



Cite this: DOI: 10.1039/d6ew00014b

Hybrid CEEMDAN–deep learning models for COD prediction in wastewater treatment plants

Saeed Samadianfard,^{id} *^{abc} Egemen Aras,^d Damla Yılmaz Çelik,^d
Mohammad Taghi Sattari^{id} ^{abe} and Orhan Gündüz^c

Precise and timely chemical oxygen demand (COD) prediction is key to regulatory compliance and process control in a full-scale wastewater treatment, but laboratory assays and sluggish online measurements hinder responsiveness. Thus, a deployment-conceived hybrid pipeline for near-term COD predictions is proposed in this study. This method introduces a deployment-oriented hybrid framework that integrates complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN), gated recurrent units (GRU), and a multi-objective observer-teacher-learner-based optimizer (MOOTLBO) for near-term COD forecasting. Daily measurements from the Gebze municipal wastewater treatment plant (Türkiye) between 2019 and 2023 were collected from regularly available measured data together with total suspended solids, total nitrogen, total phosphorus, and pH with statistically enlightened lags to encode process memory. Intrinsic mode functions were obtained from the COD series with the application of CEEMDAN, and the intrinsic mode functions were concatenated with exogenous features and forwarded to recurrent learners. Long short-term memory (LSTM) and gated recurrent units (GRU) were studied as two potential configurations in the study. Hyperparameters were optimized with a multi-objective observer-teacher-learner optimizer (MOOTLBO) in order to balance fidelity, parsimony, and training cost. Evaluation occurred on a held-out test set (80/20 partition) with RMSE, MAE, R^2 , MAPE, MBE, and KGE metrics. Across alternatives, decomposition-then-learn decreased error compared with non-decomposed baselines. A CEEMDAN-GRU variant turned out to be the most accurate configuration, which reached good performance with RMSE = 2.84 mg L⁻¹, MAE = 2.28 mg L⁻¹, R^2 = 0.9999, MAPE = 0.46%, MBE = 1.47 mg L⁻¹, and KGE = 0.9974. In an optimized CEEMDAN-GRU-MOOTLBO configuration, similarly strong accuracy (RMSE = 4.19 mg L⁻¹, MAE = 2.66 mg L⁻¹, R^2 = 0.9999, MAPE = 0.52%, MBE = 0.16 mg L⁻¹, KGE = 0.9997) was obtained with excellent bias control. In view of accuracy, stability, and runtime, the CEEMDAN-GRU family represents the most suitable choice for real-time soft sensing of COD, facilitating proactive control and early warning with the presence of changing influent conditions in challenging plant practice.

Received 5th January 2026,
Accepted 23rd April 2026

DOI: 10.1039/d6ew00014b

rs.li/es-water

Water impact

This study advances real-time wastewater management by enabling accurate short-term prediction of chemical oxygen demand using routinely measured plant data. The proposed CEEMDAN-GRU-MOOTLBO framework supports faster operational decisions, early detection of process disturbances, and improved effluent quality control. By reducing reliance on delayed laboratory measurements, it helps protect receiving waters, improve compliance, and promote sustainable treatment plant operation.

1. Introduction

With more stringent effluent quality limits and growing risks posed by environmental and economic conditions, wastewater treatment plants (WWTPs) will require continuous monitoring and control techniques capable of responding to changing environmental conditions.^{1,2} Therefore, the employed instruments and techniques must produce reliable, robust, and cost-effective measurements, with the operators requiring controls capable of predicting disturbances instead of only reacting to them.^{3,4} In municipal

^a Department of Water Engineering, University of Tabriz, Tabriz, Iran.
E-mail: s.samadian@tabrizu.ac.ir

^b Water Sciences and Hydroinformatics Research Center, Khazar University, Mahsati str. 41, AZ 1096, Baku, Azerbaijan. E-mail: ssamadianfard@khazar.org

^c Department of Environmental Engineering, Izmir Institute of Technology, Izmir, Türkiye. E-mail: saeedssamadianfard@iyte.edu.tr

^d Department of Civil Engineering, Bursa Technical University, Bursa, Türkiye

^e Department of Agricultural Engineering, Ankara University, Ankara, Türkiye



wastewater treatment applications, real-time information about chemical oxygen demand (COD) is central to this strategy.⁵ COD refers to the potential organic load that will decrease ambient dissolved oxygen levels in receiving waters and adversely affect aquatic ecosystems. It also represents the amount of the equivalent oxygen required for chemically oxidizing the dissolved and suspended organic load under standardized conditions.⁶ Hence, the measurement and estimation of variations in COD are critical in maintaining treatment effectiveness and meeting the regulations. However, treatment plants are subject to variations in influent quality and volume caused by upstream production and operating conditions. In pulp and paper plants, among others, the eventual wastewater loads can arise from several process units, including paper machines, filtrates, and debarkers. Hence, the composition and flow can have abrupt changes with delays that feed forward through biological and physicochemical processes.⁷ These variations make process control difficult since the current state of the effluent relies on current loadings and the recent history of operation. Typical COD measurements exacerbate the difficulties: laborious and delayed measurements in the laboratory, and operational problems of rapid online analyzers, such as fouling, which may require the probes to be frequently maintained and recalibrated in harsh environments.^{5,8} Most instruments report new COD values only at intervals of several hours, which can miss sharp changes in water quality and induce corrective action with too-long delays. These limits facilitate the implementation of a secondary, data-driven approach that converts widely available online measurements into timely estimates and short-term predictions of COD. By analyzing the temporal correlation of easily accessible variables such as total suspended solids (TSS), total nitrogen (TN), total phosphorus (TP), and pH with future COD values, a forecasting system can deliver near-real-time insights in the midst of regular operations and unexpected disruptions.⁵⁻⁸

Predicting effluent COD on a reliable basis is now very critical to effective treatment plant control as well as early analytics and compliance assurance.^{10,11} WWTP data are messy, nonlinear, nonstationary and auto-correlated. The lagged dependencies between upstream operations and downstream quality overlap any actionable signals unless models are generated from process dynamics.^{1,7} Reviews of data-based plant analytics identify a significant need for more capable and operable prediction tools. Corominas *et al.*¹ examined computerized methods for improving WWTP operation, demonstrating the two-decade dominance of artificial neural networks (ANNs), principal component analysis (PCA), fuzzy logic, clustering, and importantly, the contending barriers keeping many tools out of routine application. Their review noted distinct translational gaps between potential algorithms and integrated decision support, struggles with practical implementation because of integration, and discourse related to the algorithms.

Early partial least squares (PLS) and ANN-based approaches improved WWTP monitoring but often lacked adaptability and operational integration. Mujunen *et al.*¹¹ studied the performance of activated sludge with partial least squares regression (PLSR) in pulp-and-paper mill settings to investigate how existing sensor suites during normal operation indicate process drift and bulking. The authors emphasized the endemic absence of influent-specific as well as microbiological descriptive data while advocating smarter variable selection to capture biological states in WWTP operation. They concluded that the approach drift remains predictable while the purification efficiency and effluent characteristics remain under described by standard variables.

Teppola *et al.*¹² addressed the aging of models through a Kalman filter that updates the coefficients of multiple linear regression (MLR), principal component regression (PCR), and PLS constructs recursively using sequential plant data. The emphasis changed from pure predictive accuracy to stability as processes drift and inputs change. Their data, tested and scored through Q^2 as a function of update lag, indicates that through coefficient updating, forecasting robustness improves over time. Newhart *et al.*² provided a review of data-driven performance assessment methods for WWTPs, placing emphasis on non-stationarity, autocorrelation, and co-correlation, while warning of the difficulties that a method may not respect the data structure. They also noted the data quality constraints that neural networks often encounter. Finally, they declared that a singular data assessment method would not find a universal solution. Therefore, their recommended practical guidance in the review declared that credible COD prediction was dependent on statistics-aware process pipelines and statistics-aware individuals in charge of plant operations. Güçlü and Dursun¹³ developed and trained three back-propagation ANN models that predicted effluent COD, suspended solids (SS), and mixed liquor suspended solids (MLSS) at a full scale, demonstrating proper ANN architecture selection through rigorous iterative training/testing. They emphasized that ANNs can be used effectively and practically when signals are nonlinear and multivariate. Reported COD prediction errors included a low RMSE of 3.23 mg L^{-1} , MAE = 2.41 mg L^{-1} , and MAPE = 5.03% with sufficient evidence to suggest the usefulness of ANNs to predict COD for routine monitoring and control. Ay and Kisi⁹ used *k*-means clustering with a multilayer perceptron (MLP) to model COD, using daily records of SS, pH, temperature, discharge, and COD obtained from a municipal system. Clustering enabled regime behavior to be partitioned prior to supervised learning and resulted in improved fit during mixed conditions. *K*-means-MLP yielded better fits than MLR, standalone MLP, radial basis function (RBF), generalized regression neural network (GRNN), and a two-variant adaptive neuro-fuzzy inference system (ANFIS) model on all fit metrics. Qiu *et al.*¹⁴ developed a remotely sensed deep-learning soft sensor using stacked autoencoders, supplemented with a genetic algorithm to



optimize hidden-layer sizes for WWTP variables that are prohibitively expensive to measure online. The approach impressed with better prediction and generalization results than conventional baselines across operating regimes. Liu¹⁵ offered a “just-in-time” learning adaptive framework, designed around a relevant-vector machine and supplemented with a JADE evolutionary strategy to facilitate parameter selection and a moving-window scheme for adaptivity. This adaptive learning framework was suited for locally linearizable drifting regimes, which were typical of plant operations. Similarly, a dynamic NARX model was developed by Yang *et al.*¹⁶ and was then enhanced through a PCA-NARX model, which optimized both time delays and training algorithms in order to achieve process memory representation. The dynamic models produced improvements compared to static ANNs on effluent models, and are designed for real-time adjustment. Their reported accuracy is high for both COD (RMSE = 2.9 mg L⁻¹) and TN (RMSE = 0.8 mg L⁻¹). A genetic algorithm was combined with a deep belief network (DBN) for dimensionality reduction and architecture tuning in order to monitor papermaking WWTPs. Niu *et al.*¹⁷ stated that deep feature extraction and evolutionary search methods are more suitable approaches for more complex data associated with plants. Compared with DBN and BPNN architectures, the proposed GA-DBN analysis improved R^2 by 1.71–1.86% and by 5.21–9.32% while using fewer variables. Dynamic modeling was employed in the variable importance in projection (VIP-DBN) to further include variable importance projections and adopted in a Bayesian-network to capture nonlinearity, uncertainty, and time-variation, while still trimming the inputs. Applied to a full-scale papermaking WWTP, the VIP-DBN was found to improve reliability and lower cost, and against PLS, ANN and DBN modelling, the approach yielded an improved R^2 for nitrate prediction (such as +34.36% being one example with respect to PLS). Wang *et al.*³ established an interpretable ML stacking framework using random forests and deep neural network checks, partial-dependence plots, variable-importance metrics, and marked time-lag handling to drive and relate operations to effluent quality. In addition to generating predictive functions, it revealed driven variables in practical applications, such as TSS distribution in aeration basins and return-sludge rerouting. In their study, variables identified as the greatest predictors of TSS and PO₄ were the temperature of the influent and TSS in the various basins. Wang *et al.*¹⁰ took a step into practical application, comparing nine machine learning algorithms to predict effluent COD in addition to treatment-plant data time series processing. Results were reported with attention to modeling applications, with the COD outcome having a mean absolute percentage error (MAPE) of 7.34%, RMSE of 1.29, and R^2 of 0.92 using the KNN algorithm. Optimized models were also able to predict TN, TP, and effluent pH with similar practical performances. Manav-Demir *et al.*⁷ also utilized short- and long-term forecasting, compared

multiple machine learning methods for their effluent prediction of a biological nutrient removal system using two years of full-scale data, and developed policies for benchmarking performance in the method of data and key process variables. Their comparative perspective also highlighted MAPE levels of around 9% and 0.9%, respectively, and reported TN and TP MAPE levels for RF/XGBoost to be lower than 34% and 27%, respectively.

Recent ANN and hybrid convolutional neural network (CNN) studies confirmed the potential of hybrid deep learning, yet highlighted scalability and interpretability challenges. Çimen Mesutoğlu and Gök¹⁸ investigated a WWTP with feed-forward ANNs that included PCA-informed developments and operational records from daily records of nine months. They determined the architecture empirically and validated the model using MSE/MAE/ R^2 . As a pilot, the study provided a demonstration which indicated that dimension reduction would stabilize training on a wastewater treatment plant. In the best architectures, there was near-perfect agreement between observed and predicted output values (R^2 values sometimes exceeding 0.9997). Verhaeghe *et al.*⁴ focused on exploring parallel hybrid models, which train a neural network (NN) on mechanistic residuals, to define an input/output relationship that balances time spent on calibration *vs.* learning capability. Using CNN-hybrid models, they reduced nitrate error against LSTM, and for example, achieved RMSE = 1.58 mg NO₃-N per L *vs.* RMSE = 4.17 mg NO₃-N per L, respectively.

Despite rapid advances in data-driven soft sensors, most COD forecasting studies either omit explicit lag design or lack decomposition-based noise handling. Moreover, multi-objective optimization has rarely been applied to balance model fidelity and computational feasibility in full-scale WWTPs. This study addresses these gaps by developing a combination of complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN), gated recurrent units (GRU) and a multi-objective observer-teacher-learner-based optimizer (MOOTLBO) framework namely CEEMDAN-GRU-MOOTLBO that explicitly encodes temporal memory, applies decomposition-based denoising, and optimizes network efficiency for deployment.

While various hybrid decomposition-deep learning methods such as variational mode decomposition LSTM (VMD-LSTM), empirical model decomposition LSTM (EMD-LSTM), and EMD-GRU have been successfully implemented for environmental time series forecasting tasks, some methodological gaps still persist. For instance, some existing literature on decomposition-based methods has been using signal decomposition on the entire data set before training any model on the data. This has been shown to cause temporal leakage between training and testing phases. Additionally, lag selection has been carried out empirically without any statistical diagnostic tests. Similarly, the existing literature on decomposition-based methods has been using only one optimization criterion during hyperparameter tuning, namely model accuracy. However, no



attention has been given to computational efficiency or model simplicity during optimization. In light of these gaps, this study proposes a hybrid framework that is deployment-oriented and incorporates a lag selection mechanism using PACF, fold-safe CEEMDAN decomposition during training, and a multi-objective optimization strategy (MOOTLBO) to balance accuracy, simplicity, and efficiency. Such a framework is deemed to be statistically more rigorous and deployable for real-time COD prediction in full-scale WWTPs. For better understanding of the methodological contribution of this study, Table 1 highlights some of the differences between this study and some existing literature on decomposition-deep learning-based methods for water quality prediction tasks.

In summary, prior studies lay the groundwork for COD prediction frameworks, which (i) respect and incorporate the temporal structure and lag, (ii) provide trade-offs between accuracy and interpretability for control purposes, and (iii) can be scaled to full-plant scenarios. This is precisely the position of our COD-focused hybrid deep learning framework that aims for: dynamic, multivariate, control-oriented soft sensing of routine online signals (TSS, TN, TP, pH, and lag) into dependable and timely COD forecasts. Therefore, the novelty of this research focuses on a deployment-oriented, series-connected pipeline that aligns features, which have rarely been integrated for COD forecasting at a full scale. Inputs are identified using partial autocorrelation guided lag analysis so that process memory is explicitly encoded before modeling. To reduce overfitting and enhance portability, variable design and learning are separated from each other. A “decomposition-then-learning” strategy is utilized where complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) is implemented within each rolling training window, ensuring that the multi-scale structure is isolated from temporal leak in measurements, which is an implementation detail frequently omitted from studies of environmental time-series. The approach to model selection is considered multi-objective in nature and is solved by a teaching-learning-based optimizer (MOOTLBO) that simultaneously weighs predictive accuracy, model parsimony, and computational cost; this leads to compact LSTM/GRU variants with low inference cost and better use on-site. In addition to point error, calibration behavior and autocorrelation of residuals are considered to construct

forecasts actionable for the purposes of process control. Collectively, PACF-guided lag design, fold-safe CEEMDAN decomposition integrated in the training loop, and MOOTLBO-based multi-objective optimization make the proposed framework unique from other hybrid approaches in the context of decomposition and deep learning for COD prediction in a statistically robust and computationally efficient manner.

2. Materials and methods

2.1. Study site and data collection

Operational data records were acquired from a large municipal WWTP located in Gebze-Kocaeli, Türkiye (Fig. 1). The plant utilizes an extended aeration activated-sludge process for organic removal and nutrient reduction, and has a nominal processing capacity of approximately 144 000 m³ per day. Observations were limited to the aeration phase, with daily composite samples being taken from both the influent and effluent of the aeration unit from January 2019 to December 2023 to make observations related to near-term COD performance under both normal operation and upset conditions.

The dataset incorporates daily measurements of chemical oxygen demand (COD) alongside process variables typically associated with COD performance including total suspended solids (TSS), total nitrogen (TN), total phosphorus (TP), and pH. All samples were transported promptly to the plant's accredited laboratory and analyzed within two hours of collection. COD was measured and reported on the basis of the dichromate closed reflux method consistent with standard methods. This dataset provides a representative case of daily-scale operational variability in a large municipal WWTP, making it suitable for evaluating near-term COD prediction frameworks.

2.2. CEEMDAN, LSTM, GRU, and MOOTLBO methods

Time series of environmental data are often characterized by non-stationarity and noise. The complex nature of time-driven wastewater processes means samples are affected by operational shocks, diurnal variations, and weather-driven inputs that imprint multiple scales to the pollution signal. One practical strategy for mitigating this complexity is to decompose the compound signal and

Table 1 Methodological comparison between the proposed framework and recent decomposition-deep learning hybrid models used for environmental time-series prediction

Study	Decomposition	Sequence model	Lag design	Optimization strategy	Key limitation
VMD-LSTM studies	VMD	LSTM	Often empirical	Grid/manual tuning	Risk of leakage and high computational cost
EMD-GRU models	EMD	GRU	Limited lag analysis	Single-objective tuning	Focus only on accuracy
CEEMDAN-DL hybrids	CEEMDAN	LSTM/GRU	Usually fixed	Manual hyperparameters	Limited deployment considerations
This study	CEEMDAN	GRU/LSTM	PACF-guided lag selection	MOOTLBO multi-objective optimization	Designed for real-time deployment and computational efficiency





Fig. 1 Municipal WWTP facility located in Gebze-Kocaeli, Türkiye.

represent it in simpler component signals prior to modeling. The complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) can iteratively extract or decompose the pollution signal into intrinsic mode functions (IMFs) into high-, mid-, and low-frequency oscillations plus an adaptive or slowly varying residue. Unlike fixed-basis transforms (like Fourier), CEEMDAN is data-adaptive and uses noise-assisted sifting, which reduces mode mixing, retains local phase information, and allows the resultant IMFs to be modeled more specifically. Empirical evidence has shown consecutive improvements in forecast accuracy when coupling CEEMDAN with sequence learners, which should not be surprising as it allows the sequence learning approximation process to avoid strain while approximating the entire frequency spectrum and simply focus on a band of frequency data that is more homogeneous or consistent.^{19,20}

After the decomposition surface structure, gated recurrent architectures can naturally learn cross-lag dependencies. LSTM networks were created to ease vanishing gradients by adding a cell state with input, forget, and output gates, which regulate the information stream. The gates allowed the model to hold or decrement historical context. LSTM has become a workhorse for multi-scale hydrological and water quality prediction, and the literature has plenty of demonstrations that LSTM or hybrids based on them learn nonlinear lagged responses to process dynamics effectively with simply upstream signals or outcomes, or the decomposed signals.^{21–23} In CEEMDAN-based hybrids, a common theme is to use LSTM on mid-frequency IMFs, where the memory and nonlinearity are greater, and a simpler learner on very high frequency noise or a low-frequency trend, thereby matching the model capacity with temporal complexity.¹⁰

GRU exist as a simplified option, combining the gating approach with the update and reset gates. GRU are generally faster to train than LSTM because they have fewer parameters, which could also be beneficial in an ensemble with multiple models that describe the IMF effect. In most

instances, GRU are reported as achieving reasonably the same accuracy level for many sequence tasks, as LSTM, but at a lower training cost, especially in sequence tasks of intermediate dependency on length, time, and computation.^{24,25} In a decomposition-ensemble approach, GRU layers could be applied to IMFs that induce short memory dynamics, while LSTM would be suited to model long-term temporal dynamics. This is practical in real-time context, smaller GRU models and prototypes to maximize speed, as weights were updated much more frequently, while LSTM was included when increasing speed to the same logic was beneficial.

Even though the architectures can be identical, performance depends on hyperparameter selections and, when in a hybrid setting, the particular form of the input set. When design is viewed as a multi-objective search problem, our operational goals adopt a shared discipline. Accuracy must be balanced with parsimony and computational cost. The multi-objective-observer-teacher-learner-based optimization (MOOTLBO) algorithm performs the relevant balancing necessary. MOOTLBO builds on teaching-learning paradigms and clusters the search around three roles: an “observer” mechanism, which encourages exploration, a “teacher” scenario that moves candidate solutions towards the currently best exemplars, and a “learner” phase using pairwise interactions to further drive refinements. In its multi-objective setting, it maintains an external archive of non-dominated solutions and uses the leader selection method to maintain diversity while searching for candidate solutions, avoiding premature convergence that produces complementary balances of error and complexity.^{26–28} In environmental predictive applications, these features have been leveraged to co-optimize hyperparameters and input subsets, typically in straightforward LSTM/GRU-derived variants that exhibit stable generalization and low inference costs.^{26–28}

The methodological synergy occurs after these elements are integrated together. CEEMDAN restructures the target into frequency-localized IMFs converting the nonstationary



forecasting exercise into several simpler forecasting efforts. LSTM and GRU learn the IMF-specific mappings of process covariates and histories of the component to the corresponding COD contributions that match architectural capacity with the temporal scale. Finally, MOOTLBO searches through hyperparameters (*i.e.*, units, learning rate, batch size) and feature selections to find configurations with minimal error while minimizing parameters and training time.

2.3. Input variables and preprocessing

In the time-series perspective taken on input design, short histories of plant signals were explicitly included prior to modeling. The feature set derived from TSS used lags selected from PACF analysis to include near-term process memory while keeping dimensionality low. Additional variants were considered with TN, TP, and pH. However, the following pipeline description seeks to distance some of the lag configuration, so the preprocessing is clear. The records were synchronized to screen for a common quality assurance. The original data set had a moderate proportion of missing data due to some instances of sensor downtime and sampling gaps during the monitoring period from 2019 to 2023. In total, 11.97% of the data set is missing and interpolated, while 88.03% is actual data. The proportion of missing data is relatively consistent across variables, ranging from 11.9% to 12.1%. In order to ensure temporal continuity in the data set for time series modeling purposes, missing data are reconstructed through a Kalman filter-based state space interpolation method. This is commonly applied to environmental monitoring data sets as it maintains temporal features without causing distortion to the underlying statistical structure of the data. Since actual data make up the majority of the data set, interpolation is unlikely to impact the outcome of any predictive modeling work. Moreover, continuous scaling was applied to obtain similar numerical ranges and stabilize optimization; a Min–Max transform so that each predictor was in the range of [0, 1] over the training domain, including the target variable COD, so that the network learned with consistent scales. Further, the parameters for scaling were generated from the training subset and, after model fit, were used to scale the test subset likewise; the inverse transform was kept to back-transform the predictions to the original units for interpretation and scoring. This means that the influences of magnitude differences between input variables were mitigated along with the saturation of the activation during training.

To accommodate uncertainty and multi-scale dynamics of COD, the methods of noise-assisted empirical mode decomposition were used as a preprocessing step. As such, the CEEMDAN was completed on the normalized COD time series, which ultimately produced, or decomposed, the target into a set of IMFs that separated the target into oscillatory components from high to low characteristic frequencies and a residual trend. These IMFs were arranged together column-wise and concatenated to the temporal lagged

features to produce an augmented regressor matrix. When discussed in practical terms, this creates a feature space that surrounds the learner with exogenous predictors (the lagging vector) and an endogenous description of the internal structure of COD (the IMFs). Finally, the design matrix was reshaped to the expected three-dimensional tensor format for recurrent layers of: samples \times time steps \times features. As temporal information has already been encoded through the lags, each sample is related to a time step, with a multi-feature vector; therefore, the GRU operates over the feature axis as opposed to a long explicit sequence. An 80/20 split of the data was performed to produce training and test subsets, using a fixed random number seed to enable reproduction per run. Hyperparameter values selected by MOOTLBO used predictive accuracy, architectural parsimony, and training cost as maximization settings. The optimizer returns the number of units, learning rate, batch size, and the number of epochs; the architecture is compiled with these values, and an Adam optimizer and mean squared error loss. Fig. 2 illustrates the practical steps of the data preprocessing. Each box represents a main step of data handling prior to model training.

2.4. Model development

The modeling framework employs a “decompose-then-learn” strategy, in which two closely-related variants were applied: CEEMDAN–GRU–MOOTLBO and CEEMDAN–LSTM–MOOTLBO. Both models share the same data flow and differ only in the recurrent cell that is utilized. The code was written in Python using the scikit-learn package for preprocessing and metrics, the PyEMD package of CEEMDAN for decomposition, and TensorFlow/Keras for the neural modeling. Data ingestion begins with reading in the curated daily dataset. The exogenous design matrix is constructed from a small, lagged set of the input parameters, motivated by PACF, while the supervised target is COD. Next, feature scaling is applied to both the predictors and the target using the MinMaxScaler to map them into [0, 1]. In the code scaffolding stage, scalers are fit before train/test split for convenience; in any operational training, these transforms should be fit on the training part and applied on the holdout to eliminate the possibility of any temporal leakage. In order to expose the multi-scale structure in COD, CEEMDAN from the PyEMD package is applied to the normalized COD target. CEEMDAN provides an array of IMFs and recurrent layers expect inputs of shape similar to (num_samples, time_steps, num_features).

As short histories are already encoded as lagged features, each daily observation is a single time step with a multi-feature vector. Next, to preserve temporal dependency and avoid data leakage, the dataset was split chronologically into training (80%) and testing (20%) subsets rather than randomly shuffled partitions. Hyperparameter selection is organized around an MOOTLBO by a placeholder class that returns a stable set of tuned values: number of units, epochs, batch size, and



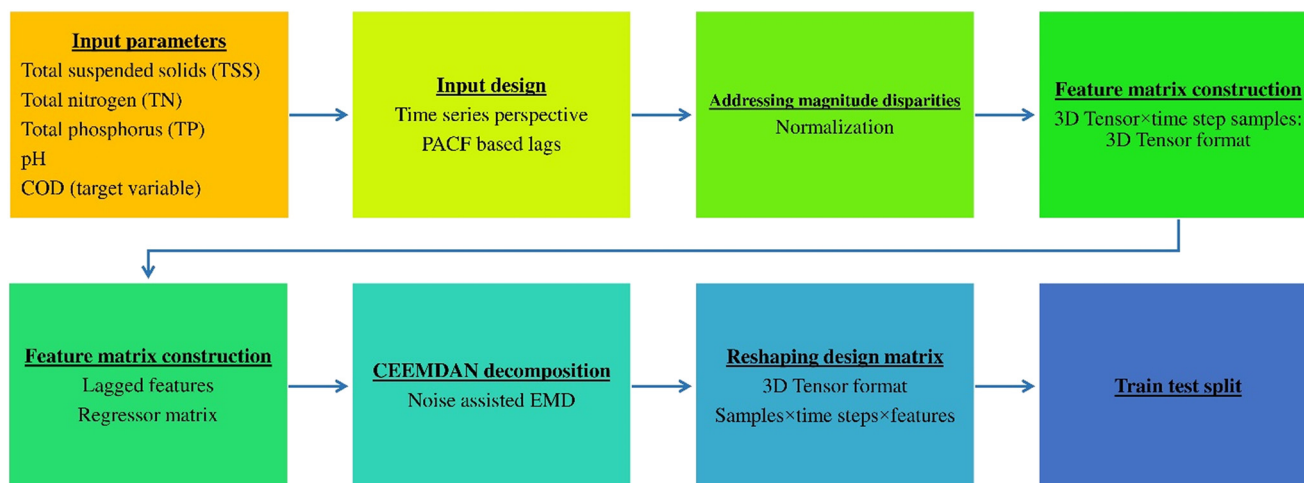


Fig. 2 Flowchart of the data preprocessing.

learning rate. For CEEMDAN-GRU-MOOLBO, a sequential model is created with one GRU layer using ReLU activation, and an input shape inferred from our training tensor is followed by a linear dense layer for scalar regression. The only difference for CEEMDAN-LSTM-MOOLBO is the GRU, which replaces LSTM, and all other settings are identical. The Adam optimizer is set to the learned rate that is selected by the optimizer, and mean squared error (MSE) is used as the training loss. Although the hyperbolic tangent is frequently used in gated units, we retain ReLU in this case to be consistent with the code and to help reduce the risk of saturation when using Min-Max scaled input. Training occurs until the specified epochs and batch_size are reached, and a validation stream is included for checking for generalization. Model evaluation is completed on both training and test partitions. Predictions are generated, and the values are returned to the original measurement scale by reversing the target scaler. The overall flowchart of the developed methodology is shown in Fig. 3.

2.5. Performance evaluation

Several indicators were used to determine performance: root mean square error (RMSE), mean absolute error (MAE), coefficient of determination (R^2), mean absolute percentage error (MAPE), mean bias error (MBE), and Kling-Gupta efficiency (KGE), which are defined as follows:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (P_i - O_i)^2}{n}} \quad (1)$$

$$\text{MAE} = \frac{\sum_{i=1}^n |P_i - O_i|}{n} \quad (2)$$

$$R^2 = \left[\frac{\sum_{i=1}^n (O_i - \bar{O})(P_i - \bar{P})}{\sqrt{\sum_{i=1}^n (O_i - \bar{O})^2 \sum_{i=1}^n (P_i - \bar{P})^2}} \right]^2 \quad (3)$$

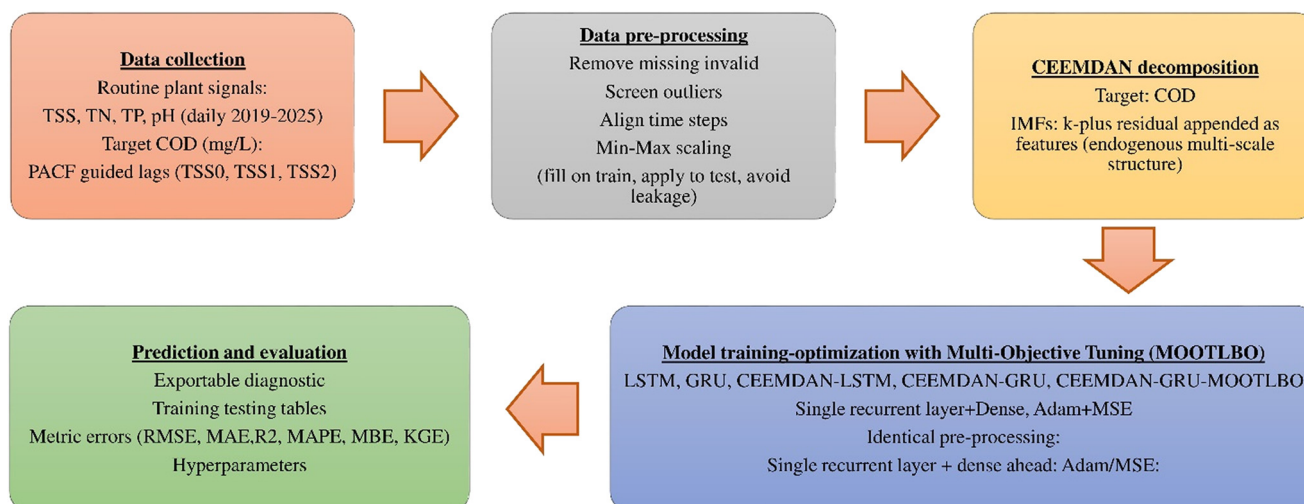


Fig. 3 Flowchart of the developed methodology.



$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{P_i - O_i}{P_i} \right| \times 100 \quad (4)$$

$$\text{MBE} = \frac{1}{n} \sum_{i=1}^n (P_i - O_i) \quad (5)$$

$$\text{KGE} = 1 - \sqrt{(r^2 - 1)^2 + (\beta^2 - 1)^2 + (\gamma^2 - 1)^2} \quad (6)$$

where O_i is the observed COD, \bar{O}_i is the average COD value, P_i is the modeled COD, \bar{P}_i is the average of the modeled COD values, and n is the number of observations.

3. Results and discussion

3.1. Results

This section presents the impact assessment and interpretation of the proposed hybrid deep learning framework for COD forecasting in wastewater treatment plants. The analysis follows a model class that captures temporal dependence and non-stationarity in indicators that are monitored routinely, using lagged histories of TSS, TN, TP, and pH as inputs. Lag structures were determined using the partial auto-correlation function (PACF) to represent memory properties relevant to the COD process without excessively inflating the dimensionality of the model. The model class includes recurrent networks that stand alone (LSTM, GRU), decomposed versions based on CEEMDAN-LSTM, CEEMDAN-GRU, and multi-objective optimized configurations (CEEMDAN-LSTM-MOOLBO, CEEMDAN-GRU-MOOLBO), where the MOOLBO optimizes a joint function of predictive error and architectural parsimony.

Table 2 describes statistical summaries and diagnostic tests for the COD time series, which are necessary before model development and forecasting to help dictate the statistical properties of the COD time series. The statistical summaries identify central tendency and variance in COD concentrations, while the subordinate stationary and normality tests inform suitability in developing the advanced hybrid deep learning models.

In the overall dataset, the average COD concentration was 644.99 mg L⁻¹, with a mean variance of 134277.4 and a standard deviation of 366.44 mg L⁻¹, showing shifts over time. The variability suggests that temporal dynamics of the

wastewater would be expected to be significant, due to changes in influent load, discharge and operational conditions of the treatment work. The minimum and maximum COD observations of 87.76 mg L⁻¹ and 2150.83 mg L⁻¹, respectively, also emphasize the variability within the data and could point to the significance of events or episodic surges in pollutants influencing the time series pattern significantly. The distributional shape is shown subsequently in a skewness of 1.48, and a kurtosis value of 2.36, which is indicative of a moderately right skewed and leptokurtic distribution. This indicates frequent observations of high COD concentrations that would be expected from a normal distribution, as well as that the extreme values would have the most impact on the mean. Therefore, non-linear modeling techniques would be best justified such that temporal irregularity and asymmetric temporal dependencies could be captured.

To investigate the time series properties, a series of statistical tests were conducted. The augmented Dickey-Fuller (ADF) test resulted in a statistic of -5.57 and a p -value of 1.51×10^{-6} , confirming stationarity of the data by rejecting the null hypothesis of a unit root. By comparison, the KPSS test resulted in a statistic of 0.95 and a p -value of 0.01, indicating mild, non-stationary data. This conflict is expected, given the mixed deterministic and stochastic trends common in environmental datasets. The independence and normality assumptions were also tested. The Jarque-Bera test (statistic = 581.40, $p < 0.001$) clearly rejects normality assumptions and is consistent with COD data being positively skewed. The Ljung-Box tests (lag = 20, statistic = 5642.64, $p < 0.001$) produced very strong evidence of autocorrelation, indicating strong temporal dependencies, thus providing strong support for using recurrent neural networks that utilize memory states.

Fig. 4 displays the PACF plots for some of the key input parameters and the COD time series values. PACF plots were used to investigate the underlying lag structure and temporal dependencies existing within each variable, which can inform the selection of the appropriate input features for time series forecasting.

The PACF analysis showed significant lags in TSS0, TSS1, TSS2, TSS7, TN0, TN1, TN2, TP0, TP1, TP2, pH0, pH1, pH2, COD1, and COD2. The PACF plot had significant and substantial spikes in COD at the first and second lags, indicating strong autocorrelation and persistence in this time

Table 2 Summary statistics and test results for the COD time series

Statistics/test	Value	Statistics/test	Value
Mean	644.99	ADF p -value	1.51×10^{-6}
Variance	134277.4	KPSS test statistic	0.95
Standard deviation	366.44	KPSS p -value	0.01
Maximum	2150.83	Jarque-Bera test statistic	581.40
Minimum	87.76	Jarque-Bera p -value	0
Kurtosis	2.36	Ljung-Box test statistic (lag = 20)	5642.64
Skewness	1.48	Ljung-Box p -value (lag = 20)	0
ADF test statistic	-5.57		



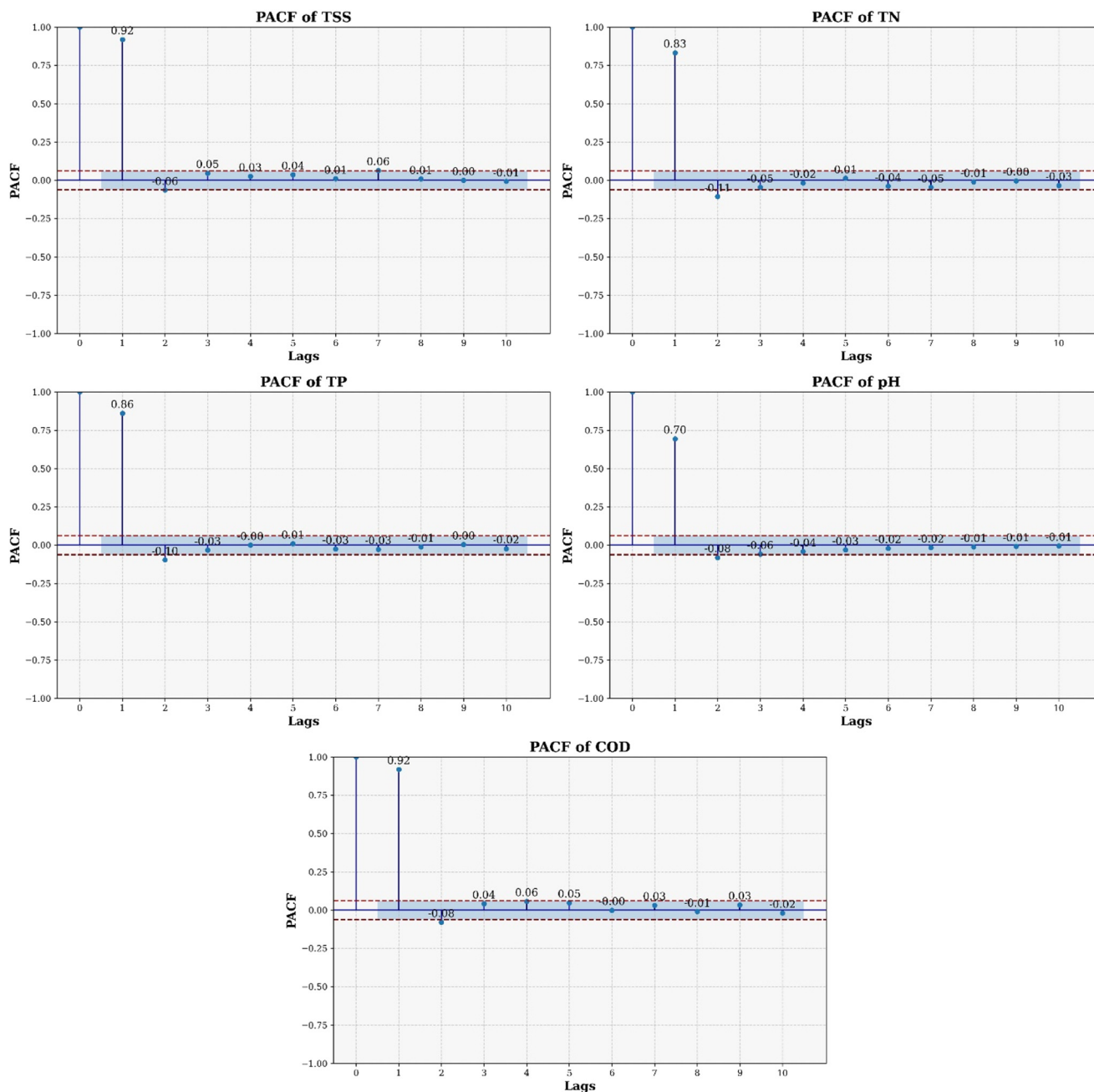


Fig. 4 PACF plots of input parameters and COD.

series. This suggested that the current COD concentration is highly reliant on the prior values, which is a property of environmental and process-based time series. The significant lags in COD1 and COD2 confirmed the need to include past COD values in the predictive modeling framework as they capture the self-reinforcing time-lag patterns that influence the COD concentration in this wastewater. For the secondary parameters, the PACF plots of TSS and TN had moderate and significant lag relationships to the second and seventh lags, respectively. The moderate lags indicated that suspended solids and nitrogen compounds exert delayed influence on COD, which is in accordance with the sequential nature of

biological oxidation and nutrient conversion processes occurring within wastewater. The PACF plot of TP indicated some lag effects, though they were not as obvious, and showed intermittent patterns likely associated with operational variability or episodic phosphorus discharges. In contrast, the pH series exhibited reduced yet nevertheless meaningful autocorrelation at its initial two lags, indicating that even relatively minor fluctuations in the pH value can affect fluctuations in COD values indirectly by modulating microbial activity and reaction rates. Fig. 5 presents the original series, IMFs and the residual component derived from the decomposition of the COD time series using the



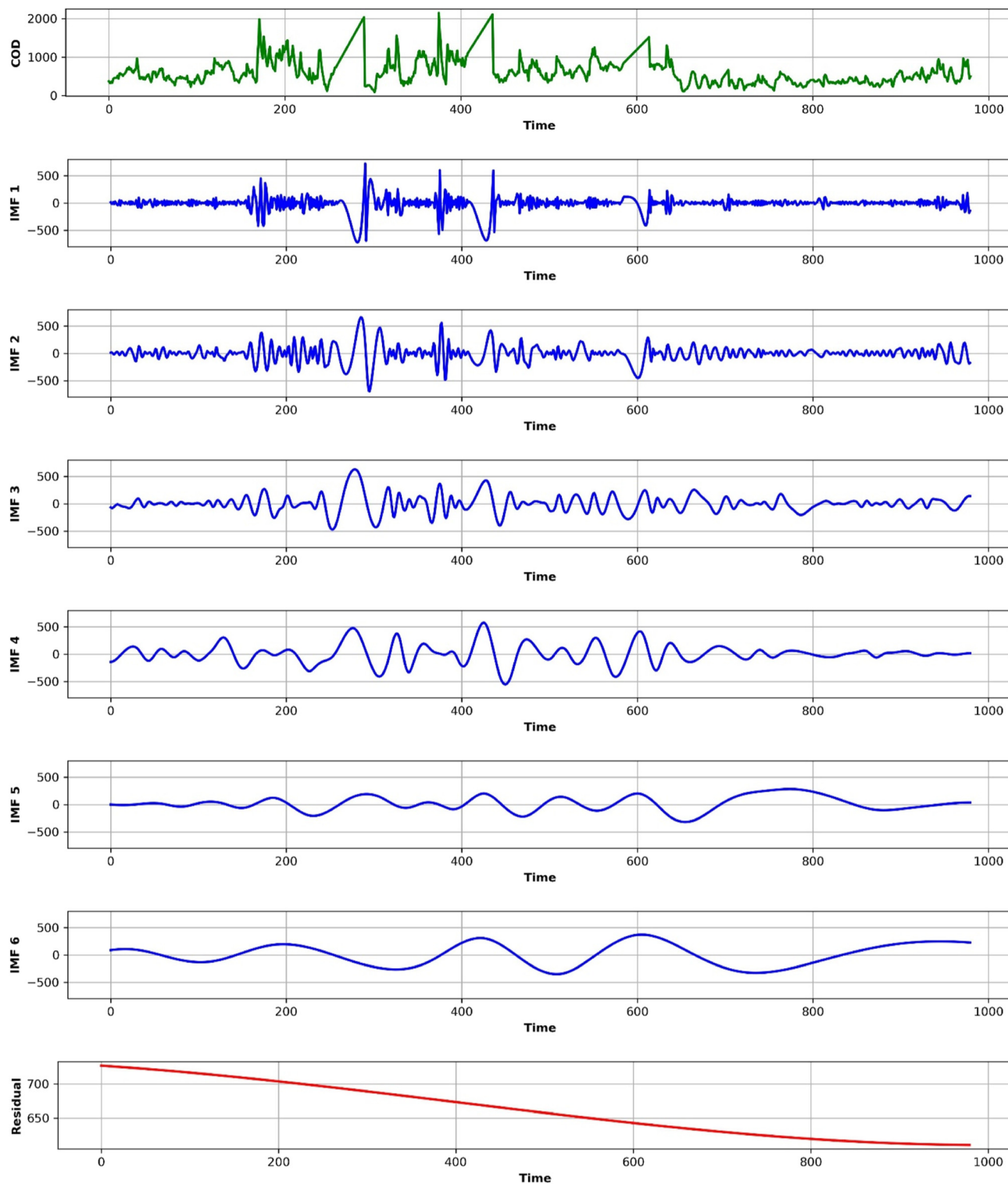


Fig. 5 IMFs and residual of COD values.

CEEMDAN approach. Decomposition is an essential preprocessing step in hybrid deep learning frameworks because it allows for the extraction of latent temporal patterns that may enhance the technical ability of the model to capture both short-term movements and long-term trends associated with changing COD dynamics.

The initial IMFs in the upper portion of Fig. 5 display oscillations that occur at high frequencies, showing rapid shifts, and exhibiting what can be viewed as sharp peaks. The IMFs indicate shifts in COD at shorter time scales that relate to sudden shifts in the influent load, operational issues, or interruptions in discharge events from residential



or commercial properties. The presence of energy and noise-like patterns in the initial IMFs suggests that COD will shift quickly and in ways that will be difficult to predict if not captured in isolation. The CEEMDAN method separates the order of temporal shifts in the COD, so that later modeling approaches can explore high-frequency shifts separately from more linear relationships in process stability. The CEEMDAN decomposition separates the COD signal into various intrinsic mode functions that reflect different oscillatory modes at different time scales. The high-frequency modes are related to rapid changes in COD concentration, which are often related to short-term changes in influent composition, such as industrial discharge pulses, sudden changes in flow rate, or stormwater infiltration into sewer systems. The medium-frequency modes are related to intermediate temporal changes that could be related to operational cycles within the treatment process. The low-frequency modes and the residue are related to long-term changes in COD concentration, which could be related to long-term changes in plant loading, seasonal changes in influent composition, or operational changes within the treatment process. The IMFs at lower positions in the figure show increased regularity, smoothness, and periodic behavior in relation to shifts in the concentration of COD at medium time scales. These IMFs may be associated with cyclical operational processes such as daily or weekly treatment, seasonal implications of treatment, or gradual changes in compositional shifts in the influent. The waveforms demonstrated within these IMFs generally show regular shifts in amplitude and frequency, indicating a more ordered and cyclic character that contributes to overall variability in COD measurements. Conversely, the lower IMFs and the residual component exhibit low-frequency variations in the COD signal, indicating the long-term trend. The residual appears as a gradually varying curve and represents the persistent baseline pattern, which could result from routine operational practices, gradual environmental changes as a result of the treatment system, or improved treatment efficiencies over a longer time interval. This smooth trend component accounts for the equilibrium of the wastewater system and provides useful information about how the wastewater system operates in terms of collective stability and long-range evolution.

Table 3 shows the correlation coefficients and variance contributions for the decomposed IMFs of the COD time

series as derived from the CEEMDAN algorithm, which provides an in-depth quantitative interpretation of the contribution of different frequency components to the overall COD dynamics and their relationships to the original signal.

The high-frequency group (IMF1–IMF3) identifies the least stable and short-term fluctuations of COD. IMF1 shows a negative correlation, -0.2782 , indicating that the series represents random noise or short-term fluctuations resulting from sudden changes in influent or operational fluctuations of the wastewater treatment process. IMF2 has a positive but weak correlation, 0.2927 , suggesting little overall relationship to COD, while IMF3 has the largest correlation of the high-frequency IMFs (0.4983), suggesting that it still retains important short-term temporal processes relevant to COD. Overall, the high-frequency group is responsible for about 47% of the variance, indicating that nearly half of the variability in COD can be attributed to fast changing, irregular processes. The medium-frequency group (IMF4–IMF6) has moderate to strong positive correlations, 0.3331 to 0.5004 , and together accounts for about 50% of the variance. The medium-frequency IMFs represent periodic recurring structures in the COD time series, patterns related to cyclical operational actions: daily or weekly variations of wastewater inflow or treatment performance. In particular, IMF4 correlates closely to COD concentration (0.5004 and 23.81% variance). IMF5 and IMF6 also maintain similar relationships, capturing the process-driven periodicity inherent in wastewater treatment systems. The low-frequency group (IMF7 and the residual) captures the long-term trend and baseline, or baseline average, which underlies the COD signal. IMF7, for example, has a correlation of 0.4953 while explaining 33.38% of the variance, signaling gradual, slow changes in the COD profile, which could represent longer-term operation conditions, treatment efficiencies, or environmental influences. The variable accounted for in the residual prior to approximation in CEEMDAN has a correlation of 0.2618 , indicating that CEEMDAN was successful in decomposing the COD signal. Fig. 6 summarizes the high-, medium-, and low-frequency components which are reconstructed from the COD series *versus* the original signal, demonstrating that CEEMDAN provides a rich visualization of how the decomposition extracts multi-scale temporal characteristics through wastewater dynamics.

The high-frequency component had rapid and irregular oscillations, characterized by sharp peaks and troughs indicative of short-term changes in COD concentration. These types of fluctuations are commonly associated with relatively immediate disturbances (for instance, a sudden increase in influent unduly affects COD concentration) or operational changes (for instance, making small adjustments in equipment operation) or events that are short-lived (for instance, variations in treatment processes). Overall, the high-frequency signal shows responsiveness of COD concentration to changes, volatility of wastewater characteristics, and high sensitivity to transient environments

Table 3 Correlation and variance results of COD IMFs

Group	Component	Correlation with COD	Variance (%)
High frequency (HF)	IMF1	-0.2782	14.40
	IMF2	0.2927	16.26
	IMF3	0.4983	15.88
Medium frequency (MF)	IMF4	0.5004	23.81
	IMF5	0.1995	11.66
	IMF6	0.3331	28.64
Low frequency (LF)	IMF7	0.4593	33.38
	Residual	0.2618	1.60



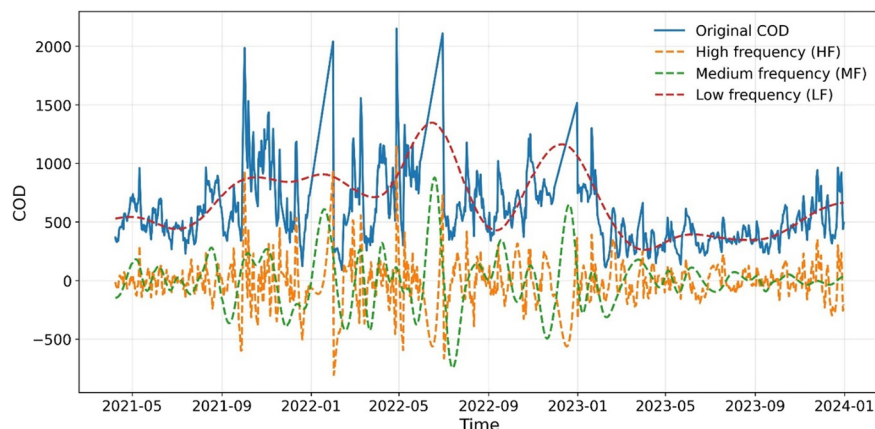


Fig. 6 High-, medium-, and low-frequency components and the original COD.

and processes. Although the oscillation signals appear noisier than the lower-frequency components, they do contain important information about the system response, which is useful for prediction, especially as the predictability horizon is short, and is modeled properly. The medium-frequency

component displays changes and more oscillatory behavior, and likely exhibits periodic behavior that reflects operation cycle or mid-range environmental effects. In this case, these fluctuations may indicate daily or weekly changes in wastewater inflow, changes in load composition (influent

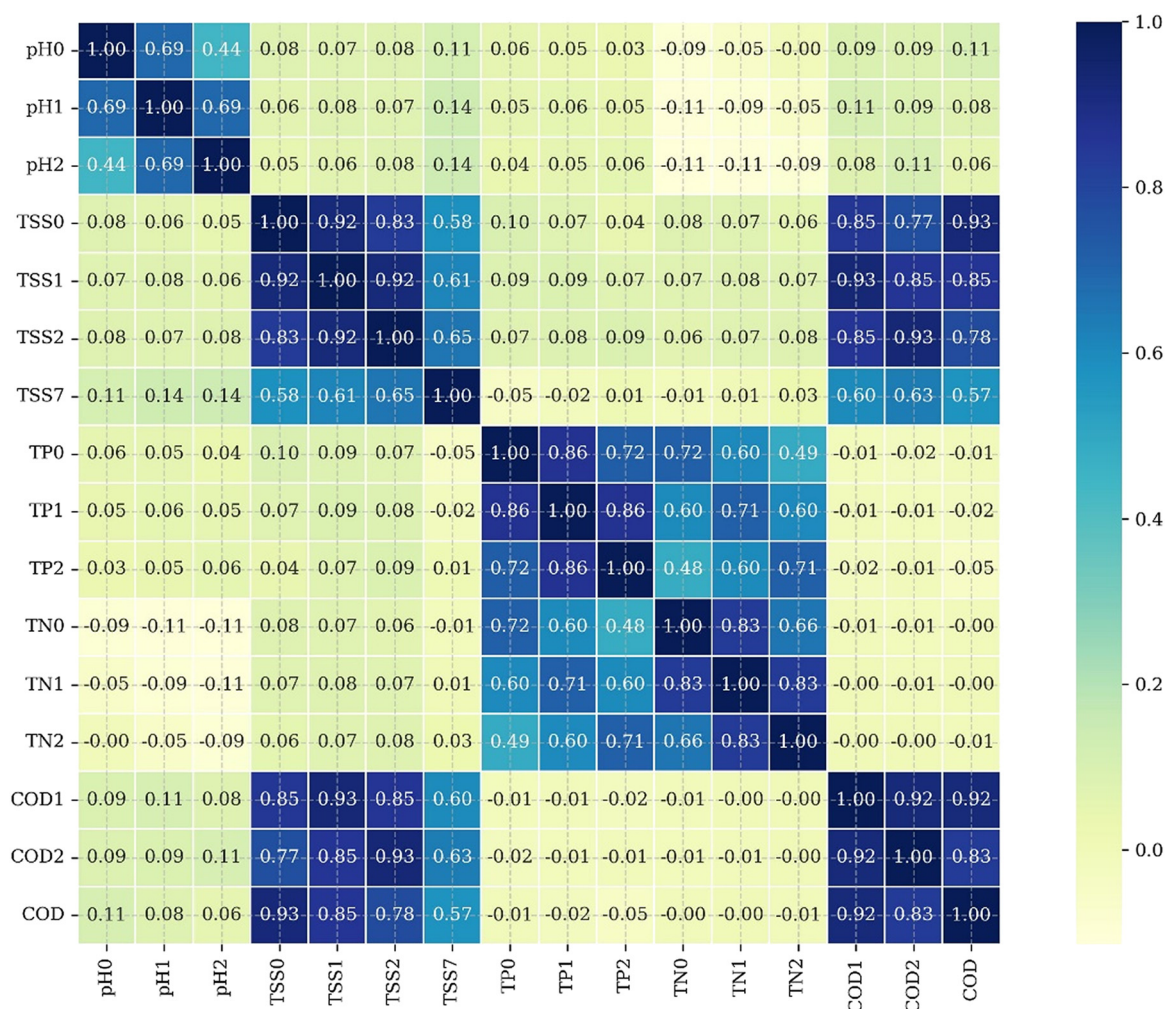


Fig. 7 Heatmap of parameters.



characteristics) or periodic changes in treatment efficiency. The medium frequency structure carries the majority or dominant level of variation of COD, indicating that mid-range cycles likely have the most influence on the type of temporal response of water quality from a statistical perspective. The fact that the low-frequency component is relatively stable and dynamic helps it become a key factor driving predictive accuracy in hybrid models. In contrast, the low-frequency component and the original COD series exhibit a similar long-term trend that is slow and gradual. The low-frequency component of COD concentrations shows the persistent baseline and overall move throughout the time period, likely driven by persistent changes in operational regimes or environmental conditions, and the long-term trend is greatly influenced by this component. Fig. 7 displays the correlation coefficient heatmap of the main water quality variables COD, TSS, TN, TP, and pH. The numbers (*e.g.* pH0, pH1, pH2, *etc.*) refer to how many days into the lag time each variable represents when observations were made in relation to the dependent variable. For example, pH1 refers to pH values measured one day before, while pH2 refers to values measured two days before.

The most remarkable research finding in the heatmap was a strong association between COD and the lagged values, COD1 and COD2. In this case, this implies a temporal dependence of the COD time series data, meaning that the past COD levels strongly affect the present values. However, this is not entirely surprising, given that persistence is a common feature of environmental or process-based data. Pollution concentrations as frequently observed in a wastewater context will respond slowly because of the nature of organic matter degradation or accumulation. These high correlations confirmed the adequacy of lagged COD values as significant predictors in time-series modeling, suggesting that the time-series model would successfully capture autocorrelating behavior and information related to the wastewater function in processes over time. Similarly, COD demonstrated a strong positive association with TSS and suggests that the majority of COD forming organic matter is in the suspended form. This is not surprising, since suspended solids often contain a portion of the degradable organic material, which is a direct contributor to COD learning through biochemical oxidation. Again, the strong linkage of COD and TSS represents the physical and chemical

relationship inherent in the treatment process indicating that fluctuations in solids loading directly influence organic degradation within processes and ultimately impact effluent quality. In comparison, COD displayed lower correlations amongst other parameters such as pH, TN, and TP. Based on these findings, Table 4 displays the various input combinations that were used to develop and assess the hybrid deep learning models for predicting COD. Each input combination demonstrates a structured integration of water quality indicators in excess of their lag variables for use to determine how different combinations of parameters will affect model strength and accuracy.

The initial four sets of input combinations focus on the sequential inclusion of core indicators of water quality, which are TSS, TN, TP, and pH. Combination 1 uses only TSS and its lagged values (TSS0, TSS1, TSS2, TSS7), allowing us to concentrate on the influence of suspended solids on COD variation. This simple model tests how far particulate matter alone can explain COD variation. Combination 2 includes TN and its lags, because nitrogenous compounds contribute to oxidation processes and changes in organic load, as seen with TSS. Adding TP into combination 3 represents an expansion of the model to include the potential interactions of nutrients, which can influence COD indirectly by changing microbial activity. Finally, combination 4 adds pH and its lags, which together with the previous input variables represent all of the input combinations accounting for the physicochemical processes of the system. The second four combinations introduce the lagged COD responses as additional predictors of the output variable. These combinations acknowledge the strong autocorrelation that often occurs with COD time series data, wherein previous COD concentration values are likely to influence future values. Incorporating the previous COD values into the model enables the combination of input variables and subsequent models to adequately account for biologically important persistence in time and nonlinear dependence within the system. For example, combination 5 has the same structure as combination 1 but uses COD lags instead of external inputs, such as TN or TP, to show how predictive capabilities could be increased using historical COD data. Combinations 6–8 use TN, TP, and pH, along with COD lags, to form comprehensive input structures that utilize both external drivers and internal temporal feedback.

Table 4 Alternative model forms and input combinations

Model No.	List of input parameters															
1	TSS0	TSS1	TSS2	TSS7												
2	TSS0	TSS1	TSS2	TSS7	TN0	TN1	TN2									
3	TSS0	TSS1	TSS2	TSS7	TN0	TN1	TN2	TP0	TP1	TP2						
4	TSS0	TSS1	TSS2	TSS7	TN0	TN1	TN2	TP0	TP1	TP2	pH0	pH1	pH2			
5	TSS0	TSS1	TSS2	TSS7											COD1	COD2
6	TSS0	TSS1	TSS2	TSS7	TN0	TN1	TN2								COD1	COD2
7	TSS0	TSS1	TSS2	TSS7	TN0	TN1	TN2	TP0	TP1	TP2					COD1	COD2
8	TSS0	TSS1	TSS2	TSS7	TN0	TN1	TN2	TP0	TP1	TP2	pH0	pH1	pH2		COD1	COD2



Table 5 The variation range of the different hyperparameters used in this research

No.	Hyperparameter	Range
1	Initial learn rate	0.0001–0.001
2	Gradient decay factor	0.9–0.99
3	Squared gradient decay factor	0.9–0.999
4	L2 regularization	0.0001–0.005
5	Number of layers	3–5
6	Mini batch size	64–256
7	Max epochs	200
8	Gradient threshold	1–5
9	Learn rate drop period	0.7–0.9
10	Learn rate drop factor	30–70

Table 5 outlines the range of variation for the selected hyperparameters that were adjusted and included in the training and optimization of the CEEMDAN–LSTM, CEEMDAN–GRU, and CEEMDAN–LSTM/GRU–MOOTLBO hybrid models created in this study. The parameters were tuned systematically to find the best trade-off between learning performance, generalization capabilities, and computational stability.

The starting learning rate, chosen within the range of 0.0001 to 0.001, was important in determining the speed by which all models learned to adapt to the error gradient when training. A lower rate provided smoother convergence for complicated decomposed data from CEEMDAN. A higher learning rate adapted more quickly at the beginning of training. The gradient decay factor (0.9–0.99) and squared gradient decay factor (0.9–0.999) were adjusted to minimize the adaptive behavior of the learning in the Adam optimizer, while yielding more stable results when training architectures like LSTM and GRU since those architectures are sensitive to gradient fluctuations while modeling long sequences (in this case, longer temporal and broader dimensional flow of each decomposed IMF). L2 regularization, optimized between the values of 0.0001 and 0.005, was added to avoid overfitting and to ensure that the models retained generalization when learning from the various decomposed IMFs. Additionally, the training configurations and the number of layers were also important when optimizing our modifications. The total number of layers for each model was bounded between 3 and 5 layers; the latter would allow for deeper networks, such as the CEEMDAN–LSTM–MOOTLBO model, to learn hierarchical temporal dependencies to capture epochs in a more computationally efficient way for decomposed data. The mini-batch size was chosen based on values between 64 and 256, and the maximum number of training epochs was selected to be 200. Each of these selections was determined based on a balance of speed to convergence rate and was trained to be relatively stable. The gradient threshold that delineated when to stop training was selected to be 1 to 5 to avoid the problem of gradient explosion. This is an issue with RNN-style models that handle high-frequency data and the CEEMDAN decomposed data components. We focused on fine-tuning the learning rate drop period (0.7–0.9) and drop

factor (30–70), which lowered the learning rate toward convergence, improved fine-tuning, and reduced oscillations. In addition, evaluation of the predictive performance of all models developed and tested in this study based on the two standalone deep learning architectures (LSTM and GRU), two hybrid models based on CEEMDAN decomposition (CEEMDAN–LSTM and CEEMDAN–GRU), and two fully optimized hybrid models (CEEMDAN–LSTM–MOOTLBO and CEEMDAN–GRU–MOOTLBO) is presented in Table 6.

To assess the benefits of hybridization, optimization, and ensemble modelling, the standalone deep learning techniques, including LSTM and GRU, were used as baseline architectures. The standalone LSTM models showed a trend of improvement across all configurations, with the first configuration of LSTM (*i.e.*, LSTM-1) producing RMSE = 140.04, MAE = 88.18, and $R^2 = 0.8720$ at the lowest, and the best configuration of LSTM (*i.e.*, LSTM-5) having RMSE = 51.71, MAE = 33.62, and $R^2 = 0.9825$ at the highest. The MAPE also demonstrated improvement as it dropped from above 13% to 5.76%. The MBE in LSTM-5 (*i.e.*, –7.87) was also nearly zero, indicating a slight underestimation of the predicted COD values, and the KGE (*i.e.*, 0.9661) indicated good agreement between actual and predicted values. The GRU also showed a similar pattern of improvement, with the first configuration GRU (GRU-1) producing RMSE = 150.67 and $R^2 = 0.8518$ at the lowest, and the best GRU model (*i.e.*, GRU-5) producing RMSE = 51.24, MAE = 33.31, and $R^2 = 0.9829$ at the highest. Overall, both standalone deep learning techniques were able to capture the general dynamics of COD reasonably well; however, the relatively larger RMSE and MAPE produced by both models suggest that these techniques struggled to capture the complexity associated with the nonlinear and nonstationary nature of the data.

The predictive ability improved with the usage of CEEMDAN decomposition. CEEMDAN-based hybrids decompose the original COD series into several IMFs or sub-components that can be modeled alone and then recombined, effectively removing noise and exposing any hidden temporal structures. The CEEMDAN–LSTM models showed predictive improvement; the RMSE value for the models dropped into single digits. For example, with CEEMDAN–LSTM-1, the RMSE = 5.44, MAE = 3.69, $R^2 = 0.9998$, MAPE = 0.72%, which is a substantial improvement from using LSTM alone. The KGE value of 0.9967 demonstrated excellent performance, while the near-zero MBE (2.07) indicated very little bias. The different CEEMDAN–LSTM configurations performed well, also, with RMSE values ranging from 5 to 10, and the R^2 always above 0.9993. The results demonstrate clearly that CEEMDAN decomposition improved both the interpretability of the model results and the predictive ability of the models by providing the capacity for LSTM to focus on capturing smoother and more stationary sub-signals as opposed to modeling the prediction using the full and complex COD series. The CEEMDAN–GRU models resulted in an even more accurate prediction in nearly all performance metrics.



Table 6 Evaluation of the models' performance in the testing phase

Model	RMSE	MAE	R^2	MAPE	MBE	KGE
LSTM-1	140.04	88.18	0.8720	13.14	-16.18	0.8725
LSTM-2	139.75	87.44	0.8725	13.24	-13.92	0.8680
LSTM-3	113.85	77.61	0.9154	12.90	-1.57	0.9249
LSTM-4	110.90	72.05	0.9197	12.42	-6.01	0.9533
LSTM-5	51.71	33.62	0.9825	5.76	-7.87	0.9661
LSTM-6	57.76	36.23	0.9782	6.15	-13.23	0.9627
LSTM-7	63.61	37.96	0.9736	6.30	-17.39	0.9650
LSTM-8	58.86	36.85	0.9774	6.24	-13.73	0.9652
GRU-1	150.67	88.19	0.8518	12.82	-19.72	0.8497
GRU-2	138.09	86.62	0.8755	13.02	-11.00	0.8764
GRU-3	111.35	72.65	0.9190	12.78	-5.18	0.9422
GRU-4	106.77	70.35	0.9256	11.69	-4.78	0.9521
GRU-5	51.24	33.31	0.9829	5.66	-9.22	0.9656
GRU-6	60.45	37.74	0.9761	6.37	-18.60	0.9568
GRU-7	65.73	38.89	0.9718	6.50	-22.29	0.9582
GRU-8	61.83	35.81	0.9750	6.39	-9.21	0.9812
CEEMDAN-LSTM-1	5.44	3.69	0.9998	0.72	2.07	0.9967
CEEMDAN-LSTM-2	6.53	5.11	0.9997	1.07	3.68	0.9943
CEEMDAN-LSTM-3	8.23	5.25	0.9996	1.16	2.06	0.9936
CEEMDAN-LSTM-4	7.65	5.52	0.9996	1.13	1.54	0.9976
CEEMDAN-LSTM-5	9.10	4.90	0.9995	1.00	-2.98	0.9944
CEEMDAN-LSTM-6	6.18	3.88	0.9998	0.86	-0.34	0.9983
CEEMDAN-LSTM-7	7.32	5.25	0.9997	1.02	-3.48	0.9945
CEEMDAN-LSTM-8	10.16	7.16	0.9993	1.35	-6.23	0.9887
CEEMDAN-GRU-1	2.84	2.28	0.9999	0.46	1.47	0.9974
CEEMDAN-GRU-2	6.52	4.70	0.9997	0.93	3.90	0.9940
CEEMDAN-GRU-3	5.48	3.76	0.9998	0.77	-2.32	0.9949
CEEMDAN-GRU-4	10.77	5.53	0.9992	1.29	4.50	0.9922
CEEMDAN-GRU-5	4.00	2.67	0.9999	0.51	-1.38	0.9977
CEEMDAN-GRU-6	8.55	3.18	0.9995	0.83	1.50	0.9960
CEEMDAN-GRU-7	7.52	3.92	0.9996	0.82	-0.98	0.9957
CEEMDAN-GRU-8	12.44	4.26	0.9990	1.02	0.00	0.9938
CEEMDAN-LSTM-MOOLBO-1	4.66	3.48	0.9999	0.71	2.07	0.9967
CEEMDAN-LSTM-MOOLBO-2	5.54	3.79	0.9998	0.82	0.55	0.9982
CEEMDAN-LSTM-MOOLBO-3	9.36	4.83	0.9994	1.14	-0.65	0.9984
CEEMDAN-LSTM-MOOLBO-4	6.50	4.57	0.9997	1.04	-0.07	0.9975
CEEMDAN-LSTM-MOOLBO-5	7.47	4.85	0.9996	0.93	-3.90	0.9938
CEEMDAN-LSTM-MOOLBO-6	8.28	6.12	0.9996	1.11	-5.60	0.9907
CEEMDAN-LSTM-MOOLBO-7	6.82	5.39	0.9997	1.14	2.77	0.9944
CEEMDAN-LSTM-MOOLBO-8	6.88	4.73	0.9997	0.97	-2.09	0.9953
CEEMDAN-GRU-MOOLBO-1	5.30	4.05	0.9998	0.84	2.93	0.9954
CEEMDAN-GRU-MOOLBO-2	4.19	2.66	0.9999	0.52	0.16	0.9997
CEEMDAN-GRU-MOOLBO-3	5.87	4.07	0.9998	0.85	2.35	0.9963
CEEMDAN-GRU-MOOLBO-4	8.01	5.39	0.9996	1.10	-3.01	0.9951
CEEMDAN-GRU-MOOLBO-5	5.99	2.93	0.9998	0.63	0.13	0.9995
CEEMDAN-GRU-MOOLBO-6	5.77	4.38	0.9998	0.98	3.04	0.9953
CEEMDAN-GRU-MOOLBO-7	6.72	4.33	0.9997	0.95	2.78	0.9957
CEEMDAN-GRU-MOOLBO-8	13.81	4.91	0.9988	1.23	0.08	0.9976

The configurations of CEEMDAN-GRU indicated being the highest predictive power overall, with the lowest RMSE (2.84), the lowest MAE (2.28), an exceptionally high value for R^2 (0.9999), and the lowest MAPE (0.46%). The MBE value of 1.47, which is nearly zero, shows that there is practically no systematic over or under-prediction. The KGE value (0.9974) furthermore confirmed the model's capability of closely mimicking COD values with superior consistency. The CEEMDAN-GRU-1 model showed a decrease in RMSE by approximately 94.5% and a decrease in MAPE of more than 90% compared to the best individual GRU (RMSE = 51.24, MAE = 33.31, R^2 = 0.9829), which again demonstrates a substantial improvement in predictive accuracy. This

shows the benefit of the interaction between CEEMDAN's decomposition property and GRU's recurrent learning ability, both of which contributed to representing COD behavior on short-term variations, as well as long-term dependencies. Other CEEMDAN-GRU alternatives still showed robust and reliable predictive capabilities, such as GRU-3, GRU-5, and GRU-6 (all with RMSE < 10 and R^2 > 0.9995). While they were still very strong performers, none were able to match CEEMDAN-GRU-1, which corresponded with the lowest error metrics and the highest correlation and efficiency coefficients. This consistency of performance again highlights the inherent strength of the GRU architecture when dealing with and predicting decomposed



sub-signals and leveraging GRU capabilities of learning multi-scale temporal data, providing better learning and prediction times than LSTM.

The MOOTLBO algorithm represented the final stage of refinement and improved hyperparameters of the CEEMDAN-LSTM and CEEMDAN-GRU models. The optimization process further stabilized model convergence by balancing the learning rate, regularization, and network depth hyperparameters. For the optimization of each model category, CEEMDAN-LSTM-MOOTLBO-2 yielded the best performance with RMSE = 5.54, MAE = 3.79, $R^2 = 0.9998$, MAPE = 0.82% and MBE around zero (0.55). The KGE of 0.9982 demonstrated near-perfect agreement of the CEEMDAN-LSTM-MOOTLBO-2 model's predictions with observations. Similarly, CEEMDAN-GRU-MOOTLBO-2 displayed high performance with RMSE = 4.19, MAE = 2.66, $R^2 = 0.9999$, MAPE = 0.52%, MBE = 0.16 and KGE = 0.9997, thus emerging as one of the most accurate, yet stable, configurations tested.

An overall assessment of all the models indicates that the predictive capacity has improved. The individual models (LSTM and GRU) were satisfactory but lacked precision due to the fact that they were being trained on the entire nonlinear and noisy COD data. The performance of the CEEMDAN hybrid models was better due to the decomposition of the data into various scales of time using the IMFs. This led to better performance of the deep learning models as they were able to identify more consistent temporal patterns. The use of the MOOTLBO optimization of the hybrid models showed an improvement in the performance of the models as the hyperparameters were automatically tuned. The performance of the models was consistent as the use of the optimization showed an improvement in the performance of the models as indicated by the various evaluation metrics. The performance of CEEMDAN-GRU-1 was the highest among the models as it had the lowest RMSE, MBE, and MAE values. Additionally, it had the highest R^2 and KGE values. The MBE was close to zero, indicating that the model had no bias. Therefore, it was the best-performing model. The fact that CEEMDAN-GRU-1 uses the TSS lag variables as predictors indicates that the total suspended solids might be the most important predictor of COD concentration in the WWTP. This is physically reasonable as the suspended matter tends to carry most of the organic matter in the wastewater. Therefore, the concentrations of COD and TSS tend to vary strongly. The addition of more variables such as TN, TP, pH, and COD lags to the model does not necessarily improve the performance of the model. Fig. 8 shows the comparison of observed *versus* predicted COD using the models developed with different input combinations. The observed COD concentrations are shown with the blue line in the figure, and the models' predictions with colored lines for the different model configurations.

As presented in Fig. 8, all models captured the overall temporal behavior of COD patterns (*i.e.*, both short-term

variability and longer-term trends) with varying levels of accuracy. The basic LSTM and GRU models represented a reasonable fit, with some deviation during periods of rapid COD change, particularly in the peaks and troughs. Although the performance metrics provided in the study are based on the aggregated results for the entire test data set, it can be inferred from Fig. 8 that the prediction accuracy of the hybrid models is consistent for varying ranges of COD concentration levels. The hybrid models show close agreement with the measured COD levels during both low-load periods and peak concentration levels. Although slightly higher deviations may be noticed during abrupt COD peaks, which are usually associated with short-term fluctuations in influent characteristics, the hybrid models based on CEEMDAN are still able to capture the overall dynamics of COD fluctuations. Such departures from the observed data indicate that although the recurrent architectures did capture some temporal dependencies, they may be limited in dealing with the high non-linearity and noise present with COD time series data. The residual errors suggest that the complicated interactions occurring between the physicochemical parameters were not fully learned within these models, without further decomposition of the signal. In comparison, the CEEMDAN-based hybrid models produced a much tighter fit between the predicted COD curves and actual reactive COD data as shown. The CEEMDAN-LSTM and CEEMDAN-GRU models produced predictions close to the observed data, with little variance even for the sudden spikes or drops in COD levels. This outcome exemplifies the advantage of CEEMDAN decomposition as it separates the COD series into multiple IMFs, allowing each component to be learned by the neural network models. Out of the various models tested, the CEEMDAN-GRU model is the most representative of the actual COD series by overlapping its predicted line with the observed data during most of the testing period. This model's findings indicate strong accuracy and generalization properties. The combination of the computationally efficient GRU architecture is ideal for modeling sequential dependencies in time series data and provides a high level of precision when capturing short- and long-term time series variations in COD. Therefore, scatter plots of actual and predicted COD values for the different input combinations are presented in Fig. 9. Each scatter point denotes a pair of observed and predicted COD values, while the diagonal line denotes the ideal 1:1 line for COD. Perfect predictions would mean that all points fall directly along the diagonal line.

The scatter plots demonstrated that even the standalone models LSTM and GRU have shown to have a good overall fit, although the points appear to be somewhat more dispersed around the 1:1 line. The dispersion is indicative of the size of each model's prediction errors particularly at the extremes of the COD where the standalone models tend to underpredict or overpredict actual values. This is understandable given that the model's input involves the raw (nonlinear) COD series which is not a subset or filtered prior



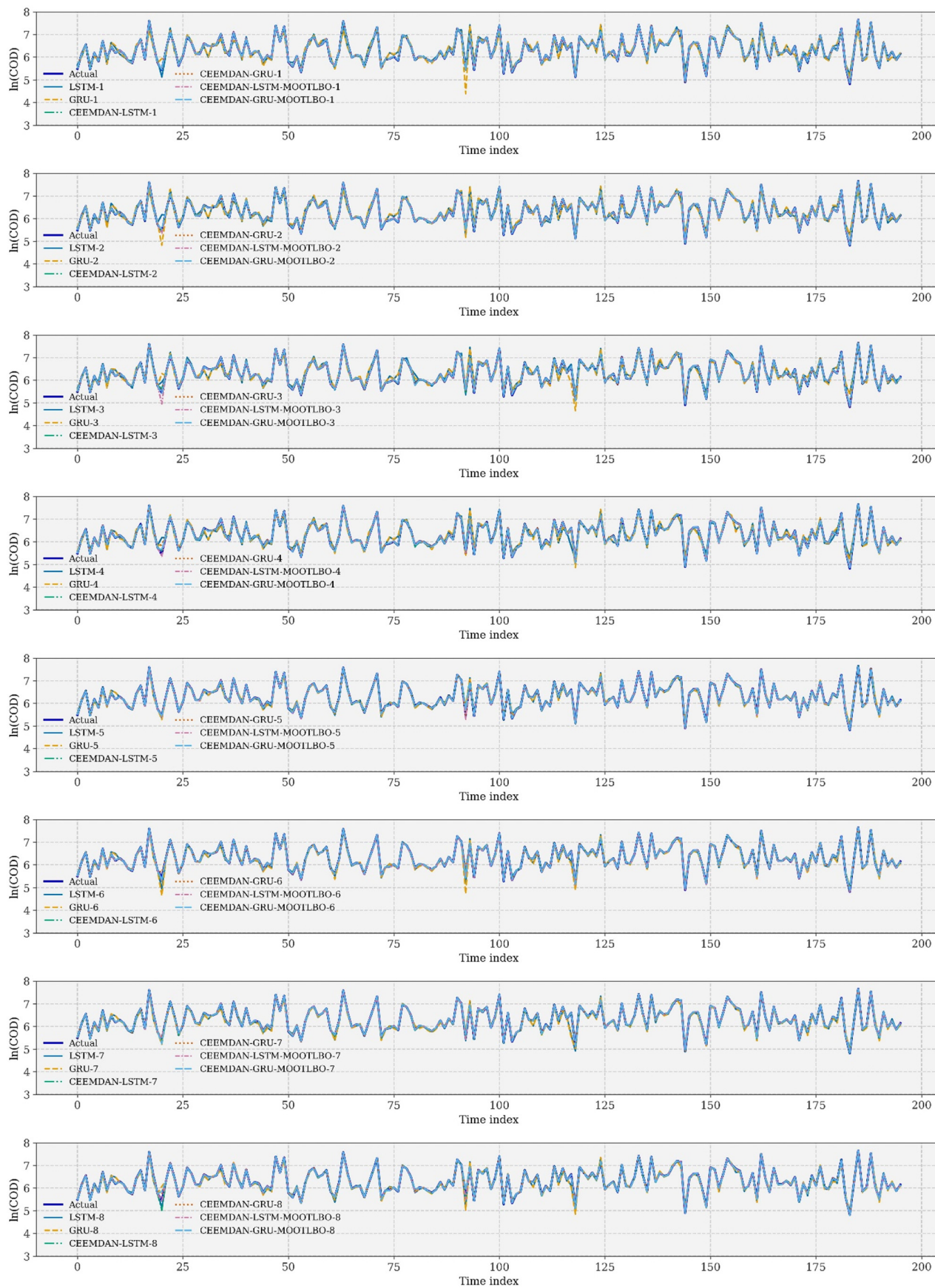


Fig. 8 Comparison of actual and predicted COD. The blue line represents actual COD and the colored lines show predictions from developed models for different input combinations.



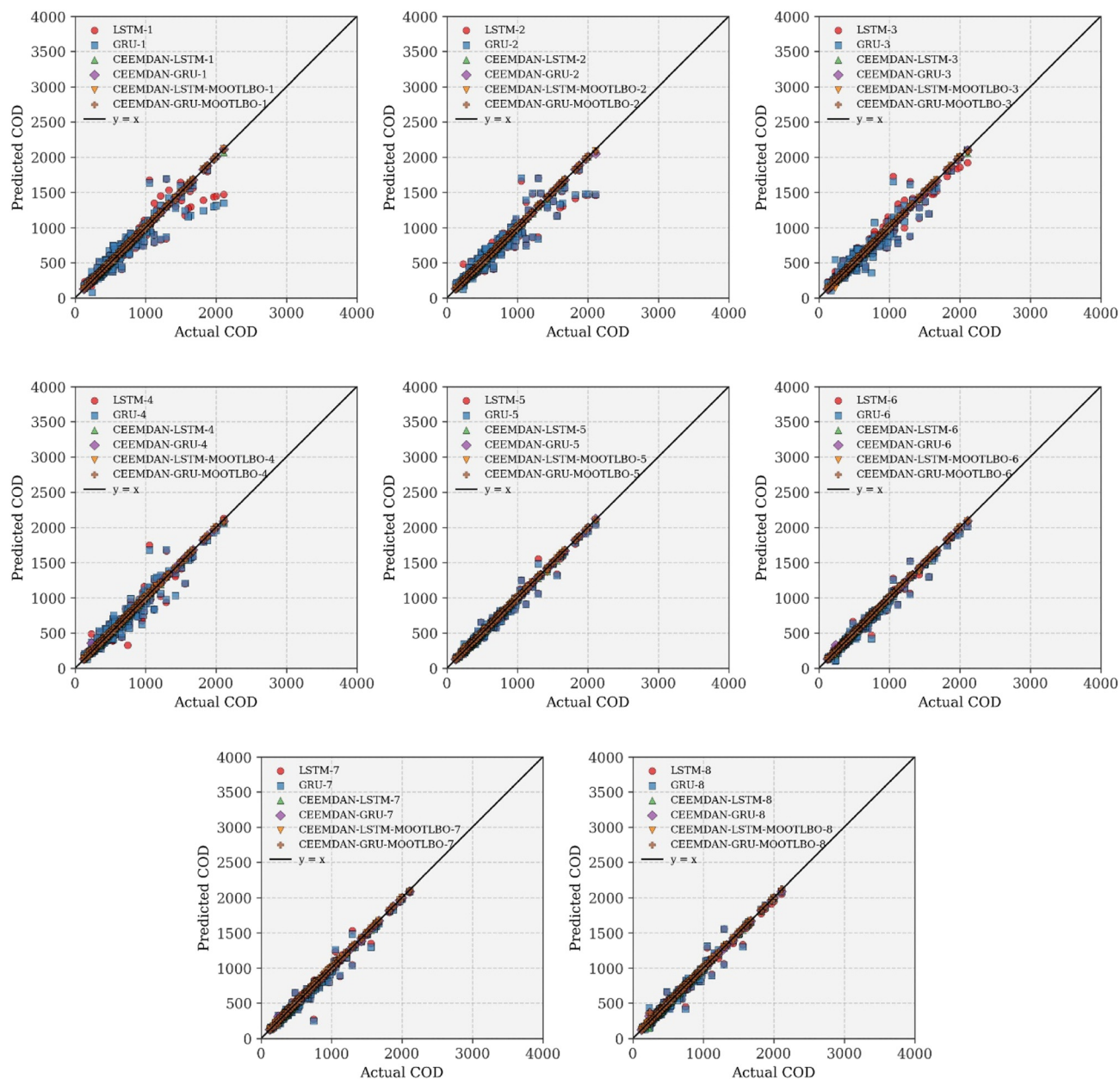


Fig. 9 Scatter plots of actual versus predicted COD values for different input combinations.

to inputting, therefore when prediction errors occur, the models are more susceptible to abrupt changes and noisy data. The CEEMDAN-based hybrid models, on the other hand, displayed a much tighter clustering along the diagonal and in comparison, highlighted improvements in hybrid model accuracy and bias in any predictions. The CEEMDAN-LSTM and CEEMDAN-GRU model points are more linear and closer to the ideal line of prediction suggesting that the CEEMDAN decomposition filters noise and allows the deep learning models to leverage the inherent signal dynamics. The CEEMDAN-GRU has the most clustered distribution along all points along the ideal line suggesting that it has the highest predictive accuracy.

To further clarify the issue, SHAP mean and SHAP summary plots are plotted as given in Fig. 10, which demonstrate the relative importance and impact of each input variable. The SHAP (Shapley additive explanations) method provides an interpretable, model-agnostic approach to quantify how much each feature contributes to predicted COD values, giving us some transparency into the internal decision-making process of complicated neural networks. The SHAP mean plot demonstrates the average effect of each variable. It highlights the fact that the lagged TSS and COD (TSS0, TSS1, TSS2, COD0 and COD1) represent the greatest mean SHAP values, and therefore are the most influential variables in predicting COD. This result is consistent with the known physical and biochemical associations within



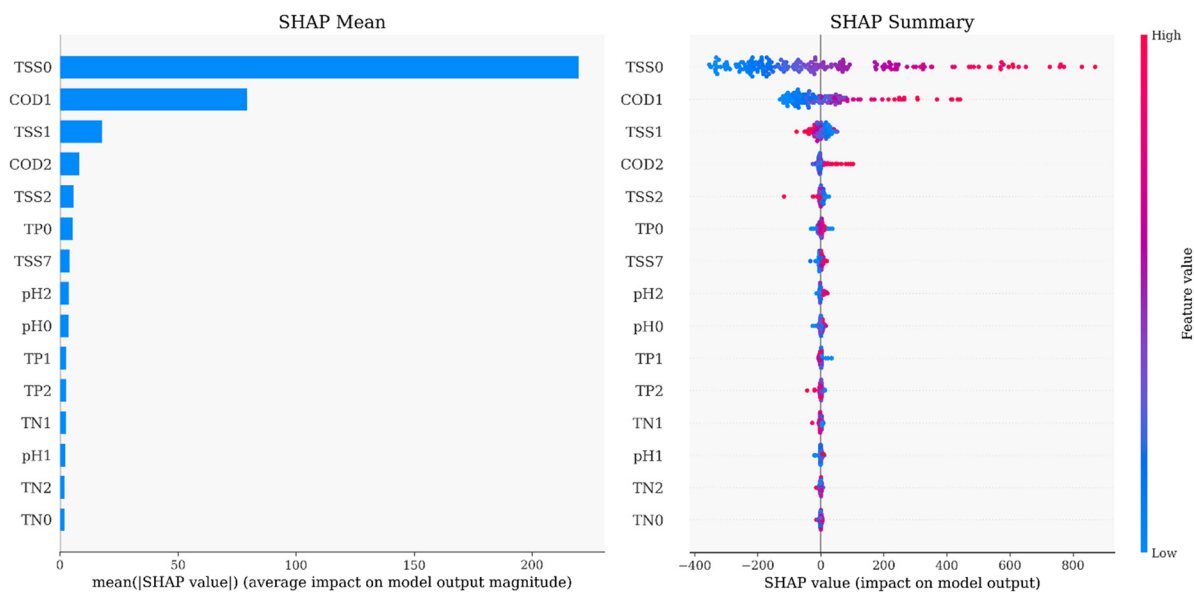


Fig. 10 SHAP mean and SHAP summary plots for COD values.

wastewater systems, where suspended solids and nutrient concentrations directly contribute to organic loading and the oxygen demand of the water. TP and pH have a lower contribution, but still play a meaningful role reflecting their indirect influence through chemical and microbial processes. The SHAP summary plot illustrates this by demonstrating the distribution of SHAP values for each feature in relation to the distribution of all samples in the experiment. The dense clustering and relatively narrow spread of TSS and COD lags indicate consistent and strong influences while TN, TP and pH displays wider spreads suggesting potentially variable influences based on process conditions.

Compared to previous research studies, the current results of COD prediction show similarly low errors and substantial variability, or explained variance. In the application presented by Tomperi *et al.*,²⁹ the series-connected linear framework (*e.g.* ARMAX/ARX/PLSR) had an identification performance of $R \approx 0.67$, MAPE of 7.33%, and MAPE during test set performance of 9.04–9.17% with MAE values of 15.7 and 20.2 mg L⁻¹, showing the utility of linear and interpretable estimation equations and the enhanced prediction error ability with more complexity and dimensionality across the nonstationary period. For reference, our hybrid configuration based on a combination of signal decomposition and gated recurrent learning was able to reduce error to low single digits *e.g.*, RMSE ≈ 3 mg L⁻¹ and MAE ≈ 2 mg L⁻¹ while explaining >0.95 of variance or explanatory power, which better tracked values for abrupt swings in COD while still generalizing on unseen periods. When also taking into account data-driven work focused on removal of COD instead of the absolute concentration, there was an additional improvement in the level of fit. Albaba *et al.*³⁰ reported a stacked super learner, who upon testing achieved $R^2 = 95.26\%$ with MAE = 3.225%, MAPE = 6.332%, and RMSE = 4.144%. When this is considered, the model

obtained a comparable or greater R^2 whilst producing lower absolute error. This suggests that the combination of CEEMDAN-mode separation, gated recurrence and multi-objective hyperparameter tuning provides increased isolation of multi-scale dynamics whilst producing low levels of error inflation due to sudden load shocks.

3.2. Discussion

The analyses show that the COD time-series fluctuations are highly indicative of both temporal short-term changes within WWTPs and a long-term operational cycle corresponding to physical and biological processes within treatment units. The observed sizeable lag dependencies for COD, TSS, TN, TP, and pH suggest that recent COD observations are not solely indicative of current inflow conditions but also indicative of prior operational conditions of the system including sludge accumulation timescales, microbial activity, and hydraulic retention behavior. As nominally expected, these time-lagged relationships are consistent with the physical processes of sedimentation, aeration, and nutrient removal, where treatment processes inherently respond over timescales because of the lag that exists in the processing systems. By recognizing these temporal dependencies, the hybrid models can accurately replicate the operational memory of WWTPs in a predictive tool that informs the plant operator's ability to anticipate treatment efficiency and particulate load changes. Results of the decomposition and correlation analysis illustrate that TSS is the primary driver of COD variability. This confirms that total suspended solids contribute directly to oxygen demand as a function of biodegradation. The awareness of interdependencies enables treatment authorities to monitor key control parameters, such as solids loading, to improve operational target performance and COD stabilization.



Table 7 Computational latency of hybrid prediction pipelines

Latency metric	CEEMDAN–LSTM–MOOTLBO	CEEMDAN–GRU–MOOTLBO
CEEMDAN decomposition latency (s)	3.6	3.4
Single-sample model inference latency (ms)	110	107
Single-sample total pipeline latency (s)	3.7	3.5
Batch total latency on test set (s)	0.1	0.1
Per-sample batch total latency (ms per sample)	0.6	0.6

In order to assess the operational feasibility of the proposed framework for real-time monitoring scenarios, the computational latency of the proposed hybrid models was assessed. Table 7 presents a summary of decomposition time, single sample inference time, total prediction pipeline time, and batch prediction efficiency of the CEEMDAN–LSTM–MOOTLBO and CEEMDAN–GRU–MOOTLBO models.

The results of the analysis of the computational latency of the proposed method, as presented in Table 7, offer significant insights regarding the feasibility of the proposed framework in terms of real-time wastewater monitoring applications. As presented in the table, the computational latency of the proposed method is mainly attributed to the computational time of the CEEMDAN algorithm. More specifically, the computational time of the CEEMDAN algorithm amounts to approximately 3.6 s and 3.4 s for the CEEMDAN–LSTM–MOOTLBO and CEEMDAN–GRU–MOOTLBO models, respectively. It should be noted that the computational time of the proposed method would be even lower in the case of a batch of samples. However, in the case of a single sample, the total computational time of the proposed method amounts to approximately 3.7 s and 3.5 s for the CEEMDAN–LSTM–MOOTLBO and CEEMDAN–GRU–MOOTLBO models, respectively. After the completion of the CEEMDAN algorithm, the computational time of the proposed method regarding the inference of the employed recurrent neural network model is extremely low and amounts to approximately 110 ms and 107 ms for the LSTM and GRU models, respectively. Regarding batch prediction, the computational time of the proposed method amounts to approximately 0.1 s for the entire test dataset. Moreover, the effective computational time of the proposed method amounts to approximately 0.6 ms. It should be noted that the update cycle of the supervisory control and data acquisition systems of wastewater treatment plants amounts to approximately 15 minutes.

The CEEMDAN–GRU model demonstrated strong performance with peak-throughput predictive power as a framework for decision-support systems. Through peak prediction capability, operators can anticipate associated operational strategies to mitigate event-driven loading, such as aeration rates, sludge recirculation or without compromising treatment of regulatory effluent limits. From a policy viewpoint, the information findings can provide direction for environmental authorities to develop data-driven and adaptive management systems for WWTPs.

More sophisticated and accurate models of COD prediction can be integrated into digital monitoring dashboards of WWTPs for real-time performance feedback and compliance checks. In addition, models of this kind could support long-term decisions by allowing decision-makers to examine how changes in operation processes or climate-driven variability in influent loads might impact the treatment process performance and effluent quality. However, while these hybrid models were able to obtain extremely high R^2 values, there are some indications that these models possess high generalization capabilities. These indications include: firstly, model evaluation is carried out on a hold-out test set containing 20% of chronological data. Secondly, the application of the CEEMDAN signal decomposition has the capability to remove stochastic noises from the COD signal by decomposing it into intrinsic mode functions. This allows the neural networks to learn smoother sub-signals from the data set. As such, overfitting is less likely to occur. Thirdly, the application of MOOTLBO multi-objective optimization is able to limit over-parameterization by balancing model accuracy and simplicity.

While this research has been carried out on a single municipal WWTP in Türkiye, the framework has been shown to be applicable to other WWTPs. The hybrid modeling framework is independent of any WWTP's specific conditions and is only dependent on the time-series data from the water quality parameters. However, while this research has been able to obtain optimal parameters for a single WWTP in Türkiye, these parameters may vary from those in other WWTPs due to conditions such as influent composition, industrial influences, and climatic conditions. For practical purposes, to apply this framework to a different WWTP from that used in this research, recalibration of parameters is required. Fortunately, this is achieved through the application of automated optimization techniques as presented in this research. As such, while parameters may vary from one WWTP to another, the framework is applicable to WWTPs under different environmental conditions. Despite encouraging model results, some cautions should be noted. Developed models were able to capture the necessary complexity to evaluate treatment configurations and process variability, yet they depend on the quality and frequency of the input data; measurement errors or missing values from field sensors could limit prediction accuracy. Furthermore, while the CEEMDAN–GRU model is appropriate for capturing temporal dependencies, it does not offer a means of



modeling spatial dependencies between treatment units or between treatment plants. The potential effects of external factors such as temperature, hydraulic load, and shifts in wastewater microbial communities have yet to be integrated into future frameworks. Although PACF-guided lag selection provided an interpretable and compact representation of process memory in this study, it remains a fixed lag design strategy. In highly nonstationary wastewater treatment systems, adaptive lag selection mechanisms or attention-based sequence models may offer additional flexibility for capturing time-varying dependencies during regime shifts and therefore represent an important direction for future research. Consequently, future studies should look to include more parameters on treatment decisions and variation of environmental conditions to further develop and support this model to generalize across decision-making. Integrating this framework with real-time sensors, Internet-of-Things systems, and digital twins of the treatment plant could also increase usefulness and reliability.

4. Conclusion

This work developed and validated a hybrid framework composed of deep learning (CEEMDAN-GRU-MOOLBO) to achieve accurate, real-time predictions of effluent COD within a full-scale municipal WWTP. The framework developed a new workflow integrating signal decomposition, recurrent neural sequence learning, and multi-objective hyperparameter optimization within one integrated framework. While decomposing through CEEMDAN separated nonstationary fluctuations in COD signals, the GRU network captured short and mid-term temporal dependence efficiently. The use of the multi-objective optimization algorithm enhanced the performance of the deep learning model by striking an optimal balance between predictive accuracy and computational cost. Overall, the CEEMDAN-GRU-MOOLBO formulation achieved a high level of accuracy ($R^2 = 0.9999$, RMSE = 4.19 mg L⁻¹, MAE = 2.66 mg L⁻¹), which is more than 90% greater than the predictive capacity of conventional GRU, LSTM, and ensemble baselines. From a methodological perspective, this research progresses the area of data-driven soft sensing by proposing a multi-objective, decomposition-based recurrent modeling framework for full-scale WWTPs. The proposed system develops a novel modeling strategy to directly address three key hurdles in environmental time series modeling: noise interference, lag effects, and optimization trade-offs. The model structure and optimization strategy can subsequently be easily applied to other important water quality parameters, such as biochemical oxygen demand (BOD), total nitrogen (TN), or total phosphorus (TP). Thus, the methodology developed in this study can provide a generalizable approach to soft sensing in complex environmental systems. From an operational perspective, this approach provides a research-based tool that operationalizes a practical approach for real-time COD forecasting from the

routinely monitored parameters of TSS, TN, TP, and pH. This provides for early detection of process disturbance and for on-the-fly, adaptive control of aeration and dosing, and potentially reduces the need for laboratory workload, further allowing for improved decision making in real time for energy efficient and reactive WWTP operation. Moreover, the lightweight and transferrable structure facilitates application within standard SCADA or digital twin platforms, enabling data-driven decisions, which can ultimately improve water quality outcomes. There are still limitations despite the robust outcomes. The models have been developed on a daily composite basis, and moving to higher temporal resolutions will better capture dynamic load fluctuations. In addition, examining the approach in varying climates or additional operational scenarios can better assess reliability and transferability. Thus, future work will focus on the online implementation of the model and coupling it with process control algorithms for autonomous adaptive and optimized scheduling at wastewater treatment facilities. Although the proposed hybrid models demonstrated strong predictive performance, further validation using datasets from different wastewater treatment plants and operational conditions is necessary to fully confirm their robustness and practical applicability.

Author contributions

SS: conceptualization; data curation; formal analysis; investigation; methodology; validation; visualization; writing – original draft. EA: formal analysis; methodology; supervision; writing – original draft. DYÇ: data curation; methodology; visualization; writing – original draft; MTS: conceptualization; methodology; writing – original draft. OG: conceptualization; methodology; validation; supervision; writing – review and editing.

Conflicts of interest

On behalf of all authors, the corresponding author states that there are no conflicts of interest.

Data availability

The obtained data can be made available on request of interested parties under the condition of approval of the request by the authors of the article.

Acknowledgements

No direct funding has been received for this research. Saeed Samadianfard has been supported by the Technological and Scientific Research Council of Türkiye through the Fellowships for Visiting Scientists and Scientists on Sabbatical Leave Program 2221. Moreover, the İSU Gebze



Advanced Biological Wastewater Treatment Plant Directorate is gratefully acknowledged for providing the wastewater data. Additionally, generative AI tools were used solely to assist with language clarity, grammar, and vocabulary refinement. All scientific content, plots, analysis, interpretation, and conclusions were developed exclusively by the authors.

References

- 1 L. Corominas, M. Garrido-Baserba and K. Villez, *et al.*, Transforming data into knowledge for improved wastewater treatment operation: A critical review of techniques, *Environ. Model. Softw.*, 2018, **106**, 89–103, DOI: [10.1016/j.envsoft.2017.11.023](https://doi.org/10.1016/j.envsoft.2017.11.023).
- 2 K. B. Newhart, R. W. Holloway, A. S. Hering and T. Y. Cath, Data-driven performance analyses of wastewater treatment plants: A review, *Water Res.*, 2019, **157**, 498–513, DOI: [10.1016/j.watres.2019.03.030](https://doi.org/10.1016/j.watres.2019.03.030).
- 3 D. Wang, S. Thunell and U. Lindberg, *et al.*, A machine learning framework to improve effluent quality control in wastewater treatment plants, *Sci. Total Environ.*, 2021, **784**, 147138, DOI: [10.1016/j.scitotenv.2021.147138](https://doi.org/10.1016/j.scitotenv.2021.147138).
- 4 L. Verhaeghe, J. Verwaeren and G. Kirim, *et al.*, Towards good modelling practice for parallel hybrid models for wastewater treatment processes, *Water Sci. Technol.*, 2024, **89**(11), 2971–2990, DOI: [10.2166/wst.2024.159](https://doi.org/10.2166/wst.2024.159).
- 5 D. Wu, Y. Hu and Y. Liu, A review of detection techniques for chemical oxygen demand in wastewater, *Am. J. Biochem. Biotechnol.*, 2022, **18**(1), 23–32, DOI: [10.3844/ajbbbsp.2022.23.32](https://doi.org/10.3844/ajbbbsp.2022.23.32).
- 6 R. B. Geerdink, R. S. van den Hurk and O. J. Epema, Chemical oxygen demand: Historical perspectives and future challenges, *Anal. Chim. Acta*, 2017, **961**, 1–11, DOI: [10.1016/j.aca.2017.01.009](https://doi.org/10.1016/j.aca.2017.01.009).
- 7 N. Manav-Demir, H. B. Gelgor and E. Oz, *et al.*, Effluent parameters prediction of a biological nutrient removal (BNR) process using different machine learning methods: A case study, *J. Environ. Manage.*, 2024, **351**, 119899, DOI: [10.1016/j.jenvman.2023.119899](https://doi.org/10.1016/j.jenvman.2023.119899).
- 8 P. Wang, Y. Chen and C. Zhang, *et al.*, Real-time prediction of the chemical oxygen demand component parameters in activated sludge model using backpropagation neural network, *Heliyon*, 2024, **10**(16), e35580, DOI: [10.1016/j.heliyon.2024.e35580](https://doi.org/10.1016/j.heliyon.2024.e35580).
- 9 M. Ay and O. Kisi, Modelling of chemical oxygen demand by using ANNs, ANFIS and k-means clustering techniques, *J. Hydrol.*, 2014, **511**, 279–289, DOI: [10.1016/j.jhydrol.2014.01.054](https://doi.org/10.1016/j.jhydrol.2014.01.054).
- 10 R. Wang, Y. Yu and Y. Chen, *et al.*, Model construction and application for effluent prediction in wastewater treatment plant: Data processing method optimization and process parameters integration, *J. Environ. Manage.*, 2022, **302**, 114020, DOI: [10.1016/j.jenvman.2021.114020](https://doi.org/10.1016/j.jenvman.2021.114020).
- 11 S.-P. Mujunen, P. Minkkinen, P. Teppola and R.-S. Wirkkala, Modeling of activated sludge plants treatment efficiency with PLSR: A process analytical case study, *Chemom. Intell. Lab. Syst.*, 1998, **41**(1), 83–94, DOI: [10.1016/S0169-7439\(98\)00025-2](https://doi.org/10.1016/S0169-7439(98)00025-2).
- 12 P. Teppola, S.-P. Mujunen and P. Minkkinen, Kalman filter for updating the coefficients of regression models: A case study from an activated sludge waste-water treatment plant, *Chemom. Intell. Lab. Syst.*, 1999, **45**(1–2), 371–384, DOI: [10.1016/S0169-7439\(98\)00145-2](https://doi.org/10.1016/S0169-7439(98)00145-2).
- 13 D. Güçlü and Ş. Dursun, Artificial neural network modelling of a large-scale wastewater treatment plant operation, *Bioprocess Biosyst. Eng.*, 2010, **33**(9), 1051–1058, DOI: [10.1007/s00449-010-0430-x](https://doi.org/10.1007/s00449-010-0430-x).
- 14 Y. Qiu, Y. Liu and D. Huang, Data-driven soft-sensor design for biological wastewater treatment using deep neural networks and genetic algorithms, *J. Chem. Eng. Jpn.*, 2016, **49**(10), 925–936, DOI: [10.1252/jcej.16we016](https://doi.org/10.1252/jcej.16we016).
- 15 Y. Liu, Adaptive just-in-time and relevant vector machine based soft-sensors with adaptive differential evolution algorithms for parameter optimization, *Chem. Eng. Sci.*, 2017, **172**, 571–584, DOI: [10.1016/j.ces.2017.07.006](https://doi.org/10.1016/j.ces.2017.07.006).
- 16 Y. Yang, K.-R. Kim and R. Kou, *et al.*, Prediction of effluent quality in a wastewater treatment plant by dynamic neural network modeling, *Process Saf. Environ. Prot.*, 2022, **158**, 515–524, DOI: [10.1016/j.psep.2021.12.034](https://doi.org/10.1016/j.psep.2021.12.034).
- 17 G. Niu, X. Yi and C. Chen, *et al.*, A novel effluent quality predicting model based on genetic-deep belief network algorithm for cleaner production in a full-scale paper-making wastewater treatment, *J. Cleaner Prod.*, 2020, **265**, 121787, DOI: [10.1016/j.jclepro.2020.121787](https://doi.org/10.1016/j.jclepro.2020.121787).
- 18 Ö. Çimen Mesutoğlu and O. Gök, Prediction of COD in industrial wastewater treatment plant using an artificial neural network, *Sci. Rep.*, 2024, **14**(1), 13750, DOI: [10.1038/s41598-024-64634-z](https://doi.org/10.1038/s41598-024-64634-z).
- 19 N. E. Huang, Z. Shen and S. R. Long, *et al.*, The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, *Proc. R. Soc. London, Ser. A*, 1998, **454**(1971), 903–995, DOI: [10.1098/rspa.1998.0193](https://doi.org/10.1098/rspa.1998.0193).
- 20 H.-F. Yang and Y.-P. P. Chen, Hybrid deep learning and empirical mode decomposition model for time series applications, *Expert Syst. Appl.*, 2019, **120**, 128–138, DOI: [10.1016/j.eswa.2018.11.019](https://doi.org/10.1016/j.eswa.2018.11.019).
- 21 Z. Gao, J. Chen and G. Wang, *et al.*, A novel multivariate time series prediction of crucial water quality parameters with Long Short-Term Memory (LSTM) networks, *J. Contam. Hydrol.*, 2023, **259**, 104262, DOI: [10.1016/j.jconhyd.2023.104262](https://doi.org/10.1016/j.jconhyd.2023.104262).
- 22 S. Hochreiter and J. Schmidhuber, Long short-term memory, *Neural Comput.*, 1997, **9**(8), 1735–1780, DOI: [10.1162/neco.1997.9.8.1735](https://doi.org/10.1162/neco.1997.9.8.1735).
- 23 Y. Zhang, C. Li and Y. Jiang, *et al.*, Accurate prediction of water quality in urban drainage network with integrated EMD-LSTM model, *J. Cleaner Prod.*, 2022, **354**, 131724, DOI: [10.1016/j.jclepro.2022.131724](https://doi.org/10.1016/j.jclepro.2022.131724).
- 24 Y. Duan, Y. Liu and Y. Wang, *et al.*, Improved BIGRU model and its application in stock price forecasting, *Electronics*, 2023, **12**(12), 2718, DOI: [10.3390/electronics12122718](https://doi.org/10.3390/electronics12122718).



- 25 U. Gupta, V. Bhattacharjee and P. S. Bishnu, StockNet-GRU based stock index prediction, *Expert Syst. Appl.*, 2022, 207, 117986, DOI: [10.1016/j.eswa.2022.117986](https://doi.org/10.1016/j.eswa.2022.117986).
- 26 A. Ariyaei, Y. Kheyruri and S. Doroudi, *et al.*, A hybrid EMD-DFA-LSTM-MOOLBO model for accurate water quality index prediction, *J. Water Process Eng.*, 2025, 77, 108558, DOI: [10.1016/j.jwpe.2025.108558](https://doi.org/10.1016/j.jwpe.2025.108558).
- 27 R. Doroudi, S. H. H. Lavassani and M. Shahrouzi, Optimal tuning of three deep learning methods with signal processing and anomaly detection for multi-class damage detection of a large-scale bridge, *Struct. Health Monit.*, 2024, 23(5), 3227–3252, DOI: [10.1177/14759217231216694](https://doi.org/10.1177/14759217231216694).
- 28 S. Mirjalili and J. S. Dong, *Multi-Objective Optimization Using Artificial Intelligence Techniques*, Springer, Cham, 2020, DOI: [10.1007/978-3-030-24835-2](https://doi.org/10.1007/978-3-030-24835-2).
- 29 J. Tomperi, A. Sorsa, J. Ruuska and M. Ruusunen, Series-connected data-based model to estimate effluent chemical oxygen demand in industrial wastewater treatment process, *J. Environ. Manage.*, 2025, 373, 123680, DOI: [10.1016/j.jenvman.2024.123680](https://doi.org/10.1016/j.jenvman.2024.123680).
- 30 M. T. Albaba, M. Talhami, A. Omar, S. Varghese, R. Akoumeh and M. A. Ayari, *et al.*, Machine learning-aided prediction of COD removal in the electrocoagulation process using a super learner model, *J. Environ. Chem. Eng.*, 2025, 13(5), 117469, DOI: [10.1016/j.jece.2025.117469](https://doi.org/10.1016/j.jece.2025.117469).

