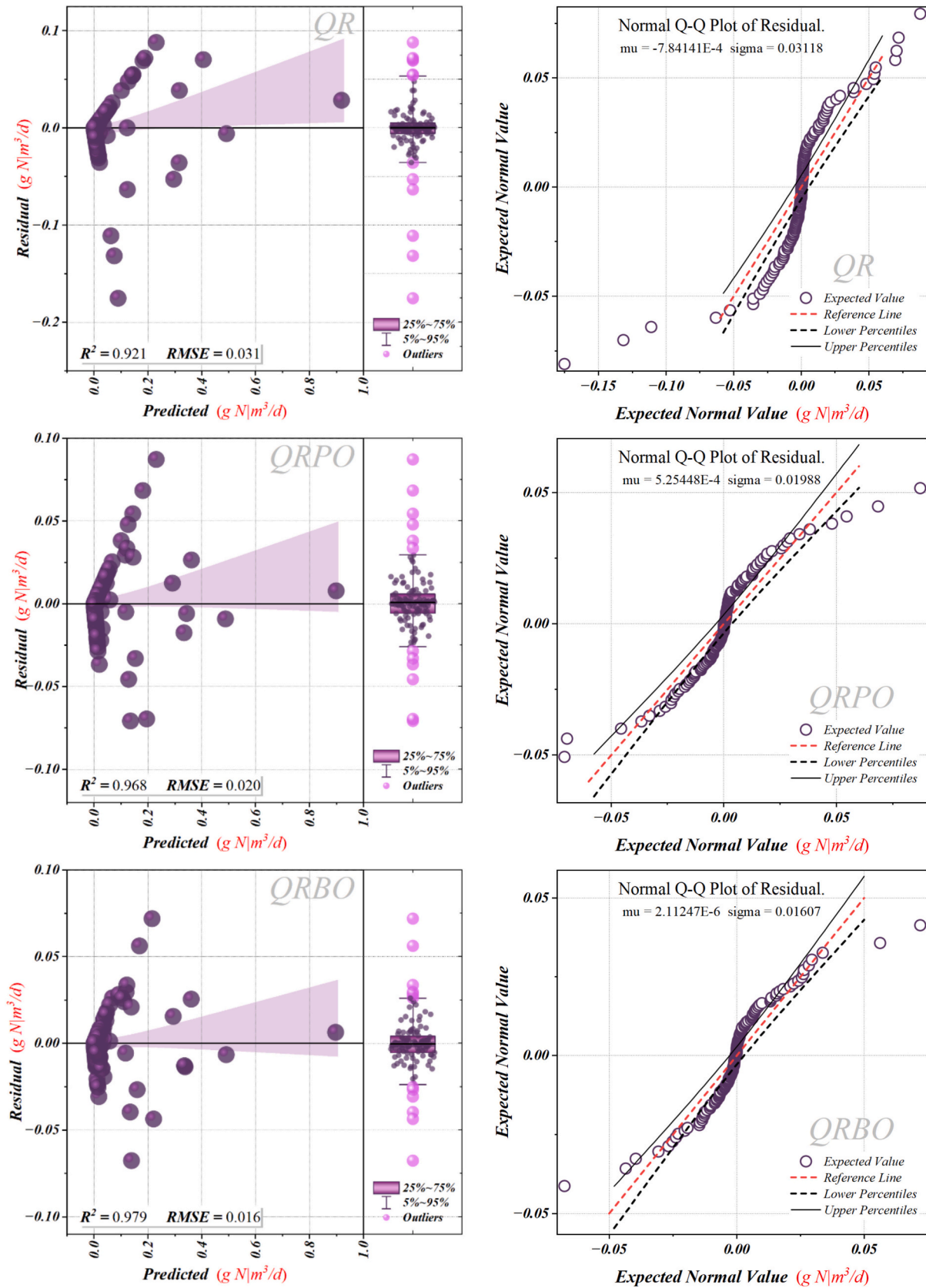


Fig. 5. Scatter and Q-Q plots of residuals for single and hybrid prediction models.

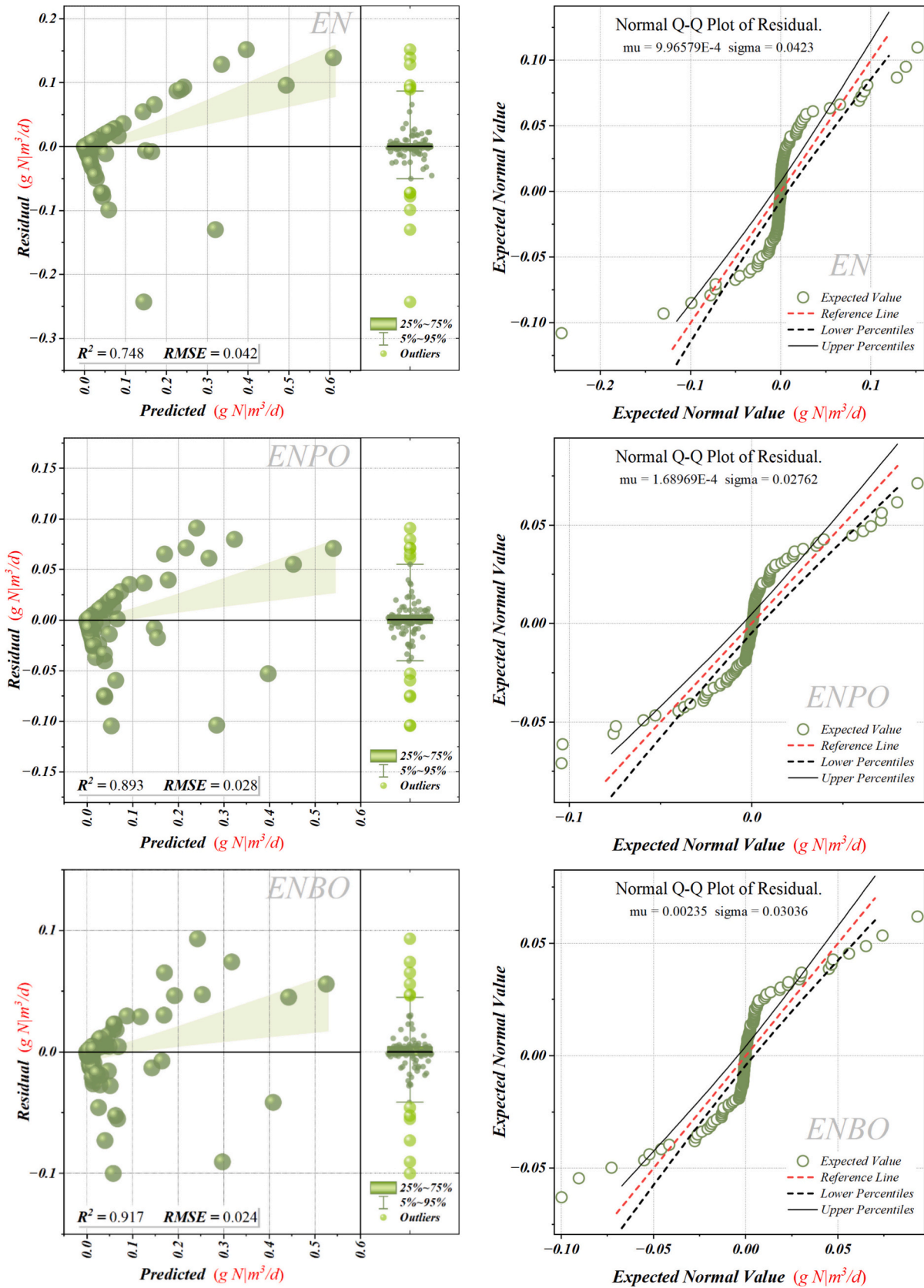


Fig. 5. (continued).

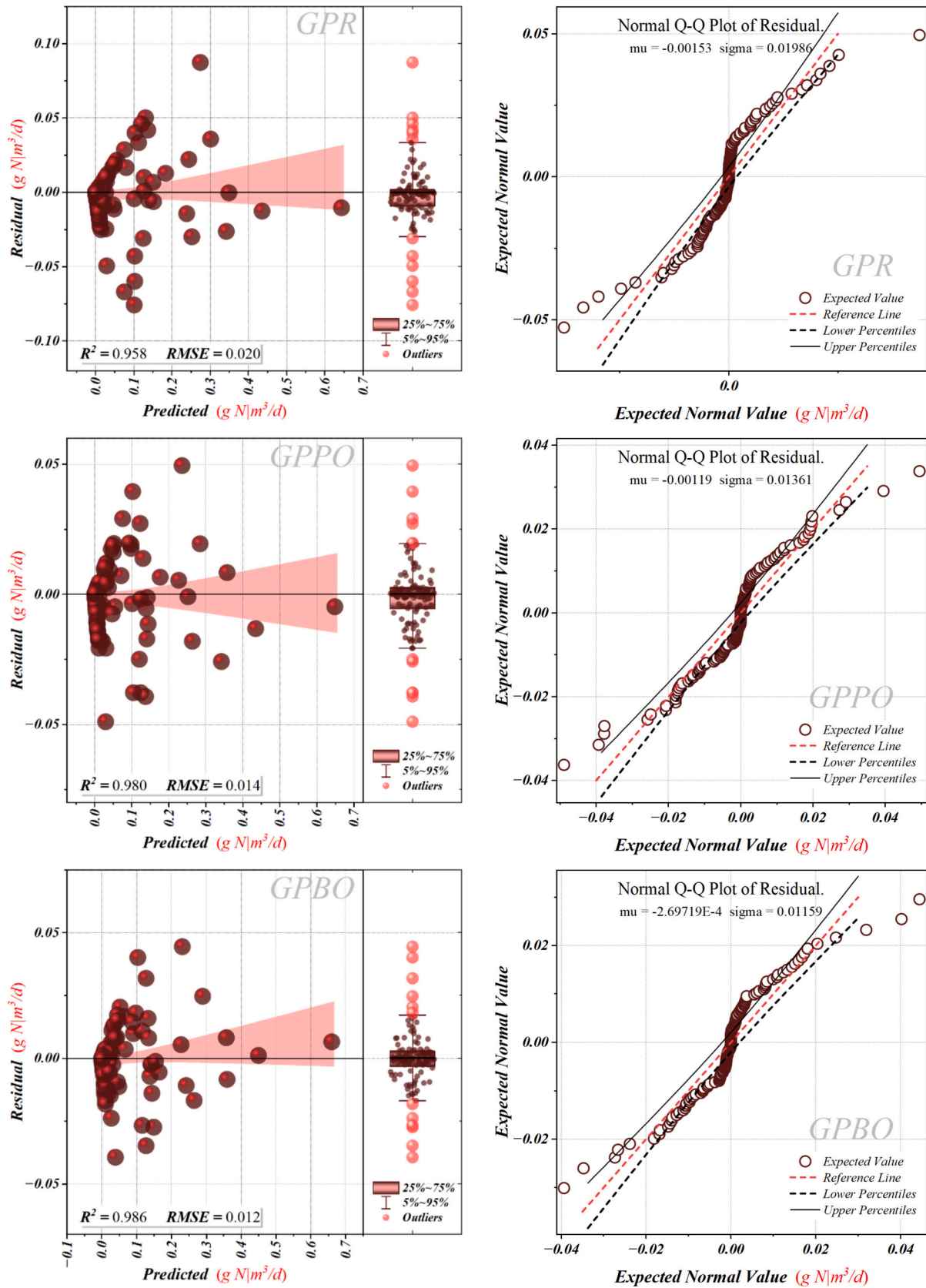


Fig. 5. (continued).

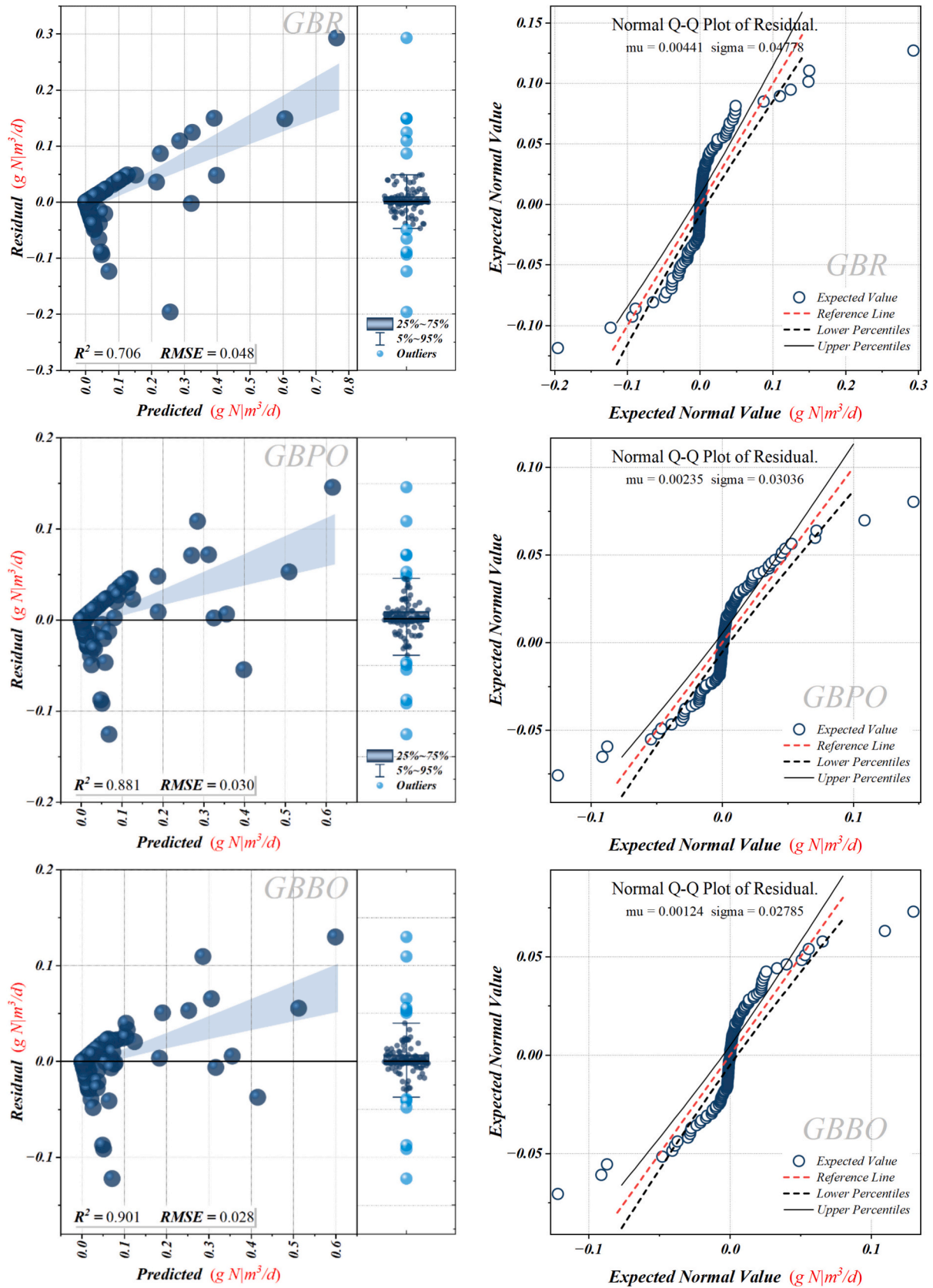


Fig. 5. (continued).

creases in R^2 and dramatic RMSE and MAE reductions in Table 4, demonstrates that optimized hybrid approaches can deliver substantially more precise N_2O emission forecasts. Such enhanced accuracy and stability (as shown by lower Euclidean and normalized distances) can directly inform more effective control strategies in wastewater treatment plants, enabling tighter regulation of aeration and dosing processes. Moreover, the robustness of these models across metrics and phases suggests that the same optimization framework could be applied to varied data sets and plant configurations, facilitating scalable, data-driven improvements in emission management. Collectively, these performance gains lay the groundwork for more reliable, real-time decision support tools that can substantially reduce greenhouse-gas outputs and advance the field of predictive wastewater process control.

4.4. Comparative visualization of models' performance

The residual scatter and Q-Q plots in Fig. 5 illustrate the relationship between predicted N_2O emission rates (x-axis) and their corresponding residuals (y-axis, defined as measured minus predicted values) for both base and hybrid models. The confidence band is shaded on each scatter plot, and Q-Q plots indicate the congruence of residual distributions with the theoretical normal distribution. Models characterized by tightly packed residuals around zero and Q-Q points on the 45° line of reference indicate unbiased prediction and normally distributed errors, which are signs of higher predictive dependability. On the other hand, greater scatter dispersion or systematic variation in the Q-Q plots reveals heteroscedasticity or non-normal residuals, signaling poorer generalizability. Combined, the graphical diagnostics complement statistical measures in that hybrid models not only improve accuracy but also eliminate systematic bias and variance in residual patterns from their respective base counterparts.

For the GPPO model, residuals are closely scattered around zero with little variance along the predicted range. The Q-Q distribution is close to the reference line, indicating normally distributed errors and absence of heteroscedasticity. This agrees with the high coefficient of determination ($R^2 = 0.980$) and low RMSE (0.014), confirming again its greater predictive accuracy and stability. In contrast, the GBPO model has more scatter dispersal with seemingly increasing variance in higher predicted values and departures of the Q-Q distribution from the ideal line. These patterns reflect the presence of heteroscedasticity and more erratic residual behavior. The lower R^2 value (0.881) and larger RMSE value (0.030) with respect to GPPO again reinforce the lower predictive reliability of this base model.

4.5. Error distribution analysis

Error distribution statistics during the test run, given in Table 5, showed distinct pattern differences between the models. Broader error ranges were observed in base models compared to their optimized

Table 5
Statistical properties of test phase errors for prediction models.

Components	Properties				
	Min (ABS)	Max (ABS)	Avg	St. Dev	Median
QR	0.212	66.248	0.753	58.628	6.796
EN	0.554	64.992	-0.978	57.368	-4.263
GPR	0.001	66.686	-5.188	54.568	-6.385
GBR	0.609	65.993	7.956	59.423	62.324
QRPO	0.876	66.412	3.266	57.344	9.535
ENPO	0.001	66.508	3.084	58.072	21.841
GPPO	0.001	66.528	0.409	51.896	-0.546
GBPO	0.665	66.524	6.969	57.007	32.821
QRBO	0.719	65.763	-4.648	51.722	-4.177
ENBO	4.326	65.546	-3.026	50.723	-7.549
GPBO	0.268	65.367	-1.726	47.787	-1.970
GBBO	0.686	65.602	-0.641	53.590	0.486

Table 6
SHAP sensitivity analysis results.

Feature	Sensitivity
Influent flowrate (m ³ /h)	0.0014
DO (g O ₂ /m ³)	0.0043
NH ₄ (g N/m ³)	0.0079
NO ₃ (g N/m ³)	0.0078
N ₂ O (g N/m ³)	0.0523
Temperature (°C)	0.0076

Table 7
LIME sensitivity analysis results.

Feature	EN	GBR	GPR	QR
0.17 < Influent flowrate (m ³ /h) ≤ 0.26	0	0.0024	0.0044	5.03E-34
0.28 < Temperature (°C) ≤ 0.56	0	0.0044	0.0351	2.22E-34
DO (g O ₂ /m ³) > 0.27	0	0.0017	0.5684	6.27E-34
N ₂ O (g N/m ³) > 0.03	0	0.0081	0.4405	2.22E-34
NH ₄ (g N/m ³) > 0.20	0	0.0037	0.5011	3.42E-34
NO ₃ (g N/m ³) > 0.20	0	0.0033	0.0685	1.04E-33

counterparts, signifying the positive impact of hybridization in reducing extreme deviations. GBR, for instance, obtained a maximum absolute error of 65.993, which GPBO reduced to 65.367, thus establishing improved robustness.

Average and median errors revealed systematic biases. Several base models, such as EN and GPR, exhibited negative mean errors, which indicate underpredicting biases. In contrast, GBR possessed a high positive bias with a mean error of 7.956. Such systematic biases were relieved in hybrid models. For example, GBBO reduced the mean error of GB-based modeling to nearly zero (-0.641), while GPPO generated a nearly unbiased mean (0.409) with a median close to zero (-0.546). These improvements suggest more balanced and credible forecasts.

Standard deviation, a measure of variability of error, was always smaller in hybrid models. GPBO had the lowest standard deviation (47.787), with very stable performance. ENBO and QRBO also had less variability than the corresponding base models, confirming the consistency dividends from optimization.

4.6. Sensitivity analysis results

SHAP and LIME sensitivity analyses in Tables 6 and 7 were used to assign quantitative values to input parameter contributions in the effects on N_2O emissions. SHAP values were obtained using the Tree SHAP algorithm for tree models (GBR, GPR), which approximates Shapley values by analyzing all coalitions of features. Feature contributions were summarized by calculating the mean absolute SHAP value across all instances, which provides a global ranking of importance that reflects each parameter's net contribution to model predictions. Concentration of dissolved N_2O was indicated to be the most significant factor by SHAP (score: 0.0523), since it is a direct quantitative measure of nitrification and denitrification activities. Its high global importance indicates that persistent monitoring of this parameter would greatly enhance model-based emission control.

LIME explanations were created based on local linear approximations that were fit in the neighborhood of each prediction. The model predictions were locally weighted sampled for each instance, and a linear regression was fit to locally approximate the complicated model. These resulting coefficients were taken to be local feature importance scores. LIME picked up on the high local importance of NH_4^+ (GPR score: 0.5011), even though its lower global SHAP score (0.0079). This would mean that NH_4^+ could be the dominant emission behavior at some operating conditions, e.g., nitrification imbalances, in support of context-based control measures.

DO and temperature possessed intermediate SHAP importance (0.0043 and 0.0076, respectively), while LIME showed DO (0.5684 for

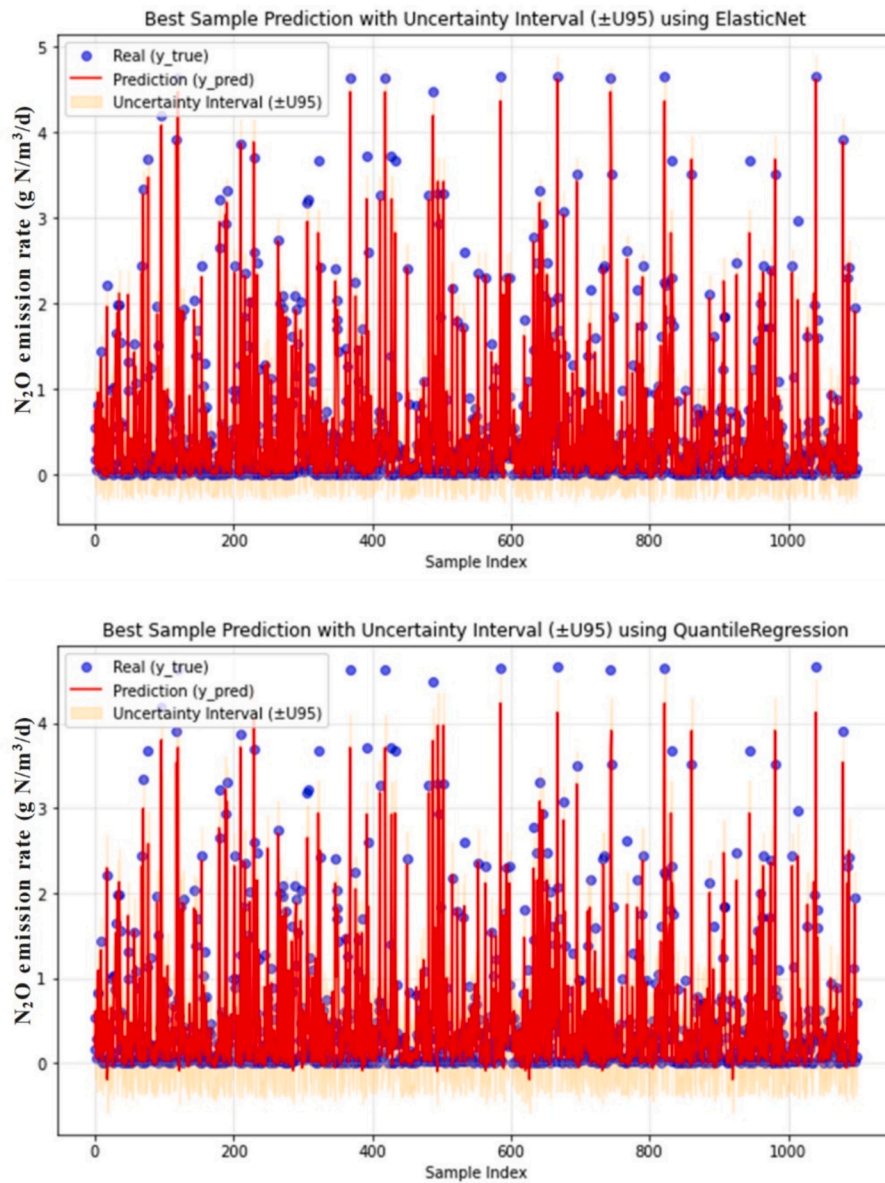


Fig. 6. Observed vs. predicted N₂O emission rates using prediction models with 95 % prediction intervals (PI).

GPR) to have locally significant roles, particularly in oxygen-sensitive treatment stages. NO₃⁻, with comparable SHAP and LIME values (0.0078 and 0.0685, respectively), occurred under conditions dominated by denitrification. Influent flowrate, even with low global SHAP value (0.0014), contributed detectable local impacts (0.0044 for GPR) during hydraulic load changes.

By combining SHAP and LIME, a complementarity interpretability framework was achieved: SHAP produces stable global rankings for long-term optimization (e.g., aeration strategies), while LIME gives instance-level explanations to aid real-time operating decisions (e.g., carbon addition or ammonium regulation), additionally raising the utility of model predictions within dynamic WWTP environments.

4.7. Monte Carlo uncertainty results

For the estimation of input uncertainty and propagation error, input measurements (DO, NH₄⁺, NO₃⁻, N₂O, Influent flowrate, and Temperature) were entered with independent Gaussian noise equivalent to 5 % of

each feature mean. Then, Monte Carlo propagation (1000 runs) was applied on the pre-trained GPR, GBR, QR, and EN models to generate empirical predictive distributions as well as 95 % prediction intervals (PIs). For the GPR model, analytical calculated prediction variances were used directly to estimate the 95 % PIs. Model performance was evaluated on an held-out test set on the basis of RMSE, R², mean width of PI, and empirical 95 % coverage. The graphical plots of predicted vs. observed N₂O emission rates with 95 % PIs are shown in Fig. 6.

The results summary of uncertainty analysis is provided in Table 8. Of the models, EN performed overall best in prediction, lowest RMSE (0.2128) and highest R² (0.9304), yet also best coverage of uncertainty (0.9723) with relatively narrow PIs (0.4172). Quantile Regression and GPR did also, although less precisely (RMSE = 0.2620 and 0.2401; R² = 0.8950 and 0.9115, respectively) yet still excellent coverage (0.9585 and 0.9604). The GBR model was worst in this setting with the highest RMSE (0.2676), lowest R² (0.8906), as well as coverage (0.9567), albeit with the biggest PIs (0.5245).

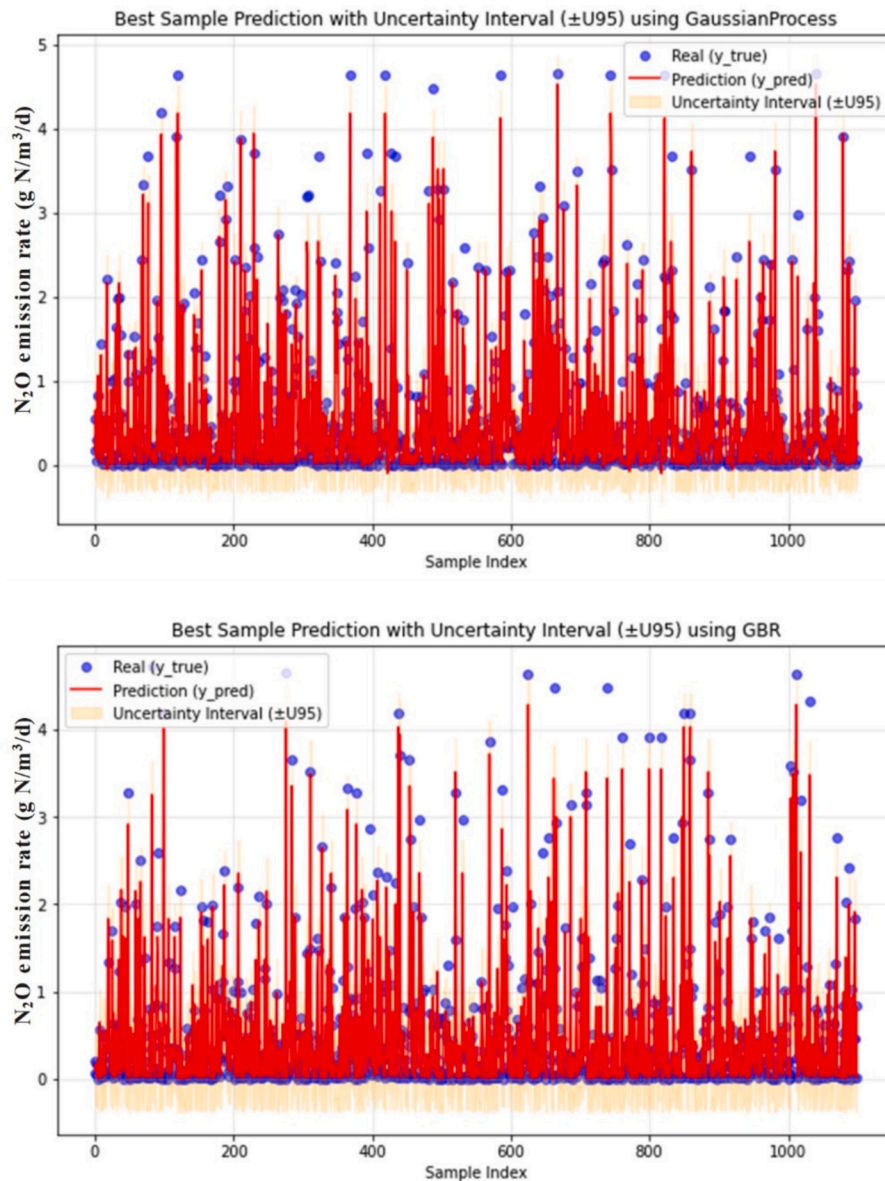


Fig. 6. (continued).

Table 8
Uncertainty analysis results.

Model	RMSE	R ²	Coverage_95	PI
ElasticNet	0.2128	0.9304	0.9723	0.4172
Quantile	0.2620	0.8950	0.9585	0.5135
GPR	0.2401	0.9115	0.9604	0.4706
GBR	0.2676	0.8906	0.9567	0.5245

5. What-if scenarios interpretations and discussion

5.1. Significance of the finding

Accurate N₂O emission estimation is important in light of the extremely powerful greenhouse effect of N₂O, with some 300-fold that of CO₂, and its ozone depleting potential. Unlike standard mechanistic models, the ML approach is sparsely calibrated and generally limited in ability to handle variability, but the nonlinear interactions between decisive parameters DO, NH₄⁺, NO₃⁻, and temperature are handled explicitly. Flexible and site-specific estimation is therefore more

feasible, overcoming a fundamental limitation of traditional approaches.

As for the lower test set R² of the optimized GBBO relative to its non-optimized counterpart (0.901 vs. 0.881 for GBBO vs. GBPO), this is due to the trade-off between overfitting and generalizability. The optimization process improves model performance in the training set and validation set through hyperparameter tuning, which may decrease test-set R² slightly because the model is more regularized and less noisy. This behavior is actually in alignment with the observation that GBBO does better in R² on training (0.923 vs. 0.896) and validation (0.955 vs. 0.929) but the test set shows a slight decline. The conduct informs us that optimization makes the model more robust and predictive, not overfitting to specific patterns in the training set.

Sensitivity analysis using SHAP and LIME revealed that dissolved N₂O concentration is most crucial, as it directly reflects nitrification and denitrification kinetics. While NH₄⁺ had relatively lower global relevance in SHAP, LIME showed that it possesses high local contributions under some operating conditions, highlighting the need for adaptive measures to varied nitrogen loading. DO, temperature, and NO₃⁻ also possessed condition-dependent contributions, particularly in oxygen-

sensitive or denitrification-driven regimes. The integration of SHAP and LIME indicates that model predictions can be utilized for WWTP long-term process design optimization and real-time control to achieve low emissions without compromising treatment efficiency.

5.2. What-if scenarios and practical implications

To examine the responsiveness of the hybrid models to realistic operational perturbations, three what-if scenarios were designed simulating changes in the most significant WWTP variables, N_2O , NH_4^+ , DO, temperature, and influent flowrate:

• Scenario 1: N_2O increase

20 % increase in N_2O was applied to all the samples, and Conditional Generative Adversarial Networks (CGAN) were used to generate realistic NH_4^+ and DO values conditioned on the updated levels of N_2O . The best-trained model was then used to predict an emission rate of N_2O .

• Scenario 2: Reducing NH_4^+ and increasing temperature

NH_4^+ was decreased by 15 % and temperature was increased by 5 °C for all samples. Realistic DO and N_2O values conditioned on the reduced levels of NH_4^+ and increased temperature produced by CGAN were used, and N_2O emission rate was estimated using the best-trained model.

• Scenario 3: Multi-variable perturbation

DO was increased by 10 %, N_2O was increased by 10 %, and influent flowrate was moderately raised by 5 %. Correlated simultaneous NH_4^+ , NO_3^- , and temperature data were generated with the assistance of CGAN, and the top-performing model was utilized to predict the emission rate of N_2O .

Statistical properties of input and output variables for all scenarios are given in Tables 9, 10 and 11, and the resulting variation of N_2O emission rates compared to the initial dataset is shown in Fig. 7. This type of analysis enables the assessment of model predictions under altered conditions and provides practical information on plant operation, process optimization, and emission abatement strategies.

• Predictive Maintenance and Process Optimization:

Low RMSE values achieved for hybrid models partially confirm their efficacy to predict N_2O peaks accurately under dynamic WWTP operational conditions. In integrated applications within SCADA systems, such models can continuously monitor the key operating parameters such as DO, NH_4^+ , and temperature, which are proven biological drivers of nitrification–denitrification processes. By predicting future emission peaks, advanced adjustment of aeration rate or sludge age can be carried

Table 9
Statistical properties of N_2O emission data variables: what-if Scenario 1.

Variable	Train (Sample size = 1234)				
	Min	Median	Mean	Std	Var
Influent flowrate (m^3/h)	0.000	0.167	0.211	0.141	0.020
DO ($\text{g O}_2/\text{m}^3$)	0.158	0.241	0.278	0.093	0.009
NH_4 ($\text{g N}/\text{m}^3$)	0.128	0.303	0.435	0.330	0.109
NO_3 ($\text{g N}/\text{m}^3$)	0.000	0.094	0.150	0.150	0.022
N_2O ($\text{g N}/\text{m}^3$)	0.000	0.006	0.046	0.118	0.014
Temperature (°C)	0.000	0.574	0.559	0.284	0.081
N_2O emission rate ($\text{g N}/\text{m}^3/\text{d}$)	0.000	0.015	0.049	0.098	0.010

Table 10

Statistical properties of N_2O emission data variables: what-if Scenario 2.

Variable	Train (Sample size = 1234)				
	Min	Median	Mean	Std	Var
Influent flowrate (m^3/h)	0.000	0.167	0.211	0.141	0.020
DO ($\text{g O}_2/\text{m}^3$)	0.158	0.212	0.209	0.012	0.000
NH_4 ($\text{g N}/\text{m}^3$)	0.000	0.108	0.141	0.102	0.010
NO_3 ($\text{g N}/\text{m}^3$)	0.000	0.094	0.150	0.150	0.022
N_2O ($\text{g N}/\text{m}^3$)	0.053	0.282	0.284	0.077	0.006
Temperature (°C)	5.000	5.574	5.559	0.284	0.081
N_2O emission rate ($\text{g N}/\text{m}^3/\text{d}$)	0.000	0.015	0.049	0.098	0.010

Table 11

Statistical properties of N_2O emission data variables: what-if Scenario 3.

Variable	Train (Sample size = 1234)				
	Min	Median	Mean	Std	Var
Influent flowrate (m^3/h)	0.000	0.175	0.221	0.148	0.022
DO ($\text{g O}_2/\text{m}^3$)	0.000	0.254	0.261	0.086	0.007
NH_4 ($\text{g N}/\text{m}^3$)	−0.119	0.134	0.142	0.089	0.008
NO_3 ($\text{g N}/\text{m}^3$)	0.004	0.096	0.114	0.089	0.008
N_2O ($\text{g N}/\text{m}^3$)	0.000	0.005	0.042	0.108	0.012
Temperature (°C)	0.189	0.864	0.899	0.341	0.116
N_2O emission rate ($\text{g N}/\text{m}^3/\text{d}$)	0.000	0.015	0.049	0.098	0.010

out, smoothing out conditions before uncontrollable N_2O accumulation. As an example, in Scenario 1 (20 % increase in N_2O), the model identified concurrent increases in NH_4^+ (mean 0.435 $\text{g N}/\text{m}^3$) and DO fluctuation (std 0.093 $\text{g O}_2/\text{m}^3$), and therefore decreasing emissions could be achieved by fine-tuning aeration to maximize the nitrifiers and avoid oxygen overfeeding.

• Strategic Planning through Scenario Analysis:

The what-ifs amplify model interpretability by pushing statistical results into operational understanding. For Scenario 2 (−15 % NH_4^+ , +5 °C), the models predicted elevated N_2O levels (mean 0.284 $\text{g N}/\text{m}^3$) despite reduced substrate, reproducing the established effect of elevated temperatures accelerating nitrifier metabolism but potentially destabilizing denitrification. This is a climate-sensitive trade-off: more conditions can contribute to more emissions even when influent loads are reduced, again emphasizing the importance of seasonal planning in temperate regions such as Copenhagen, Denmark, where the Avedøre Wastewater Treatment Plant (WWTP) operates. Similarly, Scenario 3 (multi-variable perturbation) showed how rises in DO (+10 %), N_2O (+10 %), and influent flow (+5 %) created modest increases in mean NO_3^- (0.114 $\text{g N}/\text{m}^3$) but increased variability in the form of DO and temperature. This scenario simulates stormflow and transient hydraulic conditions prevalent in European WWTPs like Avedøre, where inflow changes may decrease aeration efficiency and boost short-term N_2O emissions. Such scenario-based results allow operators to attempt “what-if” management situations, e.g., changing recirculation rates or pre-anoxic zones—before being implemented on a full scale.

• Integration with Smart Technologies and IoT:

The results indicate how hybrid approaches, particularly when

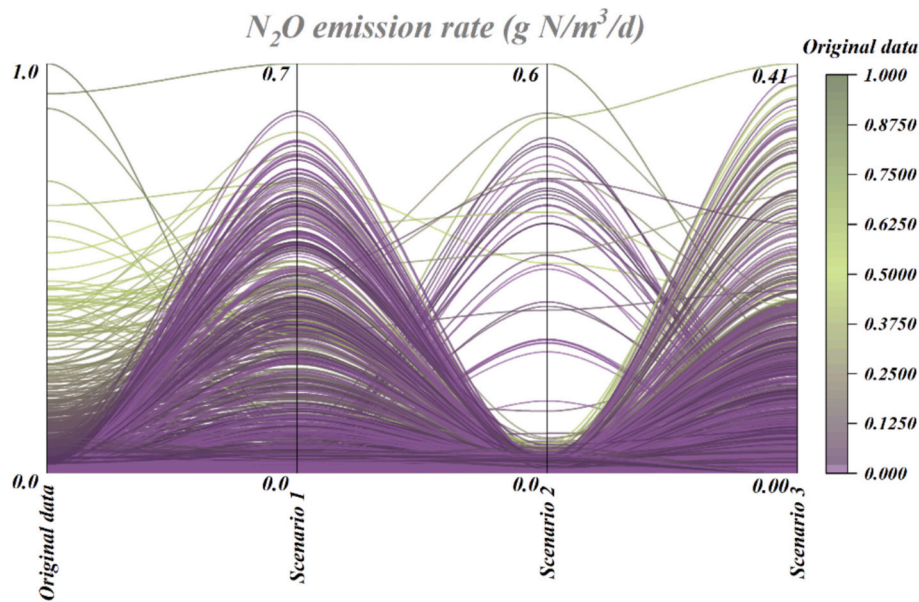


Fig. 7. Variation of WWTP N₂O emission rate: original data compared with what-if scenarios.

paired with generative methods like CGAN, can be implemented in real-time sensor systems to facilitate anticipatory emission control. For instance, DO probes with NH₄⁺ and temperature monitors can continuously feed information into the trained model, and the model then simulates likely emission reactions to anticipated perturbations. If NH₄⁺ begins to rise above threshold levels under warm temperatures (as in Scenario 2), the model would facilitate auto-dosing of carbon to balance denitrification and prevent N₂O accumulation. This provides operators with not just a monitoring tool, but also a forecast control loop, establishing distinct paths for integration into plant automation.

• Biological and Chemical Interpretation of Key Predictors:

The scenario analyses reinforce the biological mechanisms underlying N₂O formation:
Excessive NH₄⁺ loading increases substrate levels for ammonia-oxidizing bacteria, accelerating incomplete nitrification and emission of N₂O. Raising DO (Scenarios 1 and 3) initially accelerates nitrification but can suppress heterotrophic denitrification, leading to incomplete pathways and raised emissions. Raising temperature (Scenario 2) accelerates metabolic rates but can impair the balance between nitrifiers and denitrifiers and amplify emission intensity.

• Energy Efficiency, Cost Savings, and Environmental Impacts:

By quantifying emissions following different operational and climatic disturbances, the models optimized aeration and dosing plans that minimized both emissions and energy consumption. For example, Scenario 1 shows that NH₄⁺ stability under high N₂O could prevent over-aeration at a high cost but still ensure compliance. Notably, the use of emission projections in governance reporting frameworks (e.g., IPCC Tier 3 procedures) enables WWTPs to report baseline emissions alongside mitigation strategies. This improves accountability as well as facilitating co-benefits like the reduction of greenhouse gas footprint, aligned with local climate goals.

• Operator Training and Stakeholder Engagement

The what-if scenarios are simply convertible to interactive dashboards, where operators can dynamically see the effects of DO, NH₄⁺, temperature, or influent flow changes on N₂O emission. By integrating scenario outputs with SHAP-based interpretability, these dashboards provide operators a plain, “cause-and-effect” understanding of the leading causes of emission variation. This approach not only enhances operating decision-making, allowing quick reactions to aeration, carbon dosing, or flow management, but also enhances transparency and

Table 12
Comparison between this study and previous studies in literature.

Researchers	Hwangbo et al. [70]	Khalil et al. [22]	Seshan et al. [24]	Huang et al. [26]	Present study
ML Framework	DNN and LSTM (deep neural nets)	k-NN, AdaBoost, RF, DNN	LSTM encoder-decoder	CNN-BiLSTM with attention (PLO-optimized)	EN, QR, GPR, GBR; hybrids via PO and BO, Uncertainty, and XAI
Dataset	Full-scale WWTP (Avedøre, Denmark)	Full-scale WWTP (long-term field data)	Full-scale WWTP (Amsterdam West, 15 months)	Full-scale WWTP (Altenrhein, large dataset)	Full-scale WWTP (1234 observations)
Results	DNN: R ² >0.90; LSTM: R ² >0.94 (LSTM RMSE ~64 % lower than DNN)	k-NN: R ² ≈ 0.88; AdaBoost: R ² ≈ 0.94; DNN: R ² ≈ 0.90	R ² up to 0.98 (0.5–6 h forecast horizon)	R ² ≈ 0.99 (test)	Best base and hybrid: GBR: R ² = 0.706; RMSE = 0.048 GBBO: R ² = 0.901; RMSE = 0.099
Interpretability	Global sensitivity (variance)	Permutation feature importance	Permutation-based importance	SHAP analysis	SHAP and LIME analyses
Key insights	LSTM outperformed DNN for forecasting; DO, NH ₄ ⁺ , NO ₃ ⁻ identified as key drivers; hybrid DL-mechanistic modeling recommended.	Feature selection reduced inputs by 40 %; simpler ensemble models matched deep nets in accuracy while retaining interpretability.	Historical N ₂ O, influent flow, NH ₄ ⁺ , NO _x , DO were most influential; short-term trends captured well.	Water temperature was most critical; DO and flow exhibited seasonal effects; model robustness confirmed across 48 validation sets.	Dissolved N ₂ O concentration was the top predictor; NH ₄ ⁺ showed strong local effects; hybrid optimization significantly reduced prediction error and enhanced generalizability.

accountability, building trust between stakeholders. In addition, it supports evidence-based policy discussions, departing from the generic measures of emissions control towards context-specific, data-based interventions tailored to the unique hydraulic and climatic conditions of plants such as the Avedøre WWTP.

5.3. Comparative analysis of N₂O emission prediction studies

The studies summarized in Table 12 were all conducted on full-scale WWTPs and employed supervised ML or deep learning models to forecast N₂O emissions. Each work was evaluated with consistent performance metrics and was interpretable through feature-importance or sensitivity analysis techniques. Key water quality indicators were used as model inputs, and hybrid optimization strategies were applied in several cases to enhance both accuracy and generalizability. The comparison highlights model types, dataset characteristics, predictive performance, interpretability approaches, and principal findings to facilitate a clear assessment of methodological advances and practical applicability.

Across all five studies, advanced ML and deep learning frameworks have demonstrated strong capability in forecasting N₂O emissions from full-scale WWTPs, with reported R² values ranging from approximately 0.88 up to 0.99 (and R² up to 0.98 for short-term horizons of 0.5–6 h). Both tree-based ensembles and deep architectures achieve high predictive accuracy, while hybrid optimization techniques (e.g., PO) further reduce error and enhance generalizability. Interpretability (XAI) remains a critical component; global sensitivity analyses, permutation importance, and SHAP/LIME methods consistently highlight dissolved N₂O concentration, dissolved oxygen, and key nitrogen species as dominant drivers of emissions.

6. Conclusion

The study demonstrated the ability of current state-of-the-art machine learning algorithms, hybrid optimization, to forecast nitrous oxide (N₂O) emissions from WWTPs reliably and interpretably. Among the models tested, Gaussian Process Regression with Bayesian Optimization (GPBO) gave the most consistent performance with R² values higher than 0.98 and negligible error statistics during training, validation, and testing. These results validate that the optimized models are not only statistically reliable but also computationally economical to enable real-time applications, thus enabling proactive emission control.

Significantly, interpretability analysis confirmed the most significant influencing driver to be dissolved N₂O concentration, seconded by ammonium and dissolved oxygen, with condition-dependent effects as a function of loading and aeration conditions. The results bridge the gap between statistical prediction and biological explanation, giving operators transparent cause-and-effect mechanisms for decision-making. What-if scenarios also showed that operational or climatic disturbances, such as elevated temperature, hydraulic shocks, or influent N₂O enrichment, can profoundly change emissions, affirming the necessity for adaptive data-driven management.

In addition to predictive accuracy, the model yields actual environmental and economic co-benefits. Reduction of N₂O peaks enables plants to significantly reduce their greenhouse gas emissions, aiding local and national climate objectives as well as adhering to IPCC Tier 3 reporting standards. At the same time, model-based optimized dosing and aeration tactics reduce unnecessary energy consumption, one of WWTPs' most high-cost operational expenses, without triggering over-aeration, chemical loss, and possible fines for noncompliance. Such dual advantages optimize both economic effectiveness and sustainability.

Incorporating Monte Carlo-based uncertainty analysis makes the framework even more reliable, with predictions robust in varying conditions. Overall, this work advances beyond standard retrospective emission estimation, designing a scalable and understandable platform for WWTPs that incorporates precise prediction, scenario-based

resilience planning, and clear-cut environmental and economic value. It is not just a scientific achievement in predictive modeling but also an operational vehicle for operators and policymakers to gain sustainable wastewater management.

7. Limitations and future directions

While this research provides a solid basis for N₂O prediction using hybrid ML models, certain limitations need to be acknowledged. The dataset derived from a single facility (Avedøre WWTP, Denmark) restricts the broad applicability of the results. Site-specific factors such as reactor design, climate, and operational strategies may limit model portability to other WWTPs with different process designs or climatic conditions.

Future research should focus on the extension of datasets to cover a wider variety of WWTP configurations (e.g., activated sludge, sequencing batch reactors, membrane bioreactors) across different geographic and climatic zones. The inclusion of seasonal variation, influent load variation, and different control philosophies (e.g., fixed aeration vs. advanced process control) would significantly enhance model robustness and facilitate scaling to diverse contexts. The second critical direction is the quantification of environmental and economic co-benefits of projected mitigation strategies, such as greenhouse gas emission reductions, energy savings through optimized aeration, and chemical dosing cost savings. Explicit accounting for such trade-offs will make the models more actionable for policymakers and operators alike.

Pairing of real-time sensor data with high-frequency sampling and IoT-enabled systems must also be pursued to enable dynamic model updating and proactive emission control. Such systems would, in effect, enable predictive feedback loops with the capability to pre-emptively adjust aeration or carbon dosing strategies in anticipation of emission spikes. Along with single-point predictions, the development of time series prediction models (i.e., recurrent neural networks, temporal Gaussian processes, or hybrid ML-mechanistic models) must be explored to model the temporal dependencies of N₂O dynamics. Such models would allow operators to foresight emission trends hours or days in advance, allowing more effective scheduling of operational changes and offering better insight into long-term emission patterns under varying climatic and hydraulic conditions.

Methodologically, XAI tools that are more advanced can be utilized to untangle complex feature interactions and provide better interpretability to operators than the explanations now provided by SHAP and LIME. Finally, hybrid approaches that combine data-driven models with mechanistic simulation frameworks (e.g., activated sludge models) have great interdisciplinary potential. Physics-informed ML would have the capability to integrate understanding of biological and chemical processes into predictive models, improving interpretability and ensuring transferability to different plant designs and regional settings. Together, these pathways will improve both the scientific integrity and the real-world applicability of N₂O emission modeling for wastewater treatment.

CRedit authorship contribution statement

Behnam Sadaghat: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Kourosh Behzadian:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Saeid Najjar-Ghabel:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

interests or personal relationships that could have appeared to influence the work reported in this paper.

The authors declare that they have no known competing financial

Appendix I**Table 1A**

Statistical properties of the selected train/test data for EN model evaluation.

Variable	Train (Sample size = 987)					Test (Sample size = 247)				
	Min	Median	Mean	Std	Var	Min	Median	Mean	Std	Var
Influent flowrate (m ³ /h)	0.000	0.166	0.210	0.143	0.020	0.054	0.176	0.213	0.132	0.017
DO (g O ₂ /m ³)	0.020	0.231	0.238	0.081	0.007	0.000	0.234	0.235	0.064	0.004
NH ₄ (g N/m ³)	0.000	0.127	0.167	0.124	0.015	0.019	0.126	0.159	0.104	0.011
NO ₃ (g N/m ³)	0.000	0.093	0.150	0.151	0.023	0.000	0.095	0.151	0.145	0.021
N ₂ O (g N/m ³)	0.000	0.005	0.037	0.095	0.009	0.000	0.004	0.043	0.111	0.012
Temperature (°C)	0.000	0.573	0.560	0.283	0.080	0.000	0.603	0.552	0.290	0.084
N ₂ O emission rate (g N/m ³ /d)	0.000	0.016	0.050	0.098	0.010	0.000	0.012	0.047	0.097	0.010

Table 2A

Statistical properties of the selected train/test data for QR model evaluation.

Variable	Train (Sample size = 987)					Test (Sample size = 247)				
	Min	Median	Mean	Std	Var	Min	Median	Mean	Std	Var
Influent flowrate (m ³ /h)	0.000	0.167	0.212	0.146	0.021	0.054	0.167	0.206	0.115	0.013
DO (g O ₂ /m ³)	0.000	0.231	0.236	0.073	0.005	0.020	0.232	0.241	0.095	0.009
NH ₄ (g N/m ³)	0.000	0.128	0.166	0.123	0.015	0.006	0.123	0.162	0.110	0.012
NO ₃ (g N/m ³)	0.000	0.094	0.150	0.148	0.022	0.000	0.096	0.153	0.154	0.024
N ₂ O (g N/m ³)	0.000	0.005	0.038	0.096	0.009	0.000	0.005	0.039	0.108	0.012
Temperature (°C)	0.000	0.574	0.557	0.284	0.081	0.047	0.573	0.565	0.286	0.082
N ₂ O emission rate (g N/m ³ /d)	0.000	0.015	0.050	0.098	0.010	0.000	0.012	0.047	0.098	0.010

Table 3A

Statistical properties of the selected train/test data for GPR model evaluation.

Variable	Train (Sample size = 987)					Test (Sample size = 247)				
	Min	Median	Mean	Std	Var	Min	Median	Mean	Std	Var
Influent flowrate (m ³ /h)	0.000	0.166	0.209	0.137	0.019	0.048	0.178	0.220	0.154	0.024
DO (g O ₂ /m ³)	0.000	0.231	0.238	0.079	0.006	0.062	0.231	0.234	0.075	0.006
NH ₄ (g N/m ³)	0.000	0.128	0.166	0.120	0.014	0.000	0.125	0.164	0.121	0.015
NO ₃ (g N/m ³)	0.000	0.094	0.152	0.153	0.024	0.000	0.097	0.142	0.133	0.018
N ₂ O (g N/m ³)	0.000	0.005	0.039	0.102	0.010	0.000	0.006	0.035	0.082	0.007
Temperature (°C)	0.000	0.574	0.561	0.285	0.081	0.000	0.568	0.547	0.282	0.080
N ₂ O emission rate (g N/m ³ /d)	0.000	0.015	0.050	0.100	0.010	0.000	0.016	0.047	0.088	0.008

Table 4A
Statistical properties of the selected train/test data for GBR model evaluation.

Variable	Train (Sample size = 987)					Test (Sample size = 247)				
	Min	Median	Mean	Std	Var	Min	Median	Mean	Std	Var
Influent flowrate (m ³ /h)	0.000	0.169	0.211	0.143	0.020	0.051	0.163	0.211	0.132	0.017
DO (g O ₂ /m ³)	0.000	0.233	0.238	0.079	0.006	0.074	0.227	0.234	0.076	0.006
NH ₄ (g N/m ³)	0.000	0.126	0.166	0.125	0.016	0.022	0.130	0.162	0.098	0.010
NO ₃ (g N/m ³)	0.000	0.093	0.151	0.152	0.023	0.000	0.095	0.147	0.139	0.019
N ₂ O (g N/m ³)	0.000	0.005	0.037	0.096	0.009	0.000	0.005	0.044	0.109	0.012
Temperature (°C)	0.000	0.574	0.560	0.285	0.081	0.030	0.574	0.552	0.281	0.079
N ₂ O emission rate (g N/m ³ /d)	0.000	0.016	0.050	0.100	0.010	0.000	0.012	0.045	0.086	0.007

Data availability

Data will be made available on request.

References

[1] Y. Gao, A. Cabrera Serrenho, Greenhouse gas emissions from nitrogen fertilizers could be reduced by up to one-fifth of current levels by 2050 with combined interventions, *Nat. Food* 4 (2023) 170–178.

[2] H.H. Cho, V. Strezov, T.J. Evans, A review on global warming potential, challenges and opportunities of renewable hydrogen production technologies, *Sustain. Mater. Technol.* 35 (2023) e00567.

[3] R.G. Derwent, P.G. Simmonds, S.J. O'Doherty, T.G. Spain, D. Young, Natural greenhouse gas and ozone-depleting substance sources and sinks from the peat bogs of Connemara, Ireland from 1994–2020, *Environ Sci Atmos* 1 (2021) 406–415.

[4] E. Ranieri, G. D'Onghia, F. Ranieri, L. Lopopolo, S. Gregorio, A.C. Ranieri, Performance of wastewater treatment plants in emission of greenhouse gases, *Bioresour. Technol.* 404 (2024) 130897.

[5] E. Ranieri, G. D'Onghia, L. Lopopolo, P. Gikas, F. Ranieri, E. Gika, et al., Evaluation of greenhouse gas emissions from aerobic and anaerobic wastewater treatment plants in southeast of Italy, *J. Environ. Manag.* 337 (2023) 117767.

[6] S. Kollipara, T. Ahmed, A.K. Bhattiprolu, S. Chachad, In vitro and in silico biopharmaceutic regulatory guidelines for generic bioequivalence for oral products: comparison among various regulatory agencies, *Biopharm. Drug Dispos.* 42 (2021) 297–318.

[7] H. Wang, Z. Yan, X. Ju, X. Song, J. Zhang, S. Li, et al., Quantifying nitrous oxide production rates from nitrification and denitrification under various moisture conditions in agricultural soils: laboratory study and literature synthesis, *Front. Microbiol.* 13 (2023) 1110151.

[8] G. Feng, W. Sun, F. Zhang, L. Karthik, Z. Li, Inhabitancy of active *Nitrosopumilus*-like ammonia-oxidizing archaea and *Nitrospira* nitrite-oxidizing bacteria in the sponge *Theonella swinhoei*, *Sci. Rep.* 6 (2016) 24966.

[9] R. Ruser, R. Schulz, The effect of nitrification inhibitors on the nitrous oxide (N₂O) release from agricultural soils—a review, *J. Plant Nutr. Soil Sci.* 178 (2015) 171–188.

[10] M. Bahram, M. Espenberg, J. Pärn, L. Lehtovirta-Morley, S. Anslan, K. Kasak, et al., Structure and function of the soil microbiome underlying N₂O emissions from global wetlands, *Nat. Commun.* 13 (2022) 1430.

[11] B. Ma, H. Zhang, T. Huang, S. Chen, W. Sun, W. Yang, et al., Aerobic denitrification enhanced by immobilized slow-released iron/activated carbon aquagel treatment of low C/N micropolluted water: denitrification performance, denitrifying bacterial community co-occurrence, and implications, *Environ. Sci. Technol.* 57 (2023) 5252–5263.

[12] Z. Ilieva, R.A. Hamza, R. Suehring, The significance of fluorinated compound chain length, treatment technology, and influent composition on per- and polyfluoroalkyl substances removal in worldwide wastewater treatment plants, *Integr. Environ. Assess. Manag.* 20 (2024) 59–69.

[13] M. Henze, W. Gujer, T. Mino, M. Van Loosedrecht, *Activated Sludge Models ASM1, ASM2, ASM2d and ASM3*, IWA publishing, 2006.

[14] F. Hélot, T.H. Loh, D. Cheadle, Y. Gui, M. Dieudonne, An empirical study of the stochastic nature of electromagnetic field exposure in massive MIMO systems, *IEEE Access* 10 (2022) 63100–63112.

[15] Y. Gu, Y. Li, F. Yuan, Q. Yang, Optimization and control strategies of aeration in WWTPs: a review, *J. Clean. Prod.* 418 (2023) 138008.

[16] H. Xiao, K. Rasul, R. Vollgraf, Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms, *ArXiv Prepr ArXiv170807747* (2017).

[17] T. Kuroda, T. Ohtani, H. Wada, A Method for Quickly Searching Similar Waveform Patterns in Historical Process Data, 2016.

[18] L. Peterson, J. Bremer, K. Sundmacher, Hybrid modeling of the catalytic CO₂ methanation using process data and process knowledge, in: *Comput. Aided Chem. Eng.* 52, Elsevier, 2023, pp. 1489–1494.

[19] B. Gu, X. Zhang, S.K. Lam, Y. Yu, H.J.M. Van Grinsven, S. Zhang, et al., Cost-effective mitigation of nitrogen pollution from global croplands, *Nature* 613 (2023) 77–84.

[20] P. Brazzale, B. Bourbon, P. Barrucand, M. Fenelon, S. Guercini, R. Tiarca, *Wastewater Treatment in Dairy Processing-Innovative Solutions for Sustainable Wastewater Management*, 2019.

[21] M. Khalil, A.R.M. Alsayed, Y. Liu, P.A. Vanrolleghem, Machine learning for modeling N₂O emissions from wastewater treatment plants: aligning model performance, complexity, and interpretability, *Water Res.* 245 (2023) 120667.

[22] M. Khalil, A. AlSayed, Y. Liu, P.A. Vanrolleghem, An integrated feature selection and hyperparameter optimization algorithm for balanced machine learning models predicting N₂O emissions from wastewater treatment plants, *J Water Process Eng* 63 (2024) 105512.

[23] X. Xu, A. Wei, S. Tang, Q. Liu, H. Shi, W. Sun, Prediction of nitrous oxide emission of a municipal wastewater treatment plant using LSTM-based deep learning models, *Environ. Sci. Pollut. Res.* (2023) 1–20.

[24] S. Seshan, J. Poinapen, M.H. Zandvoort, J.B. van Lier, Z. Kapelan, Forecasting nitrous oxide emissions from a full-scale wastewater treatment plant using LSTM-based deep learning models, *Water Res.* 268 (2025) 122754.

[25] X. Xu, A. Wei, S. Tang, Q. Liu, H. Shi, W. Sun, Prediction of nitrous oxide emission of a municipal wastewater treatment plant using LSTM-based deep learning models, *Environ. Sci. Pollut. Res.* 31 (2024) 2167–2186, <https://doi.org/10.1007/s11356-023-31250-9>.

[26] Z. Huang, Y. Bai, H. Liu, Symmetry-inspired prediction of nitrous oxide emissions in wastewater treatment using deep learning and explainable analysis, *Symmetry (Basel)* 17 (2025), <https://doi.org/10.3390/sym17020297>.

[27] J. Yun, Y. Yu, C. Tao, M. Zhai, H. Zhang, Y. Chen, et al., Machine learning-based optimization of enhanced nitrogen removal in a full-scale urban wastewater treatment plant with ecological combination ponds, *Water Res.* 285 (2025) 123976, <https://doi.org/10.1016/j.watres.2025.123976>.

[28] H. Alnajjar, N. Al-Rousan, D. Al-Rouan, Machine Learning and Optimization Approaches for Dissolved Oxygen Prediction in River Systems, in: *1st Int. Conf. Comput. Intell. Approaches Appl* 2025, IEEE, 2025, pp. 1–6.

[29] Kurbana S, Yasmala A, Samura O, Şahina O, Kayaa GK, Atlarb K, et al. Future forecasting of dissolved oxygen concentration in wastewater treatment plants using deep learning techniques n.d.

[30] W. Ding, Z. Liu, X. Wang, Modeling of wastewater treatment and dynamic Optimization control of Dissolved Oxygen concentration in aeration tanks, *J. Environ. Eng.* 151 (2025) 4025050.

[31] G. Ye, J. Wan, Z. Deng, Y. Wang, J. Chen, B. Zhu, et al., Prediction of effluent total nitrogen and energy consumption in wastewater treatment plants: Bayesian optimization machine learning methods, *Bioresour. Technol.* 395 (2024) 130361, <https://doi.org/10.1016/j.biortech.2024.130361>.

[32] J. Jiang, X. Xiang, Q. Zhou, L. Zhou, X. Bi, S.K. Khanal, et al., Optimization of a novel engineered ecosystem integrating carbon, nitrogen, phosphorus, and sulfur biotransformation for saline wastewater treatment using an interpretable machine learning approach, *Environ. Sci. Technol.* 58 (2024) 12989–12999.

[33] H. Feng, W. Qin, G. Hu, H. Wang, Intelligent Prediction of nitrous oxide capture in designable ionic liquids, *Appl. Sci.* 13 (2023) 6900.

[34] P. Karadimos, L. Anthopoulos, Machine Learning-based energy consumption estimation of wastewater treatment plants in Greece, *Energies* 16 (2023) 7408.

[35] Y.H. Alali, F. Harrou, Y. Sun, Unlocking the potential of wastewater treatment: machine Learning based energy consumption Prediction, *Water* 15 (2023) 2349.

[36] M. Pilat-Rozek, E. Lazuka, D. Majerek, B. Szelag, S. Duda-Satarnus, G. Lagód, Application of machine Learning methods for an analysis of E-nose multidimensional signals in wastewater treatment, *Sensors (Basel)* 23 (2023).

[37] N.K. Singh, M.G. Yadav, V. Singh, H. Padhiyar, V. Kumar, S.K. Bhatia, et al., Artificial intelligence and machine learning-based monitoring and design of biological wastewater treatment systems, *Bioresour. Technol.* 369 (2023) 128486.

- [38] A.E.A. El Fels, L. Mandi, A. Kammoun, N. Ouazzani, O. Monga, M.L. Hbid, Artificial intelligence and wastewater treatment: a global scientific perspective through text mining, *Water* 15 (2023) 3487.
- [39] H.M. Mustafa, G. Hayder, S.I. Abba, A.D. Algarni, M. Mnzool, A.H. Nour, Performance evaluation of hydroponic wastewater treatment plant integrated with ensemble Learning techniques: a feature selection approach, *Processes* 11 (2023) 478.
- [40] B. Adibimanes, S. Polesek-Karczewska, F. Bagherzadeh, P. Szczuko, T. Shafighfar, Energy consumption optimization in wastewater treatment plants: machine learning for monitoring incineration of sewage sludge, *Sustain Energy Technol Assess* 56 (2023) 103040.
- [41] S. Sahu, A. Kaur, G. Singh, S.K. Arya, Harnessing the potential of microalgae-bacteria interaction for eco-friendly wastewater treatment: a review on new strategies involving machine learning and artificial intelligence, *J. Environ. Manag.* 346 (2023) 119004.
- [42] Sin G. Longterm N2O Emission Data and Deep Learning Model n.d. https://data.dtu.dk/articles/dataset/Longterm_N2O_emission_data_and_deep_learning_model/24106860.
- [43] X. Chen, A.T. Mielczarek, K. Habicht, M.H. Andersen, D. Thornberg, G. Sin, Assessment of full-scale N2O emission characteristics and testing of control concepts in an activated sludge wastewater treatment plant with alternating aerobic and anoxic phases, *Environ. Sci. Technol.* 53 (2019) 12485–12494, <https://doi.org/10.1021/acs.est.9b04889>.
- [44] A.S. Yaro, F. Maly, P. Prazak, Outlier detection in time-series receive signal strength observation using Z-score method with S n scale estimator for indoor localization, *Appl. Sci.* 13 (2023) 3900.
- [45] T. Miyato, T. Kataoka, M. Koyama, Y. Yoshida, Spectral normalization for generative adversarial networks, *ArXiv Prepr ArXiv180205957* (2018).
- [46] R. Muthukrishnan, C.K. James, The effect of multicollinearity on feature selection, *Indian J. Sci. Technol.* 17 (2024) 3664–3668.
- [47] S. Buhai, Quantile regression: overview and selected applications, *Ad Astra* 4 (2005) 1–17.
- [48] A. Guma, M. Al-Rakhami, M.M. Al Rahhal, F.R. Albogamy, E. Al Maghayreh, H. AlSalman, Prediction of COVID-19 confirmed cases using gradient boosting regression method, *Comput. Mater. Contin.* 66 (2021) 315–329.
- [49] C. Zhang, Y. Zhang, X. Shi, G. Almpianidis, G. Fan, X. Shen, On incremental learning for gradient boosting decision trees, *Neural. Process. Lett.* 50 (2019) 957–987.
- [50] B. Wang, T. Chen, Gaussian process regression with multiple response variables, *Chemom. Intell. Lab. Syst.* 142 (2015) 159–165.
- [51] J.O. Ogutu, T. Schulz-Streeck, H.-P. Piepho, Genomic selection using regularized linear regression models: ridge regression, lasso, elastic net and their extensions, in: *BMC Proc.* 6, Springer, 2012, pp. 1–6.
- [52] J. Qiu, An analysis of model evaluation with cross-validation: techniques, applications, and recent advances, *Adv Econ Manag Polit Sci* 99 (2024) 69–72.
- [53] A. Gogna, A. Tayal, Metaheuristics: review and application, *J Exp Theor Artif Intell* 25 (2013) 503–526. *J. Exp. Theor. Artif. Intell.*
- [54] K. Eggensperger, M. Feurer, F. Hutter, J. Bergstra, J. Snoek, H. Hoos, et al., Towards an empirical foundation for assessing bayesian optimization of hyperparameters, *NIPS Work. Bayesian Optim. Theory Pract.* 10 (2013) 1–5.
- [55] A. Kumar, S. Patil, A. Kovacevic, S.A. Ponnusami, Performance prediction and Bayesian optimization of screw compressors using Gaussian process regression, *Eng. Appl. Artif. Intell.* 133 (2024) 108270, <https://doi.org/10.1016/j.engappai.2024.108270>.
- [56] B. Abdollahzadeh, N. Khodadadi, S. Barshandeh, P. Trojovský, F. S. Gharehchopogh, E.-S.M. El-kenawy, et al., Puma optimizer (PO): a novel metaheuristic optimization algorithm and its application in machine learning, *Clust. Comput.* (2024) 1–49.
- [57] Z. Wang, Q. Wang, Z. Liu, T. Wu, A deep learning interpretable model for river dissolved oxygen multi-step and interval prediction based on multi-source data fusion, *J. Hydrol.* 629 (2024) 130637, <https://doi.org/10.1016/j.jhydrol.2024.130637>.
- [58] J. Wu, Z. Wang, J. Dong, Z. Yao, X. Chen, H. Fan, Multi-step ahead dissolved oxygen concentration prediction based on knowledge guided ensemble learning and explainable artificial intelligence, *J. Hydrol.* 636 (2024) 131297, <https://doi.org/10.1016/j.jhydrol.2024.131297>.
- [59] B. Panigrahi, S. Razavi, L.E. Doig, B. Cordell, H.V. Gupta, K. Liber, On robustness of the explanatory power of machine learning models: insights from a new explainable AI approach using sensitivity analysis, *Water Resour. Res.* 61 (2025) e2024WR037398.
- [60] A.M. Salih, Z. Raisi-Estabragh, I.B. Galazzo, P. Radeva, S.E. Petersen, K. Lekadir, et al., A perspective on explainable artificial intelligence methods: SHAP and LIME, *Adv. Intell. Syst.* 7 (2025) 2400304.
- [61] A. Willie, Explainable AI (XAI): Enhancing Transparency and Interpretability in Machine Learning Models, 2024.
- [62] P.P. Thulasiram, Explainable Artificial Intelligence (Xai): Enhancing Transparency and Trust in Machine Learning Models, 2025.
- [63] L. Cummins, A. Sommers, S. Mittal, S. Rahimi, M. Seale, J. Jabour, et al., Explainable anomaly detection: Counterfactual driven what-if analysis, in: *2024 IEEE 6th Int. Conf. Cogn. Mach. Intell., IEEE*, 2024, pp. 144–151.
- [64] Z. Li, X. Cao, X. Gao, K. Tian, K. Wu, M. Anis, et al., Simulating the unseen: crash prediction must learn from what did not happen, *ArXiv Prepr ArXiv250521743* (2025).
- [65] R. Graziosi, M. Ronzani, A. Buliga, C. Di Francescomarino, F. Folino, C. Ghidini, et al., Generating multiperspective process traces using conditional variational autoencoders, *Proc. Sci.* 2 (2025) 8. *Proc. Sci.*
- [66] V. Viswan, N. Shaffi, M. Mahmud, K. Subramanian, F. Hajamohideen, Explainable artificial intelligence in Alzheimer's disease classification: a systematic review, *Cogn. Comput.* 16 (2024) 1–44.
- [67] V. Vimbi, N. Shaffi, M. Mahmud, Interpreting artificial intelligence models: a systematic review on the application of LIME and SHAP in Alzheimer's disease detection, *Brain Informatics* 11 (2024) 10.
- [68] S. Lundberg, A unified approach to interpreting model predictions, *ArXiv Prepr ArXiv170507874* (2017).
- [69] P. Dey, S. Roy, J.R. Bathi, A. Mishra, Evaluation of uncertainty in stream flow prediction using Monte Carlo simulation for watershed-scale hydrological modeling, *J. Hydrol. Eng.* 29 (2024) 5023026.
- [70] S. Hwangbo, R. Al, X. Chen, G. Sin, Integrated model for understanding N2O emissions from wastewater treatment plants: a deep learning approach, *Environ. Sci. Technol.* 55 (2021) 2143–2151.