Chemosphere, Volume 376, May 2025, Article number 144299
DOI: 10.1016/j.chemosphere.2025.144299

Enhancing process monitoring and control in novel carbon capture and utilization biotechnology through artificial intelligence modeling: An advanced approach toward sustainable and carbon-neutral wastewater treatment

Stefano Cairone [a], Giuseppina Oliva [a], Fabiana Romano [a], Federica Pasquarelli [a], Aniello Mariniello [a], Antonis A. Zorpas [b], Simon J.T. Pollard [c], Kwang-Ho Choo [d], Vincenzo Belgiorno [a], Tiziano Zarra [a], Vincenzo Naddeo [a,*]

[a] Sanitary Environmental Engineering Division (SEED), Department of Civil Engineering, University of Salerno, Via Giovanni Paolo II #132, 84084 Fisciano, SA, Italy

[b] Laboratory of Chemical Engineering and Engineering Sustainability, Faculty of Pure and Applied Sciences, Open University of Cyprus, Giannou Kranidioti 89, Latsia, 2231 Nicosia, Cyprus

[c] Cranfield University, Water Science Institute, Faculty of Engineering and Applied Sciences, Cranfield, Bedfordshire MK43 0AL, UK

[d] Department of Environmental Engineering, Kyungpook National University (KNU), 80 Daehak-ro, Bukgu, Daegu 41566, Republic of Korea

*Corresponding author: vnaddeo@unisa.it

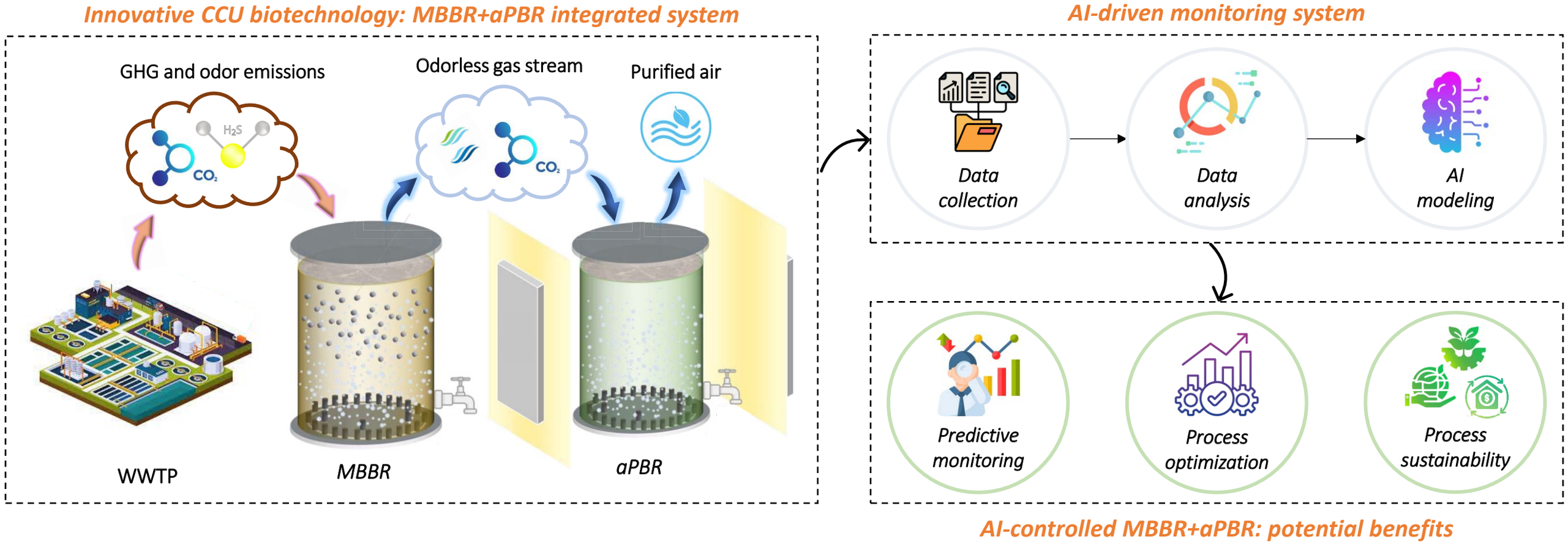
Acknowledgments

This work was supported by the University of Salerno through FARB projects (300393FRB22OLIVA, 300393FRB22NADDE, 300393FRB23NADDE, 300393FRB22ZARRA) and the PhD scholarship awarded to Stefano Cairone (PhD Course in "Risk and Sustainability on Civil Engineering, Architectural and Environmental Engineering Systems" at the Department of Civil Engineering). The authors extend their gratitude to the technical staff of the Sanitary Environmental Engineering Division (SEED) of the University of Salerno, particularly Domenico Giaquinto and Paolo Napodano, for their valuable support. The authors sincerely thank Prof. Ngai Yin Yip (Department of Earth and Environmental Engineering, Columbia University, New York, USA) for his contribution in reviewing the manuscript and providing valuable feedback on this study. Giuseppina Oliva acknowledges her grant funded by the National Operational Programme on Research and Innovation 2014-2020, managed by the Italian Ministry of University and Research (MUR). Additionally, the outcomes of this study have benefited from insights and developments within the SPORE-MED project, part of the PRIMA program funded by the European Union (Agreement 2322).

Highlights

- AI modeling can improve process monitoring and control of CCU biotechnologies.
- The novel MBBR+aPBR system can contribute to the carbon-neutrality of WWTPs.
- AI models were tested to predict CO_2 and H_2S emissions from MBBR+aPBR.
- LSBoost showed the highest prediction accuracy.
- Model selection and optimization are crucial for effective AI modeling.

Graphical Abstract



Enhancing process monitoring and control in novel carbon capture and utilization biotechnology through artificial intelligence modeling: An advanced approach toward sustainable and carbon-neutral wastewater treatment

Abstract

Integrating carbon capture and utilization (CCU) technologies into wastewater treatment plants (WWTPs) is essential for mitigating greenhouse gas (GHG) emissions and enhancing environmental sustainability, but further advancements in process monitoring and control are critical to optimizing treatment performance. This study investigates the application of artificial intelligence (AI) modeling to enhance process monitoring and control in a novel integrated CCU biotechnology with a moving bed biofilm reactor (MBBR) sequenced with an algal photobioreactor (aPBR). This system reduces GHG and odour emissions simultaneously. Several machine learning (ML) models, including artificial neural networks (ANNs), support vector machines (SVM), random forest (RF), and least-squares boosting (LSBoost), were tested. The LSBoost was the most suitable for modeling the MBBR+aPBR system, exhibiting the highest accuracy in predicting CO_2 (R^2 = 0.97) and H_2S (R^2 = 0.95) emissions from the MBBR. LSBoost also achieved the highest accuracy for predicting CO_2 (R^2 = 0.85) and H_2S (R^2 = 0.97) outlet concentrations from the aPBR. These findings underscore the importance of aligning AI algorithms to the characteristics of the treatment technology. The proposed AI models outperformed conventional statistical methods, demonstrating their ability to capture the complex, nonlinear dynamics typical of processes in environmental technologies. This study highlights the potential of AI-driven monitoring and control systems to significantly improve the efficiency of CCU biotechnologies in WWTPs for climate change mitigation and sustainable wastewater management.

Keywords: advanced process control; bioprocess modeling; carbon neutrality; gaseous emission control; integrated algal biotechnology; supervised machine learning; odour treatment technology.

1. Introduction

Climate change is a critical global challenge requiring immediate action (Pahunang et al., 2021). Reducing greenhouse gas (GHG) emissions is a pressing goal across all industrial sectors, including wastewater treatment (Y. He et al., 2023), which accounts for about 1.6 % of global emissions and about 5 % of global non-CO_2 GHG emissions (Lu et al., 2018). This is compounded by the increase in global wastewater production (Qadir et al., 2020), which could further amplify the contribution of wastewater treatment plants (WWTPs) to GHG emissions.

GHG emissions from WWTPs can be direct or indirect. Direct emissions primarily arise from the biological degradation and biotransformation of carbonaceous and nitrogenous pollutants in water and sludge lines; while indirect emissions stem from energy consumption (often derived from non-renewable sources) and the production and transport of chemicals and fuels used in treatment (Lu et al., 2018; Mannina et al., 2018; Yu et al., 2022). Among indirect GHG emissions, electricity use during operation and maintenance represents the largest portion of conventional WWTPs' carbon footprint, emphasizing the need for strategies that improve energy efficiency and self-sufficiency (Guven et al., 2022; X. He et al., 2023; Rani et al., 2022). Advanced approaches for reducing direct GHG emissions are equally critical in the pursuit of carbon neutrality in WWTPs (Lancioni et al., 2024; Pasquarelli et al., 2024a). The growing research interest in these topics is evidenced by the rising number of publications over the past few years. A TITLE-ABS-KEY search using the terms «(“carbon neutral*” OR “carbon-neutral*” OR “self sufficie*” OR “self-sufficie*”) AND (“wastewater treatment” OR “waste water treatment”)» in the Scopus database revealed 849 documents, with 591 papers (around 70% of the total) published since 2021 (**Figure S1**). In particular, several studies have emphasized the urgent need to achieve multiple, sometimes competing objectives of advanced wastewater treatment and low GHG emissions, with a strong focus on achieving carbon neutrality in WWTPs (Cairone et al., 2024e, 2024d; Corpuz et al., 2024; Yang et al., 2022, 2023, 2024b, 2024a). Carbon-neutral wastewater treatment requires the adoption of state-of-the-art

technologies that ensure high contaminant removal while minimizing GHG emissions. One promising strategy is integrating carbon capture and utilization (CCU) in WWTPs to capture and convert CO_2 emissions into valuable resources while substantially reducing GHG emissions (Corpuz et al., 2021; Maselli et al., 2024; Senatore et al., 2022, 2021). CCU technologies not only mitigate the environmental impact of wastewater treatment but also align with circular economy principles by transforming CO_2 into valuable products. This dual functionality makes CCU particularly advantageous, offering environmental and economic benefits. Moreover, CCU addresses the growing global emphasis on resource recovery and the transition toward net-zero emissions, critical goals for sustainable wastewater management. These features make CCU an attractive and forward-thinking solution also for circular economy-driven WWTPs. Various CCU technologies have been proposed, including biological and enzymatic methods, chemical and electrochemical conversion processes, membrane-based separation, liquid absorption, and solid adsorption. Although many of these are energy- and material-intensive, biological technologies stand out for their efficiency, cost-effectiveness, and low energy requirements (Alli et al., 2024; Yang et al., 2024c).

Among the CCU biotechnologies proposed in the prior art, a particularly innovative approach involves combining a moving bed biofilm reactor (MBBR) with an algal photobioreactor (aPBR) in an integrated MBBR+aPBR system (Oliva et al., 2023; Pasquarelli et al., 2024a). This system not only reduces GHG emissions but abates odorous substances such as hydrogen sulfide (H_2S), which present serious social concerns for WWTPs. The effective management of odour emissions is essential given the elevated sensitivity of surrounding communities to odours (De Paola et al., 2024). In the MBBR+aPBR system, bacterial activity within the MBBR degrades and oxidizes undesirable compounds like H_2S, and produces CO_2, which is utilized by microalgae in the aPBR, thus forming a CCU system. This approach promotes carbon neutrality and aligns with circular economy principles because the harvested microalgae can be processed into biofuels and other high-value products (Oliva et al., 2024; Zarra et al., 2024).

Recent studies also highlight the importance of developing strategies for automation, optimization and the advanced control of treatment processes. Adapting operational conditions in real-time to the needs of the treatment process can enhance WWTP performance and optimize resource use (Cairone et al., 2024a). Integrating real-time monitoring with predictive modeling offers substantial benefits. Artificial intelligence (AI) is now emerging as a powerful tool for creating advanced models that can optimize WWTP operations, overcoming the limitations of conventional mechanical/mathematical models (Cairone et al., 2024b, 2024c; Kamali et al., 2021; Nam et al., 2023; Newhart et al., 2019; Zhang et al., 2023). While conventional models struggle with the manifold complexity and uncertainties of operating conditions, AI models can outperform them in identifying intricate data patterns and nonlinear relationships between inputs and outputs, leading to more accurate predictions (Li et al., 2019; Niu et al., 2022; Zhang et al., 2019). These predictions can, in turn, provide useful insights for real-time process optimization and control, even in the complex and dynamic environments typical of bioprocesses (Cheng et al., 2023). By optimizing operational conditions, AI-based process control can yield significant environmental, economic, and social benefits (Alhajeri et al., 2024; Cairone et al., 2024a; Serna-Carrizales et al., 2024). These benefits include increased contaminant removal, reduced GHG and odour emissions, optimized energy and chemical consumption, enhanced public health protection, and reduced operating costs (Cairone et al., 2024a).

This study explores the potential of AI modeling for process monitoring and control in the novel MBBR+aPBR integrated system, providing valuable insights for sustainable and carbon-neutral wastewater treatment. To the best of the authors' knowledge, no prior studies have investigated AI modeling to support the monitoring system of CCU biotechnologies implemented in WWTPs. Several AI models were tested and compared for their performance in predicting CO_2 and H_2S outlet concentrations from the MBBR+aPBR system. The potential benefits, challenges, and future perspectives of the proposed approach are discussed.

2. Materials and methods

2.1 Materials and experimental set-up

The laboratory-scale MBBR+aPBR integrated system consisted of an MBBR with an operating volume of 10 L, inoculated with activated sludge, and an aPBR with an operating volume of 26 L, inoculated with microalgae (**Figure 1**). This configuration provides complementary treatment capabilities. The MBBR degrades organic matter and odorous compounds efficiently, including H_2S. The aPBR abates these odorous compounds further while facilitating the fixation of CO_2 produced by bacterial activity in the MBBR, using it for metabolic processes during algal growth.

A synthetic solution simulating dairy wastewater (DWW) was fed continuously into the bioreactors. Pure H_2S was injected into the system using an automatic syringe pump (New Era Pump Systems, Inc., Model 300), its concentration regulated by dilution with ambient air using a compressor (Stanley DST 100/8/6) equipped with a flowmeter set to 2 L/min. One-third of the MBBR's operating volume was filled with plastic carriers (Kaldnes Ring, Amitec, Italy) to support biomass growth and increase the contact time between the raw gas and biomass. The aPBR was stirred continuously at 300 rpm and the temperature was maintained at 25 ± 1 °C. This promotes optimal growth conditions for the microalgae. In the aPBR, light plays a critical role in driving photosynthesis, which aids CO_2 sequestration and O_2 production, contributing to the overall efficiency of the treatment process (Guedes et al., 2023; Vergara et al., 2016). To promote photosynthesis in the algal biomass, artificial light was provided by four white LED bulbs (luminous intensity: 8258 LUX) arranged around the aPBR, with a 12-hour light/12-hour dark cycle. An extensive description of the system is available in previous studies (Oliva et al., 2023; Pasquarelli et al., 2024a).

2.2 AI modeling

Machine learning (ML) algorithms were implemented to predict the performance of the MBBR+aPBR integrated system. Specifically, artificial neural networks (ANNs), support vector

machine (SVM), random forest (RF), and least-squares boosting (LSBoost) were tested. These were selected for their suitability in regression tasks and their ability to effectively handle complex data patterns. Modeling was conducted using MATLAB® (Release 2024a, MathWorks, Inc., Natick, Massachusetts, USA). To enhance model accuracy and robustness, hyperparameter optimization was performed for each tested model using a grid search approach.

The modeling process began with separate analyses of the MBBR (Section 3.1) and aPBR (Section 3.2) to identify the best-performing supervised learning algorithm for each treatment unit. After selecting the optimal algorithms, the models were integrated by using the outputs (Section 2.2.1) from the MBBR model as inputs for the aPBR model (Section 3.3). This integration allowed for an accurate representation of the overall system by capturing the interactions between the two treatment units.

2.2.1 Data collection and preprocessing

Data quality and quantity are key factors in ML modeling. The current study relies on data from the laboratory-scale MBBR+aPBR integrated system described in Section 2.1. Key operational parameters, including pH, temperature, dissolved oxygen (DO), saturated oxygen, total suspended solids (TSS), and concentrations of CO_2 and H_2S at the inlet and outlet, were monitored to evaluate the performance of the MBBR+aPBR system. Temperature, pH, DO, and saturated O_2 were measured using a multi-parameter probe (Hanna Instruments), cleaned regularly, and calibrated following the manufacturer's guidelines. CO_2 and H_2S concentrations were measured using a multiparameter gas analyzer (MultiRAE, Recom Industriale S.r.l., Italy) and a gas chromatograph equipped with a thermal conductivity detector (GC–TCD, Thermo Fischer Scientific, USA). The standard method was applied to determine TSS content (APAT, 2003). All parameters were measured daily during the experimental period to ensure an accurate representation of system dynamics.

Data preprocessing involved multiple steps to improve dataset quality and enhance the accuracy of the AI models (Maharana et al., 2022; Rios Fuck et al., 2024). Outlier detection was performed using the interquartile range (IQR) method, with outliers defined as data points falling below $Q_1 - 1.5 \times$

IQR or above $Q_3 + 1.5 \times IQR$, where Q_1 and Q_3 represent the first and third quartiles, respectively, and $IQR = Q_3 - Q_1$ (Bahramian et al., 2023). In the missing data management step, detected outliers were removed and replaced using linear interpolation between neighboring values, assuming a linear trend between adjacent data points. This approach aligns with previous studies (Bu et al., 2024; Zaghloul and Achari, 2022). Notably, the dataset exhibited high quality, with only four data points identified as outliers. A summary of statistical parameters (mean values, standard deviations, and coefficient of variation) for the MBBR and aPBR data is presented in **Table 1**. The cleaned dataset was then split into training and testing sets to develop robust models capable of generalizing to unseen data (Xu and Goodacre, 2018). Aligning to previous studies in AI modeling, the dataset was divided, with 75% of the data utilized for training and the remaining 25% for testing. This strategy, widely adopted in AI studies (Kovacs et al., 2022; Nam et al., 2021; Nourani et al., 2021), ensures a sufficient amount of data is allocated for training while reserving an appropriate portion for testing. The training dataset underwent a feature engineering process, including feature selection, to exclude irrelevant features. This step aimed to retain features critical for modeling by removing those that do not substantially contribute (or may even negatively impact) model outcomes. Feature selection reduces noise, improves model learning efficiency, and minimizes computational overhead (Zhu et al., 2023). Multicollinearity and correlation analyses were conducted to identify highly correlated variables, either among features or between features and response variables, using the Pearson correlation coefficient (r). Removing highly correlated features is necessary to prevent redundancy in the model, while retaining features with higher r values relative to the response variables is generally beneficial, as they provide meaningful information for prediction. However, excluding features solely because they do not exhibit high r values relative to the response variables may not be appropriate, as this approach does not account for potential nonlinear relationships that the Pearson correlation analysis cannot capture. Based on the correlation matrix results (**Figure S2**), saturated oxygen and DO were found to be highly correlated for both MBBR and aPBR systems, which is reasonable. Consequently, saturated oxygen was excluded from the input variables, leaving six selected features: temperature,

pH, DO, TSS, CO_2 inlet, and H_2S inlet. Following feature selection, feature scaling was applied to standardize or normalize the ranges of predictor variables. This step improves model performance by minimizing overfitting risks and reducing errors caused by differences in parameter scales (Aghdam et al., 2023; Sadoune et al., 2023).

2.2.2 Model evaluation

Once the model is developed and optimized, its performance is assessed and compared with other models based on specific metrics for regression tasks. Specifically, model performance was evaluated using, as metrics, the Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Coefficient of Determination (R^2). These metrics provide a comprehensive assessment of the models' accuracy and goodness of fit by comparing predicted values against the actual test set data. To improve model interpretability, the best-performing model was analyzed using permutation feature importance (PFI), a method that ranks features based on their contribution to predicting the response variable. This approach evaluates the change in prediction error when a feature's values are randomly permuted, with larger increases in error indicating greater importance.

Comparisons with two well-established statistical methods, multiple linear regression (MLR) and partial least squares (PLS), were conducted to assess whether the proposed AI-based modeling approach offers superior predictive performance over traditional methods.

2.2.3 Artificial neural network (ANN)

ANNs are computational models inspired by the structure and function of the human brain. They consist of interconnected nodes (or neurons) organized into input, hidden, and output layers. The training process enables the ANN to learn from data by adjusting neuron weights. In this study, a feedforward neural network (FF-NN), specifically a multilayer perceptron (MLP), was employed. In an FF-NN, information flows unidirectionally from input to output neurons, never moving backward. The architecture of an ANN, including the number of hidden layers and the number of neurons per

layer, is crucial because it determines the network's capacity to learn and generalize from data. To optimize the ANN architecture, the number of hidden layers and neurons per layer were systematically varied between 1 and 25, yielding 625 potential configurations. The optimal configuration was selected based on minimizing MSE, RMSE, and MAE while maximizing the R^2 value. Grid search was conducted to optimize the learning rate hyperparameter, with tested values of 0.01, 0.1, 0.25, and 0.5. The sigmoidal transfer function ("tansig") and the Levenberg-Marquardt training algorithm ("trainlm") were chosen for their efficacy (Cifuentes-Cabezas et al., 2023; Hosseinzadeh et al., 2020). The optimized ANN was used to make predictions, which were compared to actual values for accuracy assessment.

2.2.4 Support vector machine (SVM)

SVMs are supervised learning models used for classification and regression tasks. They work by separating data into two classes using a hyperplane, or a set of hyperplanes, in a high-dimensional space, aiming to find the optimal hyperplane that maximizes the margin, i.e., the largest distance between the hyperplane and the nearest data points (called support vectors) from opposite classes. The data can be separated using hyperplanes described by linear or nonlinear functions. Both linear and nonlinear kernel functions were tested in this study, including linear, Gaussian radial basis function (RBF), and polynomial kernels (degree of 2, 3 and 4). Hyperparameter optimization was performed through grid search and cross-validation, varying values for Gamma (0.001, 0.01, 0.1, and 1) and the box constraint (or regularization, C = 0.1, 1, 10, 100, 250, 500, 600, 750, 800, and 1000). A k-fold cross-validation (with k = 2, 5, and 10) was conducted to identify the most robust configuration.

2.2.5 Random forest (RF)

RF is an ensemble learning method suitable for classification and regression tasks. RF models generate multiple decision trees, each trained on a random subset of the training dataset. The final prediction is made by aggregating the outputs of all individual trees. The number of decision trees is

a key hyperparameter affecting model accuracy. Although increasing the number of trees in an RF model generally improves its accuracy, improvement often reaches a plateau beyond a certain number of trees, yielding diminishing returns while increasing computational costs. Therefore, it is crucial to balance model accuracy with computational efficiency. In this study, RF models were optimized by varying the number of decision trees (50, 100, 250, 500, 600, 700, 800, and 1000). A k-fold cross-validation (with k = 2, 5, and 10) was conducted to evaluate each configuration and determine the optimal model.

2.2.6 Least-squares boosting (LSBoost)

LSBoost is a gradient boosting algorithm designed for regression tasks. It builds an ensemble of weak learners, typically decision trees, sequentially. Each subsequent tree in the ensemble is trained to correct the residual errors of the previous trees, optimizing the model by minimizing the least squares error. Key hyperparameters, including the learning rate (0.01, 0.1, 0.25, 0.5, and 1), the number of boosting trees (50, 100, 250, 500, 600, 700, 800, and 1000), and the subsample (0.4, 0.6, 0.8, 1), were optimized through grid search. A 5-fold cross-validation was used to identify the best-performing LSBoost model configuration.

3. Results and discussion

3.1 MBBR modeling

This section presents the results of the AI models used to predict the performance of the MBBR.

The ANN model exhibited high accuracy in predicting CO_2 outlet concentrations (**Figure 3a**) but showed reduced performance for H_2S outlet concentrations (**Figure 2a**). Specifically, while the ANN model provided relatively accurate predictions for early H_2S samples, its performance diminished for later samples. This reduction may reflect the difficulty in capturing the temporal variability or nonlinear nature of the H_2S inlet and outlet concentrations, indicating that improvements in model

249 architecture or additional input features may be necessary to account for these dynamics. The
250 optimization process identified the optimal ANN architecture for H_2S predictions, comprising 14
251 hidden layers with 14 neurons each, whereas CO_2 predictions were best modeled with 19 hidden
252 layers containing 15 neurons each. Additionally, an ANN capable of simultaneously predicting both
253 CO_2 (**Figure 3b**) and H_2S (**Figure 2b**) outlet concentrations (multi-task ANN) was tested and
254 optimized. The best configuration for this combined prediction task consisted of 16 hidden layers
255 with 7 neurons per layer. The ability to predict both CO_2 and H_2S outlet concentrations with a single
256 ANN model suggests that a multi-task approach could effectively balance predictive performance
257 across distinct target outputs, potentially reducing the need for separate models.

258 The SVM model exhibited high accuracy in predicting both H_2S (**Figure 2c**) and CO_2 (**Figure 3c**)
259 outlet concentrations. The SVM's capacity to capture the patterns in both emissions highlights its
260 potential in specialized tasks but also underscores the challenges of applying the same model
261 architecture to both emissions without additional refinement. In fact, optimal hyperparameters for the
262 SVM were determined as Gamma = 1 and C = 500 for H_2S predictions, while Gamma = 1 and C =
263 800 were preferable for CO_2. This difference in hyperparameter values reflects the necessity of fine-
264 tuning SVM models to meet the specific objectives of the modeling task.

265 The RF model showed relatively poor accuracy in predicting H_2S outlet concentrations (**Figure 2d**)
266 but performed better for CO_2 emissions (**Figure 3d**). The optimal RF configuration involved 500
267 decision trees for H_2S predictions and 600 decision trees for CO_2 predictions. The observed
268 discrepancy in performance between H_2S and CO_2 suggests that further optimization of the RF model
269 is needed, possibly by experimenting with different tree depths or incorporating additional feature
270 engineering. This may help the model better capture the complexity of H_2S emissions patterns, which
271 appear more challenging for RF to predict accurately. These conditions can be attributed to the
272 substantial variability in H_2S inlet concentrations (Cv = 0.28), resulting from its real-time artificial
273 generation through an extemporaneous chemical reaction that is sensitive to environmental conditions

(Morse et al., 1987; Reddy et al., 2019). This variability provides an opportunity to verify the system's robustness in handling large fluctuations in inlet load. On the other hand, outlet concentrations exhibited great variability, which is challenging to predict using "conventional" AI modeling approaches.

LSBoost demonstrated strong performance in predicting both H_2S (**Figure 2e**) and CO_2 (**Figure 3e**) outlet concentrations. LSBoost's sequential learning approach, which focuses on minimizing residual errors, contributed to its superior performance by effectively capturing complex emission patterns. The optimal configuration consisted of 800 decision trees and a learning rate 0.5, which was applied to both H_2S and CO_2 predictions. The robustness of the LSBoost model highlights its potential as a reliable and versatile tool for environmental system modeling, particularly in scenarios involving complex and nonlinear data.

Table S1 summarizes the evaluation metrics for the AI models in predicting H_2S and CO_2 outlet concentrations from the MBBR. According to the results, LSBoost achieved the best performance in predicting CO_2 emissions (MSE = 0.006; R^2 = 0.967), followed by SVM (MSE = 0.011; R^2 = 0.938), ANN (MSE = 0.012; R^2 = 0.933), and RF (MSE = 0.022; R^2 = 0.872). For H_2S emissions, LSBoost again outperformed the other models (MSE = 0.011; R^2 = 0. 945), followed by SVM (MSE = 0.013; R^2 = 0.934), ANN (MSE = 0.049; R^2 = 0.758), and RF (MSE = 0.096; R^2 = 0.528). These findings underscore LSBoost's versatility, as it demonstrated high accuracy in predicting both CO_2 and H_2S outlet concentrations. Similarly, SVM showed accurate predictions for both CO_2 and H_2S outlet concentrations, while ANN exhibited specialized performance for CO_2 but reduced accuracy for H_2S. In contrast, RF exhibited relatively low accuracy in predicting both CO_2 and H_2S outlet concentrations. This emphasizes the importance of selecting the appropriate model based on specific emission characteristics. The tested AI models consistently outperformed the statistical methods for predicting both H_2S and CO_2 outlet concentrations from the MBBR (**Table S1**). Specifically, while the MLR (**Figure S3b**) and PLS (**Figure S4b**) methods performed reasonably well for CO_2

predictions (MSE = 0.03; R^2 = 0.8), their accuracy dropped dramatically for H_2S emissions (**Figure S3a** and **Figure S4a**), with both methods yielding an R^2 of only 0.06 (MSE = 0.2). This disparity may be attributed to the relatively strong correlation between CO_2 outlet concentrations and most features, compared to the absence of high correlations between H_2S outlet concentrations and the features (**Figure S2a**). These results highlight the limitations of conventional statistical methods in capturing complex, nonlinear relationships among variables, which are common in environmental technologies.

Analyzing the PFI results of the best-performing LSBoost model predicting MBBR performance, it is evident that all features had approximately equal importance for H_2S outlet predictions, with H_2S inlet and TSS showing slightly greater importance compared to other features (**Figure S7a**). In contrast, for CO_2 outlet predictions, the LSBoost model for MBBR primarily relied on DO and H_2S inlet, which exhibited higher importance relative to the other features (**Figure S7b**). It is essential to emphasize that a feature's importance reflects its contribution to the model's predictive mechanism and does not necessarily correspond to its relevance in the underlying biophysicochemical processes. Therefore, feature importance analysis serves as a tool for model interpretability rather than a direct representation of real-world processes. This point will be further discussed in Section 4.

3.2 aPBR modeling

This section discusses the results of the AI models used to predict the performance of the aPBR.

The ANN model showed high accuracy in predicting H_2S outlet concentrations (**Figure 4a**) but lower accuracy for CO_2 outlet concentrations (**Figure 5a**). The optimal architecture for H_2S predictions consisted of 4 hidden layers with 8 neurons per layer, while the best architecture for CO_2 predictions involved 3 hidden layers with 11 neurons per layer. Additionally, a multi-task ANN capable of simultaneously predicting both CO_2 (**Figure 5b**) and H_2S (**Figure 4b**) outlet concentrations was optimized. The optimal configuration, with 15 hidden layers and 23 neurons each, suggests that multi-task approaches can efficiently balance predictive performance across output targets, which may be valuable in modeling complex environmental technologies.

The SVM model achieved high accuracy for H_2S outlet concentrations (**Figure 4c**) but performed poorly for CO_2 outlet concentrations (**Figure 5c**). The optimal hyperparameters for both emissions were Gamma = 1 and C = 800. While the SVM showed strong performance for H_2S, further adjustments are needed to improve CO_2 predictions, possibly by refining the model's feature selection or architecture.

The RF model exhibited low accuracy for both H_2S (**Figure 4d**) and CO_2 (**Figure 5d**) outlet concentrations. The optimal RF configuration included 700 decision trees for H_2S predictions and 800 decision trees for CO_2. However, the consistently low performance indicates that RF may not be well-suited for this task without further optimization, such as modifying tree depths or incorporating more relevant features. Specifically, in real-scale applications, it would be feasible to integrate predicted H_2S inlet concentrations into the model, accounting for process conditions (Zwain et al., 2020). By incorporating H_2S inlet concentrations to the aPBR as an additional parameter to be predicted, the model can provide a more comprehensive representation of the H_2S dynamics within the MBBR+aPBR system. This integration enables a complete prediction of the complex processes involving H_2S dynamics across the system, enhancing the reliability of outlet concentration predictions. Consequently, including H_2S inlet concentrations strengthens the model's overall predictive accuracy and supports optimizing the system's performance.

LSBoost demonstrated very high accuracy in predicting both H_2S (**Figure 4e**) and CO_2 (**Figure 5e**) outlet concentrations. The optimal configuration involved 800 decision trees and a learning rate of 0.5 for both outlet concentrations.

Table S2 summarizes the evaluation metrics for AI models in predicting H_2S and CO_2 outlet concentrations from the aPBR. LSBoost achieved the best performance in predicting both H_2S (MSE = 0.002; R^2 = 0.965) and CO_2 (MSE = 0.001; R^2 = 0.851) outlet concentrations, underscoring its efficacy in modeling complex emissions patterns. The multi-task ANN capable of simultaneously predicting CO_2 and H_2S outlet concentrations also performed relatively well. This may be attributed

to its capacity to identify shared underlying patterns and interdependencies between CO_2 and H_2S outlet concentrations, highlighting the benefit of multi-task learning in complex systems. In contrast, ANN and SVM demonstrated specialized performance for CO_2, with R^2 values of 0.897 and 0.934, respectively, but performed poorly for H_2S, with R^2 values of 0.660 and 0.679, respectively. Consistent with observations for MBBR modeling (Section 3.2), RF exhibited relatively low accuracy in predicting both CO_2 (MSE = 0.013; R^2 = 0.750) and H_2S (MSE = 0.003; R^2 = 0.467) outlet concentrations. These findings further confirm the importance of carefully selecting the algorithm most suitable for specific modeling goals. The tested AI models consistently outperformed the statistical methods in aPBR modeling (**Table S2**). Specifically, both the MLR (**Figure S5**) and PLS (**Figure S6**) methods performed poorly for both H_2S (R^2 = 0.510 and 0.575, respectively) and CO_2 emissions (R^2 = 0.146 and 0.106, respectively). As highlighted in the MBBR modeling (Section 3.1), these results emphasize the ability of AI/ML models to handle complex, nonlinear relationships that traditional statistical methods may not effectively capture.

Consistent with the observations for MBBR modeling (Section 3.1), the PFI results revealed that all features had approximately equal importance in the best-performing LSBoost model for predicting H_2S outlet concentrations from the aPBR (**Figure S8a**). As observed for MBBR modeling (Section 3.1), the LSBoost model for predicting CO_2 outlet concentrations from the aPBR attributed varying importance to each feature, with temperature and DO showing the highest importance in this case (**Figure S8b**). As discussed in Section 3.1, it is important to reiterate that these results do not necessarily reflect the role of these parameters in the actual treatment process. Instead, they represent the parameters' contribution to the model's predictive mechanisms (Section 4).

3.3 MBBR+aPBR integrated system modeling

Based on the results discussed in section 3.1, the LSBoost algorithm was selected for modeling the MBBR due to its superior prediction performance for both CO_2 and H_2S outlet concentrations. The outputs from the optimal LSBoost model were used to predict the performance of the aPBR, enabling

the estimation of emissions from the MBBR+aPBR integrated system. As shown in section 3.2, the LSBoost model also achieved the highest accuracy for H_2S and CO_2 outlet concentrations from the aPBR. Therefore, the LSBoost+LSBoost integrated ML approach was employed to model the entire process, predicting H_2S and CO_2 outlet concentrations at the end of the treatment (**Figure 6**). The results, summarized in **Table S3**, highlight the high prediction accuracy for both H_2S (MSE = 0.01; R^2 = 0.90) and CO_2 (MSE = 0.00; R^2 = 0.83) outlet concentrations.

The main strength of this integrated modeling approach is its capability to predict the entire treatment process's behaviour without the need for direct measurement of gaseous emissions from the MBBR, thereby reducing costs and time associated with gas analysis. However, inaccuracies in predicting emissions from the MBBR could propagate to the aPBR predictions, reducing overall accuracy. Specifically, a comparison between using measured concentrations of gaseous compounds at the MBBR outlet and those estimated through modeling shows a slight reduction in predictive performance when relying on modeled data. This reduction might be attributed to the high fluctuation of H_2S inlet concentrations and the influence of various mechanisms on the CO_2 mass balance within the sequenced system, making computed estimations complex (Pasquarelli et al., 2024b). Among these mechanisms, sulfate oxidation driven by sulfur-oxidizing bacteria (SOB) plays a key role, as it competes with CO_2 fixation in algal biomass, ultimately contributing to carbon dioxide reduction (Syed et al., 2006). Additionally, the oxidation of sulfate in the liquid medium may create mixotrophic conditions by accelerating the breakdown of biodegradable organic matter present in the DWW influent (José Carlos De et al., 2024). The release of CO_2 from the oxidation of organic matter in DWW could therefore be considered significantly affected by SOB metabolism (Klatt and Polerecky, 2015; Pudi et al., 2022). This effect becomes more pronounced as H_2S inlet concentrations increase, resulting in a simultaneous rise in CO_2 release from the breakdown of carbonyl compounds, along with enhanced nutrient uptake (Pasquarelli et al., 2024a). Furthermore, pH is a parameter that can significantly influence CO_2 assimilation and dissolution. In the investigated conditions, pH remained

within a narrow range, without significantly impacting the CO_2 mass balance (Meier et al., 2018; Omri et al., 2011).

Despite this complexity, the promising results from the LSBoost+LSBoost approach demonstrate its potential as an effective tool for enhancing the monitoring and control of gaseous emissions in CCU biotechnologies. Since this 'integrated modeling' approach relies on MBBR model outputs as input variables for predicting aPBR performance, some inherent inaccuracies are unavoidable, leading to a decrease in overall accuracy compared to a model based solely on measured data. However, this approach could be further optimized by incorporating more advanced techniques or algorithms to reduce these inaccuracies in predicting the MBBR outlet concentrations used as inputs for the aPBR. This would, in turn, improve the overall accuracy of the 'integrated modeling' approach in predicting the performance of the MBBR+aPBR integrated system.

4. AI-controlled MBBR+aPBR integrated system: potential benefits and current challenges

AI models have demonstrated promising potential in predicting the performance of the MBBR+aPBR integrated system for controlling CO_2 and H_2S. More broadly, these models could support the development of AI-controlled MBBR+aPBR systems, potentially enabling improvements in real-time monitoring and process optimization. In such smart systems, AI modeling could offer valuable insights to dynamically adjust critical operational parameters, such as aeration rates in the MBBR, nutrient supply, and light intensity in the aPBR, based on predictive models' outputs. This adaptive approach can potentially enhance treatment efficiency while simultaneously reducing energy consumption and operating costs. Additionally, AI-driven predictive control may help stabilize system performance under fluctuating environmental conditions and varying loads, supporting more consistent treatment efficiency.

The monitored operating parameters (pH, temperature, DO, saturated oxygen, and TSS) play distinct yet interconnected roles in the treatment processes of the MBBR+aPBR system. The operating pH influences not only microbial activity and chemical reactions in the bioreactors but also nutrient assimilation, making pH control crucial to ensure an optimal environment for both bacteria and microalgae. For example, an acidic pH can inhibit the growth of microbes responsible for H_2S biodegradation in the MBBR (Lebrero et al., 2014), while in the aPBR, pH significantly affects microalgal growth by affecting the solubility and availability of CO_2 and carbonates (Yan et al., 2016). Maintaining optimal temperature is equally important, as it affects the metabolic rates of the microbial community in the MBBR and microalgal growth in the aPBR. Additionally, temperature influences the solubility of gases like O_2 and CO_2, which impacts mass transfer and overall system efficiency (Barten et al., 2021). DO and saturated O_2 are critical for aerobic degradation in the MBBR, where O_2 supports microbial respiration and the degradation of organic matter. In the aPBR, O_2 is a by-product of photosynthesis, but excessively high DO concentrations can limit microalgal growth (Fu et al., 2019). TSS concentrations serve as an indicator of biomass concentration and treatment efficiency. Low TSS values may suggest insufficient biological activity, while excessively high concentrations can compromise system performance by reducing aPBR effectiveness due to limited light penetration (Foladori et al., 2020; Soaudy et al., 2023). Overall, each of these operating parameters plays a crucial role in treatment performance, and implementing efficient monitoring and adjustment strategies for process parameters can significantly enhance treatment efficiency. Similarly, monitoring CO_2 and H_2S concentrations at both the inlet and outlet is essential for assessing the performance of the MBBR+aPBR system, as these gases have been selected to reflect the system's effectiveness in controlling GHGs and odour emissions, respectively. While this discussion highlights the relevant role of each monitored parameter in the processes occurring within the MBBR+aPBR integrated system, the results of the feature importance analysis (**Figure S7** and **Figure S8**) reveal that the 'reasoning' behind AI models' predictive mechanisms may sometimes be inconsistent with domain knowledge. For instance, the model may attribute lower importance to variables that are

widely recognized as key parameters for a given process. To enhance model interpretability and explainability, there has been growing interest in adopting more sophisticated interpretability methods and explainable AI techniques (XAI), such as Shapley Additive exPlanations, (SHAP) and Local Interpretable Model-agnostic Explanations (LIME) (Majnooni et al., 2024; Onay and Onay, 2024). However, it is important to emphasize that aligning model interpretability and explainability with the actual mechanisms driving a biophysicochemical phenomenon remains a challenging task that requires critical evaluation by experts (Zhu et al., 2023).

The integration of AI into the MBBR+aPBR system introduces advanced monitoring and early warning capabilities, which are key strengths of this approach. Developing digital twins by integrating AI models with real-time monitoring systems could enable continuous emissions tracking (e.g., CO_2 and H_2S), support predictive tasks, and detect anomalies that may help identify process inefficiencies or potential system failures. This proactive strategy would support rapid intervention, potentially minimizing downtime and improving system resilience. Consequently, an AI-controlled MBBR+aPBR system could operate more efficiently, with the possibility of reduced energy demands, improved contaminant removal efficiency, and a lower carbon footprint, contributing to a more sustainable wastewater treatment process.

Despite these potential advantages, several challenges remain in integrating AI into treatment technologies, such as MBBR+aPBR system. One critical issue is the need for high-quality and comprehensive datasets to ensure accurate model training and validation. In many WWTPs, data are either limited or insufficiently processed or analyzed (Bahramian et al., 2023). To address this, robust sensor networks and advanced data management systems are essential to provide reliable inputs for AI models, enabling accurate predictions and dynamic control adjustments. Creating global, comprehensive, and standardized databases that integrate data from multiple WWTPs offers a potential solution, promoting the applicability of AI-driven strategies. Several valuable features can be incorporated into the modeling to enhance the prediction performance. Among these, H_2S inlet

concentrations can be estimated using process monitoring data, which allows for improved predictions of H_2S outlet concentrations, as extensively discussed in the existing literature (Baawain et al., 2019; Hamad et al., 2024; Santos et al., 2006; Tian et al., 2021; Zwain et al., 2020). Significant challenges are also related to the "black-box" nature of AI models, which hinders operators' understanding of the decision-making process behind AI-driven outcomes (Holzinger et al., 2022; Tsang and Benoit, 2023). CO_2 assimilation and release are instead influenced by a combination of different interdependent mechanisms, including the degradation of carbonyl compounds, which increase CO_2 emissions while simultaneously enhancing nutrient utilization and potentially improving CO_2 fixation (Oliva et al., 2023; Pasquarelli et al., 2024b). Although a black-box approach may not explicitly capture such mechanisms during modeling, the collection of a sufficient amount of high-quality data and the implementation of appropriate preprocessing and feature engineering strategies in this study enabled accurate predictions of CO_2 emissions from the sequenced system. This paves the way for AI to model processes difficult to describe with conventional mathematical approaches. Future research should focus on developing and optimizing explainable and interpretable AI models, along with user-friendly interfaces that allow operators to visualize and comprehend the logic behind AI predictions. Infrastructure compatibility is also a crucial hurdle. Many existing WWTPs are not designed to integrate sophisticated AI-driven control strategies, and upgrading these facilities may require substantial investment in implementing advanced control systems, smart sensors, and automation infrastructure (Cairone et al., 2024a). While the initial cost may be high, the long-term environmental and economic benefits, including reduced operational costs and improved treatment performance, make such investment feasible. Additionally, training WWTP operators in AI technologies and introducing new roles, such as data analysts and AI/ML specialists, will be crucial for the effective implementation of AI-driven control strategies (Cairone et al., 2024b).

While the findings of this study demonstrate the potential of AI modeling in enhancing process monitoring and control of the MBBR+aPBR integrated system, several limitations and uncertainties

must be considered. First, the models were evaluated using a relatively limited dataset from a laboratory-scale setup, which may not fully capture the variability of real-world operating conditions in full-scale WWTPs. Additionally, although LSBoost exhibited the highest accuracy in predicting CO_2 and H_2S emissions from the MBBR+aPBR system, the generalizability of this model to other CCU systems or varying plant conditions remains uncertain. The study also did not explore the impact of external factors, such as fluctuations in environmental conditions or raw wastewater composition, on model performance. Furthermore, the study focused on a single biotechnology system, and further research is needed to assess the applicability of these AI approaches across diverse CCU technologies. In future studies, the proposed approach can be extended to other wastewater treatment and CCU technologies, helping to demonstrate the broader potential of AI in optimizing operational parameters across different treatment systems and driving advancements toward carbon-neutral wastewater treatment. Cooperation among researchers and plant managers will be essential in addressing current challenges and fully realizing the benefits of AI-driven control systems for optimizing wastewater treatment processes.

5. Conclusion

This study demonstrates the effectiveness of AI models in accurately predicting the performance of CCU biotechnologies, focusing on the novel MBBR+aPBR integrated system. The optimized AI models outperformed conventional statistical methods. Among the tested models, LSBoost emerged as the most accurate, achieving high prediction accuracy for CO_2 and H_2S outlet concentrations from the MBBR+aPBR system. The varying predictive performances across the models deployed underscore the importance of selecting the appropriate AI model tailored to the specific characteristics of the treatment technology and the target outputs.

The proposed AI-driven approach provides valuable insights into developing accurate predictive tools, contributing significantly to enhanced process monitoring and control systems. The advanced

predictive capabilities of these AI models offer the potential for real-time monitoring, enabling operators to receive early warnings about critical operational conditions. This approach allows timely interventions, reducing potential issues that could otherwise impact treatment performance. Integrating an advanced AI-driven monitoring and control system within the MBBR+aPBR setup can lead to substantial benefits, including improved degradation of undesirable compounds such as H_2S and reductions in CO_2 emissions. These advancements may significantly mitigate the negative environmental impacts associated with WWTPs.

To fully exploit the potential of the proposed approach, future research should address limitations, such as enhancing prediction accuracy and validating the models across a broader range of datasets. Comparative studies with other ML techniques will be essential in further improving the practical utility of these models. This research direction is crucial for advancing scientific and technical progress toward achieving carbon neutrality and sustainability in wastewater treatment.

References

Aghdam, E., Mohandes, S.R., Manu, P., Cheung, C., Yunusa-Kaltungo, A., Zayed, T., 2023. Predicting quality parameters of wastewater treatment plants using artificial intelligence techniques. Journal of Cleaner Production 405, 137019. https://doi.org/10.1016/j.jclepro.2023.137019

Alhajeri, N.S., Tawfik, A., Nasr, M., Osman, A.I., 2024. Artificial intelligence-enabled optimization of Fe/Zn@biochar photocatalyst for 2,6-dichlorophenol removal from petrochemical wastewater: A techno-economic perspective. Chemosphere 352, 141476. https://doi.org/10.1016/j.chemosphere.2024.141476

Alli, Y.A., Bamisaye, A., Bamidele, M.O., Etafo, N.O., Chkirida, S., Lawal, A., Hammed, V.O., Akinfenwa, A.S., Hanson, E., Nwakile, C., Kazeem, K.O., Ayanwunmi, R.J., Ige, A.S., Parga Torres, J.R., Al Nageim, H., 2024. Transforming waste to wealth: Harnessing carbon dioxide for sustainable solutions. Results in Surfaces and Interfaces 17, 100321. https://doi.org/10.1016/j.rsurfi.2024.100321

Baawain, M., Al-Mamun, A., Omidvarborna, H., Al-Sulaimi, I.N., 2019. Measurement, control, and modeling of H2S emissions from a sewage treatment plant. Int. J. Environ. Sci. Technol. 16, 2721–2732. https://doi.org/10.1007/s13762-018-1997-z

Bahramian, M., Dereli, R.K., Zhao, W., Giberti, M., Casey, E., 2023. Data to intelligence: The role of data-driven models in wastewater treatment. Expert Systems with Applications 217, 119453. https://doi.org/10.1016/j.eswa.2022.119453

Barten, R., Djohan, Y., Evers, W., Wijffels, R., Barbosa, M., 2021. Towards industrial production of microalgae without temperature control: The effect of diel temperature fluctuations on microalgal physiology. Journal of Biotechnology 336, 56–63. https://doi.org/10.1016/j.jbiotec.2021.06.017

559 Bu, Y., Liu, C., Pan, Z., Li, W., Wei, Y., Cai, W., Huang, Z., Ma, B., 2024. An attempt to augment
560 performance of machine learning models in a pilot-scale urban wastewater treatment system.
561 Journal of Water Process Engineering 63, 105410.
562 https://doi.org/10.1016/j.jwpe.2024.105410
563 Cairone, S., Hasan, S.W., Choo, K.-H., Lekkas, D.F., Fortunato, L., Zorpas, A.A., Korshin, G., Zarra,
564 T., Belgiorno, V., Naddeo, V., 2024a. Revolutionizing wastewater treatment toward circular
565 economy and carbon neutrality goals: Pioneering sustainable and efficient solutions for
566 automation and advanced process control with smart and cutting-edge technologies. Journal
567 of Water Process Engineering 63, 105486. https://doi.org/10.1016/j.jwpe.2024.105486
568 Cairone, S., Hasan, S.W., Choo, K.-H., Li, C.-W., Zarra, T., Belgiorno, V., Naddeo, V., 2024b.
569 Integrating artificial intelligence modeling and membrane technologies for advanced
570 wastewater treatment: Research progress and future perspectives. Science of The Total
571 Environment 944, 173999. https://doi.org/10.1016/j.scitotenv.2024.173999
572 Cairone, S., Hasan, S.W., Choo, K.-H., Li, C.-W., Zorpas, A.A., Ksibi, M., Zarra, T., Belgiorno, V.,
573 Naddeo, V., 2024c. Enhancing membrane fouling control in wastewater treatment processes
574 through artificial intelligence modeling: research progress and future perspectives. Euro-
575 Mediterr J Environ Integr. https://doi.org/10.1007/s41207-024-00659-0
576 Cairone, S., Hegab, H.M., Khalil, H., Nassar, L., Wadi, V.S., Naddeo, V., Hasan, S.W., 2024d. Novel
577 eco-friendly polylactic acid nanocomposite integrated membrane system for sustainable
578 wastewater treatment: Performance evaluation and antifouling analysis. Science of The Total
579 Environment 912, 168715. https://doi.org/10.1016/j.scitotenv.2023.168715
580 Cairone, S., Mineo, A., Pollice, A., Belgiorno, V., Mannina, G., Naddeo, V., 2024e. Innovative
581 Membrane Bioreactors for Advanced and Sustainable Wastewater Treatment, in: Mannina, G.,
582 Ng, H.Y. (Eds.), Frontiers in Membrane Technology. Springer Nature Switzerland, Cham, pp.
583 120–126. https://doi.org/10.1007/978-3-031-63357-7_20

584 Cheng, Y., Bi, X., Xu, Y., Liu, Y., Li, J., Du, G., Lv, X., Liu, L., 2023. Artificial intelligence
585 technologies in bioprocess: Opportunities and challenges. Bioresource Technology 369,
586 128451. https://doi.org/10.1016/j.biortech.2022.128451
587 Cifuentes-Cabezas, M., Bohórquez-Zurita, J.L., Gil-Herrero, S., Vincent-Vela, M.C., Mendoza-Roca,
588 J.A., Álvarez-Blanco, S., 2023. Deep Study on Fouling Modelling of Ultrafiltration
589 Membranes Used for OMW Treatment: Comparison Between Semi-empirical Models,
590 Response Surface, and Artificial Neural Networks. Food Bioprocess Technol 16, 2126–2146.
591 https://doi.org/10.1007/s11947-023-03033-0
592 Corpuz, M.V.A., Borea, L., Senatore, V., Castrogiovanni, F., Buonerba, A., Oliva, G., Ballesteros, F.,
593 Zarra, T., Belgiorno, V., Choo, K.-H., Hasan, S.W., Naddeo, V., 2021. Wastewater treatment
594 and fouling control in an electro algae-activated sludge membrane bioreactor. Science of The
595 Total Environment 786, 147475. https://doi.org/10.1016/j.scitotenv.2021.147475
596 Corpuz, M.V.A., Cairone, S., Natale, M., Giannattasio, A., Iuliano, V., Grassi, A., Pollice, A.,
597 Mannina, G., Buonerba, A., Belgiorno, V., Naddeo, V., 2024. Sustainable control of
598 microplastics in wastewater using the electrochemically enhanced living membrane
599 bioreactor. Journal of Environmental Management 370, 122649.
600 https://doi.org/10.1016/j.jenvman.2024.122649
601 De Paola, S., Mottola, S., Oliva, G., Naddeo, V., De Marco, I., 2024. Enhanced Mitigation of VOCs,
602 CO2 and Odour Emissions from Organic Fraction of Solid Waste Using Innovative
603 Biodegradable Patches Generated by Supercritical Carbon Dioxide Impregnation. Case
604 Studies in Chemical and Environmental Engineering 100866.
605 https://doi.org/10.1016/j.cscee.2024.100866
606 Foladori, P., Petrini, S., Andreottola, G., 2020. How suspended solids concentration affects
607 nitrification rate in microalgal-bacterial photobioreactors without external aeration. Heliyon
608 6, e03088. https://doi.org/10.1016/j.heliyon.2019.e03088

609 Fu, J., Huang, Y., Liao, Q., Xia, A., Fu, Q., Zhu, X., 2019. Photo-bioreactor design for microalgae: A
610 review from the aspect of CO2 transfer and conversion. Bioresource Technology 292, 121947.
611 https://doi.org/10.1016/j.biortech.2019.121947
612 Guedes, V.C., Palma, G.M., Horta, A.C.L., 2023. An evaluation of light wavelengths, intensity and
613 control for the production of microalgae in photobioreactors: a review. Braz. J. Chem. Eng.
614 https://doi.org/10.1007/s43153-023-00388-x
615 Guven, H., Ersahin, M.E., Ozgun, H., 2022. Chapter 7 - Energy self-sufficiency in wastewater
616 treatment plants: perspectives, challenges, and opportunities, in: An, A., Tyagi, V., Kumar, M.,
617 Cetecioglu, Z. (Eds.), Clean Energy and Resource Recovery. Elsevier, pp. 105–122.
618 https://doi.org/10.1016/B978-0-323-90178-9.00019-6
619 Hamad, N., Nile, B.K., Algretaweerd, H., 2024. Modeling of H2S emission in sulfate-rich sewage lift
620 station by TOXCHEM model. AIP Conference Proceedings 3009, 030115.
621 https://doi.org/10.1063/5.0191749
622 He, X., Li, Z., Xing, C., Li, Y., Liu, M., Gao, X., Ding, Y., Lu, L., Liu, C., Li, C., Wang, D., 2023.
623 Carbon footprint of a conventional wastewater treatment plant: An analysis of water-energy
624 nexus from life cycle perspective for emission reduction. Journal of Cleaner Production 429,
625 139562. https://doi.org/10.1016/j.jclepro.2023.139562
626 He, Y., Li, Y., Li, X., Liu, Yingrui, Wang, Y., Guo, H., Hou, J., Zhu, T., Liu, Yiwen, 2023. Net-zero
627 greenhouse gas emission from wastewater treatment: Mechanisms, opportunities and
628 perspectives. Renewable and Sustainable Energy Reviews 184, 113547.
629 https://doi.org/10.1016/j.rser.2023.113547
630 Holzinger, A., Saranti, A., Molnar, C., Biecek, P., Samek, W., 2022. Explainable AI Methods - A Brief
631 Overview, in: Holzinger, A., Goebel, R., Fong, R., Moon, T., Müller, K.-R., Samek, W. (Eds.),
632 xxAI - Beyond Explainable AI: International Workshop, Held in Conjunction with ICML
633 2020, July 18, 2020, Vienna, Austria, Revised and Extended Papers, Lecture Notes in

634 Computer Science. Springer International Publishing, Cham, pp. 13–38.
635 https://doi.org/10.1007/978-3-031-04083-2_2
636 Hosseinzadeh, A., Zhou, J.L., Altaee, A., Baziar, M., Li, X., 2020. Modeling water flux in osmotic
637 membrane bioreactor by adaptive network-based fuzzy inference system and artificial neural
638 network. Bioresource Technology 310, 123391.
639 https://doi.org/10.1016/j.biortech.2020.123391
640 José Carlos De, M.-J., Patrícia Camara De, B., Ribeiro, R.P., Kligerman, D.C., Oliveira, J.L.M., 2024.
641 Nitrogen removal and nitrous oxide emission from moving bed biofilm reactor (MBBR) under
642 different loads and airflow. Journal of Environmental Chemical Engineering 12, 112574.
643 https://doi.org/10.1016/j.jece.2024.112574
644 Kamali, M., Appels, L., Yu, X., Aminabhavi, T.M., Dewil, R., 2021. Artificial intelligence as a
645 sustainable tool in wastewater treatment using membrane bioreactors. Chemical Engineering
646 Journal 417, 128070. https://doi.org/10.1016/j.cej.2020.128070
647 Klatt, J.M., Polerecky, L., 2015. Assessment of the stoichiometry and efficiency of CO2 fixation
648 coupled to reduced sulfur oxidation. Front. Microbiol. 6.
649 https://doi.org/10.3389/fmicb.2015.00484
650 Kovacs, D.J., Li, Z., Baetz, B.W., Hong, Y., Donnaz, S., Zhao, X., Zhou, P., Ding, H., Dong, Q., 2022.
651 Membrane fouling prediction and uncertainty analysis using machine learning: A wastewater
652 treatment plant case study. Journal of Membrane Science 660, 120817.
653 https://doi.org/10.1016/j.memsci.2022.120817
654 Lancioni, N., Szelag, B., Sgroi, M., Barbusiński, K., Fatone, F., Eusebi, A.L., 2024. Novel extended
655 hybrid tool for real time control and practically support decisions to reduce GHG emissions
656 in full scale wastewater treatment plants. Journal of Environmental Management 365, 121502.
657 https://doi.org/10.1016/j.jenvman.2024.121502

Lebrero, R., Estrada, J.M., Muñoz, R., Quijano, G., 2014. Deterioration of organic packing materials commonly used in air biofiltration: Effect of VOC-packing interactions. Journal of Environmental Management 137, 93–100. https://doi.org/10.1016/j.jenvman.2013.11.052

Li, Z., Dai, J., Chen, H., Lin, B., 2019. An ANN-based fast building energy consumption prediction method for complex architectural form at the early design stage. Build. Simul. 12, 665–681. https://doi.org/10.1007/s12273-019-0538-0

Lu, L., Guest, J.S., Peters, C.A., Zhu, X., Rau, G.H., Ren, Z.J., 2018. Wastewater treatment for carbon capture and utilization. Nat Sustain 1, 750–758. https://doi.org/10.1038/s41893-018-0187-9

Maharana, K., Mondal, S., Nemade, B., 2022. A review: Data pre-processing and data augmentation techniques. Global Transitions Proceedings, International Conference on Intelligent Engineering Approach(ICIEA-2022) 3, 91–99. https://doi.org/10.1016/j.gltp.2022.04.020

Majnooni, S., Fooladi, M., Nikoo, M.R., Al-Rawas, G., Haghighi, A.T., Nazari, R., Al-Wardy, M., Gandomi, A.H., 2024. Smarter water quality monitoring in reservoirs using interpretable deep learning models and feature importance analysis. Journal of Water Process Engineering 60, 105187. https://doi.org/10.1016/j.jwpe.2024.105187

Mannina, G., Butler, D., Benedetti, L., Deletic, A., Fowdar, H., Fu, G., Kleidorfer, M., McCarthy, D., Steen Mikkelsen, P., Rauch, W., Sweetapple, C., Vezzaro, L., Yuan, Z., Willems, P., 2018. Greenhouse gas emissions from integrated urban drainage systems: Where do we stand? Journal of Hydrology 559, 307–314. https://doi.org/10.1016/j.jhydrol.2018.02.058

Maselli, G., Oliva, G., Nesticò, A., Belgiorno, V., Naddeo, V., Zarra, T., 2024. Carbon capture and utilisation (CCU) solutions: Assessing environmental, economic, and social impacts using a new integrated methodology. Science of The Total Environment 948, 174873. https://doi.org/10.1016/j.scitotenv.2024.174873

Meier, L., Stará, D., Bartacek, J., Jeison, D., 2018. Removal of H2S by a continuous microalgae-based photosynthetic biogas upgrading process. Process Safety and Environmental Protection 119, 65–68. https://doi.org/10.1016/j.psep.2018.07.014

684 Morse, J.W., Millero, F.J., Cornwell, J.C., Rickard, D., 1987. The chemistry of the hydrogen sulfide
685 and iron sulfide systems in natural waters. Earth-Science Reviews 24, 1–42.
686 https://doi.org/10.1016/0012-8252(87)90046-8
687 Nam, K., Heo, S., Kim, S., Yoo, C., 2023. A multi-agent AI reinforcement-based digital multi-solution
688 for optimal operation of a full-scale wastewater treatment plant under various influent
689 conditions. Journal of Water Process Engineering 52, 103533.
690 https://doi.org/10.1016/j.jwpe.2023.103533
691 Nam, K., Heo, S., Rhee, G., Kim, M., Yoo, C., 2021. Dual-objective optimization for energy-saving
692 and fouling mitigation in MBR plants using AI-based influent prediction and an integrated
693 biological-physical model. Journal of Membrane Science 626, 119208.
694 https://doi.org/10.1016/j.memsci.2021.119208
695 Newhart, K.B., Holloway, R.W., Hering, A.S., Cath, T.Y., 2019. Data-driven performance analyses of
696 wastewater treatment plants: A review. Water Research 157, 498–513.
697 https://doi.org/10.1016/j.watres.2019.03.030
698 Niu, C., Li, X., Dai, R., Wang, Z., 2022. Artificial intelligence-incorporated membrane fouling
699 prediction for membrane-based processes in the past 20 years: A critical review. Water
700 Research 216, 118299. https://doi.org/10.1016/j.watres.2022.118299
701 Nourani, V., Asghari, P., Sharghi, E., 2021. Artificial intelligence based ensemble modeling of
702 wastewater treatment plant using jittered data. Journal of Cleaner Production 291, 125772.
703 https://doi.org/10.1016/j.jclepro.2020.125772
704 Oliva, G., Buonerba, A., Grassi, A., Hasan, S.W., Korshin, G.V., Zorpas, A.A., Belgiorno, V., Naddeo,
705 V., Zarra, T., 2024. Microalgae to biodiesel: A novel green conversion method for high-quality
706 lipids recovery and in-situ transesterification to fatty acid methyl esters. Journal of
707 Environmental Management 357, 120830. https://doi.org/10.1016/j.jenvman.2024.120830
708 Oliva, G., Pahunang, R.R., Vigliotta, G., Zarra, T., Ballesteros, F.C., Mariniello, A., Buonerba, A.,
709 Belgiorno, V., Naddeo, V., 2023. Advanced treatment of toluene emissions with a cutting-edge

algal bacterial photo-bioreactor: Performance assessment in a circular economy perspective. Science of The Total Environment 878, 163005. https://doi.org/10.1016/j.scitotenv.2023.163005

Omri, I., Bouallagui, H., Aouidi, F., Godon, J.-J., Hamdi, M., 2011. H2S gas biological removal efficiency and bacterial community diversity in biofilter treating wastewater odor. Bioresource Technology 102, 10202–10209. https://doi.org/10.1016/j.biortech.2011.05.094

Onay, A., Onay, M., 2024. Enhancing the content of phycoerythrin through the application of microplastics from *Porphyridium* cruentum produced in wastewater using machine learning methods. Journal of Environmental Management 371, 123266. https://doi.org/10.1016/j.jenvman.2024.123266

Pahunang, R.R., Buonerba, A., Senatore, V., Oliva, G., Ouda, M., Zarra, T., Muñoz, R., Puig, S., Ballesteros, F.C., Li, C.-W., Hasan, S.W., Belgiorno, V., Naddeo, V., 2021. Advances in technological control of greenhouse gas emissions from wastewater in the context of circular economy. Science of The Total Environment 792, 148479. https://doi.org/10.1016/j.scitotenv.2021.148479

Pasquarelli, F., Oliva, G., Mariniello, A., Buonerba, A., Li, C.-W., Belgiorno, V., Naddeo, V., Zarra, T., 2024a. Carbon neutrality in wastewater treatment plants: An integrated biotechnological-based solution for nutrients recovery, odour abatement and CO2 conversion in alternative energy drivers. Chemosphere 354, 141700. https://doi.org/10.1016/j.chemosphere.2024.141700

Pasquarelli, F., Oliva, G., Mariniello, A., Buonerba, A., Zorpas, A.A., Ng, H.Y., Belgiorno, V., Naddeo, V., Zarra, T., 2024b. Integration of Moving Bed Biofilm Reactor (MBBR) and algal PhotoBioReactors (aPBR) for achieving carbon neutrality in wastewater treatment. Science of The Total Environment 955, 177012. https://doi.org/10.1016/j.scitotenv.2024.177012

Pudi, A., Rezaei, M., Signorini, V., Andersson, M.P., Baschetti, M.G., Mansouri, S.S., 2022. Hydrogen sulfide capture and removal technologies: A comprehensive review of recent

developments and emerging trends. Separation and Purification Technology 298, 121448. https://doi.org/10.1016/j.seppur.2022.121448

Qadir, M., Drechsel, P., Jiménez Cisneros, B., Kim, Y., Pramanik, A., Mehta, P., Olaniyan, O., 2020. Global and regional potential of wastewater as a water, nutrient and energy source. Natural Resources Forum 44, 40–51. https://doi.org/10.1111/1477-8947.12187

Rani, A., Snyder, S.W., Kim, H., Lei, Z., Pan, S.-Y., 2022. Pathways to a net-zero-carbon water sector through energy-extracting wastewater technologies. npj Clean Water 5, 1–17. https://doi.org/10.1038/s41545-022-00197-8

Reddy, S., Nadgouda, S.G., Tong, A., Fan, L.-S., 2019. Metal sulfide-based process analysis for hydrogen generation from hydrogen sulfide conversion. International Journal of Hydrogen Energy 44, 21336–21350. https://doi.org/10.1016/j.ijhydene.2019.06.180

Rios Fuck, J.V., Cechinel, M.A.P., Neves, J., Campos de Andrade, R., Tristão, R., Spogis, N., Riella, H.G., Soares, C., Padoin, N., 2024. Predicting effluent quality parameters for wastewater treatment plant: A machine learning-based methodology. Chemosphere 352, 141472. https://doi.org/10.1016/j.chemosphere.2024.141472

Sadoune, H., Rihani, R., Marra, F.S., 2023. DNN model development of biogas production from an anaerobic wastewater treatment plant using Bayesian hyperparameter optimization. Chemical Engineering Journal 471, 144671. https://doi.org/10.1016/j.cej.2023.144671

Santos, J.M., Sá, L.M., Reis Júnior, N.C., Gonçalves, R.F., Siqueira, R.N., 2006. Modelling hydrogen sulphide emission in a WWTP with UASB reactor followed by aerobic biofilters. Water Sci Technol 54, 173–180. https://doi.org/10.2166/wst.2006.861

Senatore, V., Buonerba, A., Zarra, T., Oliva, G., Belgiorno, V., Boguniewicz-Zablocka, J., Naddeo, V., 2021. Innovative membrane photobioreactor for sustainable CO2 capture and utilization. Chemosphere 273, 129682. https://doi.org/10.1016/j.chemosphere.2021.129682

Senatore, V., Oliva, G., Buonerba, A., Zarra, T., Borea, L., Hasan, S.W., Belgiorno, V., Naddeo, V., 2022. An integrated algal membrane photobioreactor as a green-transition technology for the

762 carbon capture and utilization. Journal of Environmental Chemical Engineering 10, 107344.
763 https://doi.org/10.1016/j.jece.2022.107344
764 Serna-Carrizales, J.C., Zárate-Guzmán, A.I., Flores-Ramírez, R., Díaz de León-Martínez, L., Aguilar-
765 Aguilar, A., Warren- Vega, W.M., Bailón-García, E., Ocampo-Pérez, R., 2024. Application of
766 artificial intelligence for the optimization of advanced oxidation processes to improve the
767 water quality polluted with pharmaceutical compounds. Chemosphere 351, 141216.
768 https://doi.org/10.1016/j.chemosphere.2024.141216
769 Soaudy, M.R., Ghonimy, A., Greco, L.S.L., Chen, Z., Dyzenchauz, A., Li, J., 2023. Total suspended
770 solids and their impact in a biofloc system: Current and potentially new management
771 strategies. Aquaculture 572, 739524. https://doi.org/10.1016/j.aquaculture.2023.739524
772 Syed, M., Soreanu, G., Falletta, P., Béland, M., 2006. Removal of hydrogen sulfide from gas streams
773 using biological processes - A review. CANADIAN BIOSYSTEMS ENGINEERING 48.
774 Tian, L., Han, C., Zhang, J., Ouyang, Y., Xi, J., 2021. Development of an H2S emission model for
775 wastewater treatment plants. J Air Waste Manag Assoc 71, 1303–1311.
776 https://doi.org/10.1080/10962247.2021.1939195
777 Tsang, W.K., Benoit, D.F., 2023. Interpretability and Explainability in Machine Learning, in: Ohsawa,
778 Y. (Ed.), Living Beyond Data: Toward Sustainable Value Creation, Intelligent Systems
779 Reference Library. Springer International Publishing, Cham, pp. 89–100.
780 https://doi.org/10.1007/978-3-031-11593-6_5
781 Vergara, C., Muñoz, R., Campos, J.L., Seeger, M., Jeison, D., 2016. Influence of light intensity on
782 bacterial nitrifying activity in algal-bacterial photobioreactors and its implications for
783 microalgae-based wastewater treatment. International Biodeterioration & Biodegradation
784 114, 116–121. https://doi.org/10.1016/j.ibiod.2016.06.006
785 Xu, Y., Goodacre, R., 2018. On Splitting Training and Validation Set: A Comparative Study of Cross-
786 Validation, Bootstrap and Systematic Sampling for Estimating the Generalization

Performance of Supervised Learning. J. Anal. Test. 2, 249–262. https://doi.org/10.1007/s41664-018-0068-2

Yan, C., Zhu, L., Wang, Y., 2016. Photosynthetic CO2 uptake by microalgae for biogas upgrading and simultaneously biogas slurry decontamination by using of microalgae photobioreactor under various light wavelengths, light intensities, and photoperiods. Applied Energy 178, 9–18. https://doi.org/10.1016/j.apenergy.2016.06.012

Yang, Y., Li, X., Zhu, H., Xu, X., Bao, L., 2022. Chemical removal of m-cresol: a critical review. Reviews in Chemical Engineering 38, 1023–1044. https://doi.org/10.1515/revce-2021-0001

Yang, Y., Lin, H., Long, Y., Mei, Y., Chen, J.P., 2024a. Development of catalytic zero-valent iron incorporated PAN catalytic film for efficient degradation of organic matters. npj Clean Water 7, 1–10. https://doi.org/10.1038/s41545-024-00333-6

Yang, Y., Liu, M., You, X., Li, Y., Lin, H., Chen, J.P., 2024b. A novel bimetallic Fe-Cu-CNT catalyst for effective catalytic wet peroxide oxidation: reaction optimization and mechanism investigation. Chemical Engineering Journal 479, 147320. https://doi.org/10.1016/j.cej.2023.147320

Yang, Y., Tang, S., Chen, J.P., 2024c. Carbon capture and utilization by algae with high concentration CO2 or bicarbonate as carbon source. Science of The Total Environment 918, 170325. https://doi.org/10.1016/j.scitotenv.2024.170325

Yang, Y., Tang, S., Lin, H., Fu, H., Mei, Y., Long, Y., 2023. Catalytic Reaction Intensification by a Novel Cryogenic Auxiliary Synthesized Fe-PAN Membrane. Ind. Eng. Chem. Res. 62, 20677–20688. https://doi.org/10.1021/acs.iecr.3c03497

Yu, X.-L., Zhao, Z.-T., Zhao, H.-B., He, S.-S., Cui, C.-H., Sun, H.-J., Zhao, Y.-L., Bai, S.-W., Dong, J., Pang, J.-W., Ding, J., Ren, N.-Q., Yang, S.-S., 2022. Mapping research on carbon neutrality in WWTPs between 2001 and 2021: A scientometric and visualization analysis. Sustainable Horizons 3, 100022. https://doi.org/10.1016/j.horiz.2022.100022

812 Zaghloul, M.S., Achari, G., 2022. Application of machine learning techniques to model a full-scale
813 wastewater treatment plant with biological nutrient removal. Journal of Environmental
814 Chemical Engineering 10, 107430. https://doi.org/10.1016/j.jece.2022.107430
815 Zarra, T., Senatore, V., Zorpas, A.A., Oliva, G., Voukkali, I., Belgiorno, V., Naddeo, V., 2024.
816 Advanced membrane photobioreactors in algal CO2 biofixation and valuable biomass
817 production: Integrative life cycle assessment and sustainability analysis. Sustainable
818 Chemistry and Pharmacy 40, 101658. https://doi.org/10.1016/j.scp.2024.101658
819 Zhang, S., Jin, Y., Chen, W., Wang, J., Wang, Y., Ren, H., 2023. Artificial intelligence in wastewater
820 treatment: A data-driven analysis of status and trends. Chemosphere 139163.
821 https://doi.org/10.1016/j.chemosphere.2023.139163
822 Zhang, Y., Gao, X., Smith, K., Inial, G., Liu, S., Conil, L.B., Pan, B., 2019. Integrating water quality
823 and operation into prediction of water production in drinking water treatment plants by genetic
824 algorithm enhanced artificial neural network. Water Research 164, 114888.
825 https://doi.org/10.1016/j.watres.2019.114888
826 Zhu, J.-J., Yang, M., Ren, Z.J., 2023. Machine Learning in Environmental Research: Common Pitfalls
827 and Best Practices. Environ. Sci. Technol. 57, 17671–17689.
828 https://doi.org/10.1021/acs.est.3c00026
829 Zwain, H.M., Nile, B.K., Faris, A.M., Vakili, M., Dahlan, I., 2020. Modelling of hydrogen sulfide
830 fate and emissions in extended aeration sewage treatment plant using TOXCHEM
831 simulations. Scientific Reports 10, 22209. https://doi.org/10.1038/s41598-020-79395-8
832

Glossary of AI modeling terms

Activation (or transfer) functions: Functions that determine the activation level of each node and connect neurons within the network by propagating the output of nodes from one layer to the next.

Artificial neural network (ANN): A computational model inspired by the structure and function of the human brain, consisting of interconnected nodes (or neurons) organized into layers.

ANN architecture (or ANN configuration): The structure of an ANN, including the number of layers, the number of neurons per layer, and the types of connections between neurons.

Box constraint (or regularization): Hyperparameter that controls the maximum penalty on margin-violating observations, balancing model accuracy and complexity to prevent overfitting.

Classification task: A machine learning task that involves assigning labels (or categories) to input data based on learned patterns.

Coefficient of Determination (R^2): A statistical measure indicating the goodness of fit for a model's predictions compared to actual data, with values closer to 1 representing a better fit.

Cross-validation: A technique for evaluating the performance of a model by training and testing it on new (unseen) data.

Decision tree: An algorithm that uses tree-like structure to make sequential decisions to reach a specific outcome.

Deep learning: A subfield of machine learning that uses deep neural networks (multi-layered ANNs) to capture complex patterns from large datasets.

Feature engineering: The process of selecting, transforming, and creating the most relevant variables from raw data to be used as inputs (features) in a machine learning model.

Gamma: Hyperparameter that determines the influence of each training example on the decision boundary.

Gradient boosting algorithm: An algorithm that builds a sequence of weak learners (typically decision trees), with each learner correcting errors from the previous ones.

Grid search: A hyperparameter tuning technique that tests multiple combinations of parameters to identify the best configuration for a model.

Hyperparameters: Parameters defined before training that govern model learning process.

Hidden layers (ANN): Intermediate layers in an ANN that process data from the input layer before passing it to the output layer.

Input layer (ANN): The first layer in an ANN, composed of input neurons, that brings the initial data to the model.

Input variables: Features or variables in a dataset used by a model to make predictions.

Kernel functions: Mathematical functions used to transform data into higher-dimensional spaces.

Learning rate: A hyperparameter controlling the rate at which an algorithm updates the parameters during training.

Least-squares boosting (LSBoost): A gradient boosting algorithm that builds an ensemble of decision trees to minimize the residual errors from previous trees.

Levenberg-Marquardt algorithm ("trainlm"): A training algorithm used to optimize weights and biases of a model.

Linear kernel: The simplest and most straightforward kernel function that assumes the data are linearly separable.

Machine learning: A branch of AI focused on algorithms that learn patterns from data, enabling them to make predictions or decision based on learned representations.

Mean absolute error (MAE): A metric that calculates the average absolute difference between predicted and actual values.

Mean squared error (MSE): A metric that computes the average of squared differences between predicted and actual values.

Model training: The process of feeding data to a model to learn underlying patterns and relationships.

Output layer (ANN): The final layer in an ANN that produces predictions or outputs based on the learned model.

Output variables: Target variables that a model aims to predict using input data.

Polynomial kernel: A kernel function that transforms data into a higher-dimensional space using polynomial functions. It is particularly useful with non-linear decision boundaries.

Radial basis function (RBF) kernel (or Gaussian kernel): A kernel function used for non-linear separation of data by measuring the similarity between data points based on their distance in the input space.

Root mean squared error (RMSE): The square root of MSE, a metric that measures the average difference between predicted and actual values.

Random forest (RF): An ensemble learning method that combines many decision trees for classification or regression tasks.

Regression task: A machine learning task for predicting values of continuous dependent variables based on independent variables.

Subsample: Hyperparameters that control the fraction of observations used for training each tree.

Supervised learning: machine learning technique that relies on labeled data, meaning that input variables (predictors) are used to predict output variables (responses), whose 'true' values are readily

available in the training data, allowing the model's performance to be evaluated using appropriate metrics.

Support vector machine (SVM): A supervised learning algorithm for classification and regression that separates data into classes by finding the optimal hyperplane or set of hyperplanes.

Support vectors: Data points closest to the hyperplane in an SVM model, influencing its position and orientation.

Sigmoidal transfer function ("tansig"): An "S"-shaped function that introduces nonlinearity into the model.

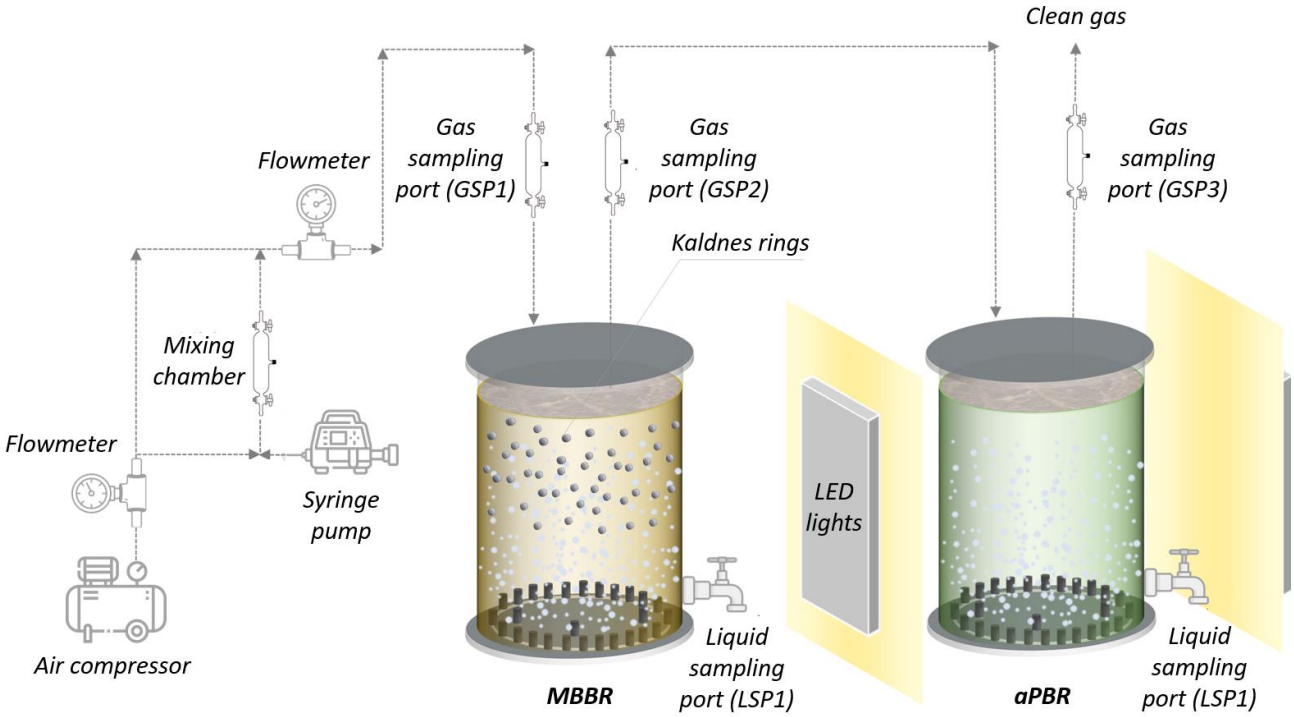
Training algorithm: An algorithm that adjusts model parameters during training.

Training dataset: The subset of data used to train a model, allowing it to learn the patterns/relationships required for a given task.

Testing dataset: The subset of data used to evaluate the performance of a trained model, assessing its ability to generalize to unseen data.

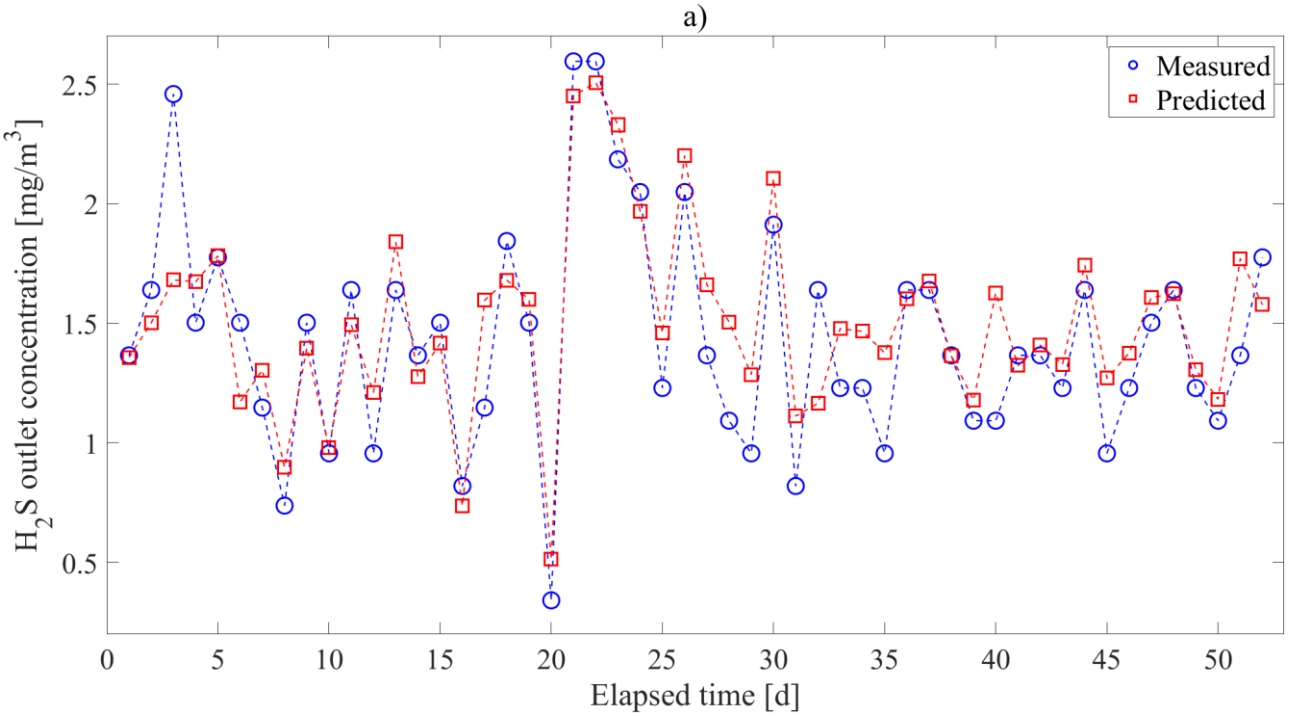
912 **Figures**

913



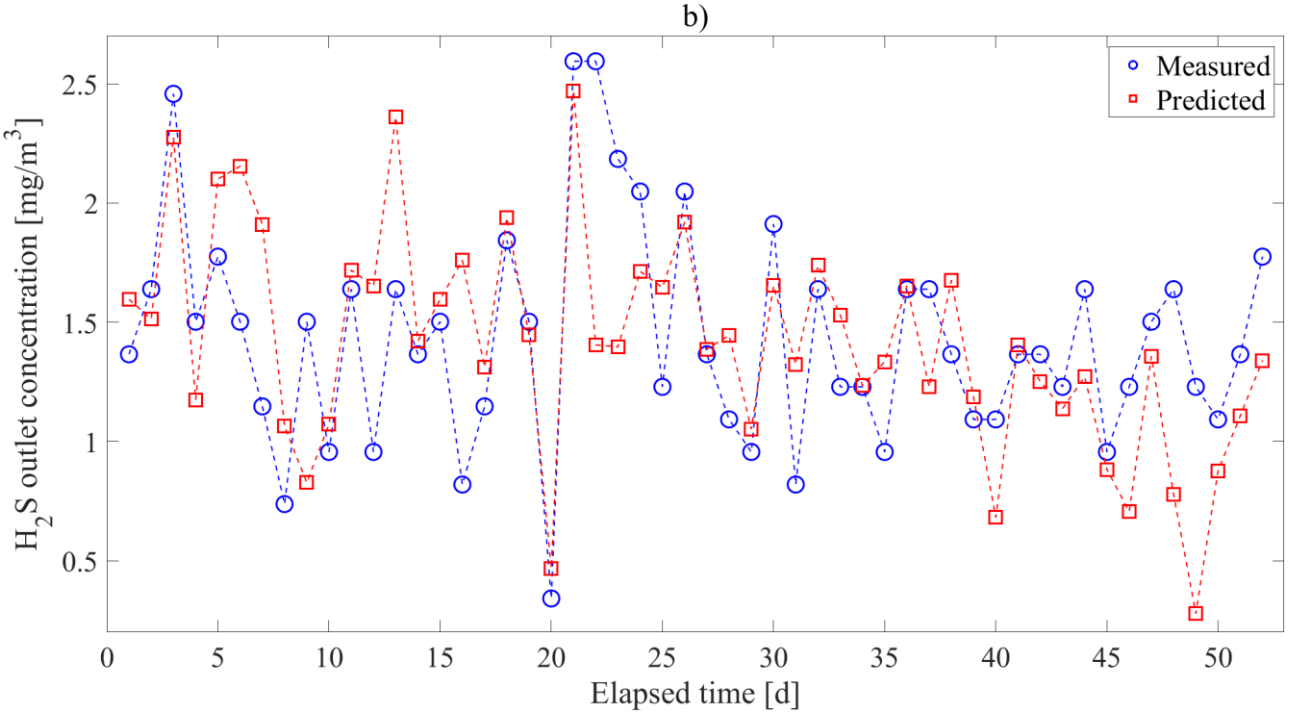
914

915 **Figure 1.** Schematic representation of the MBBR+aPBR integrated system.

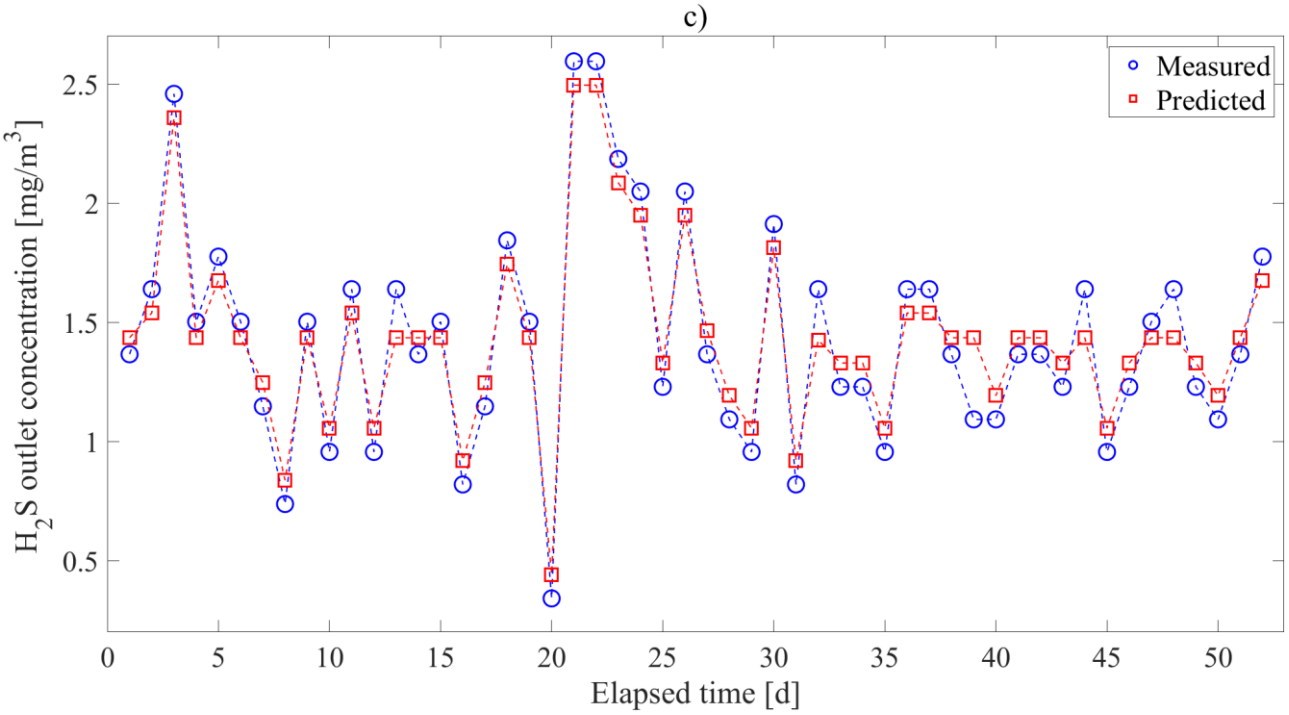

a)
Measured
Predicted
H_2S outlet concentration [mg/m^3]
Elapsed time [d]

916

917

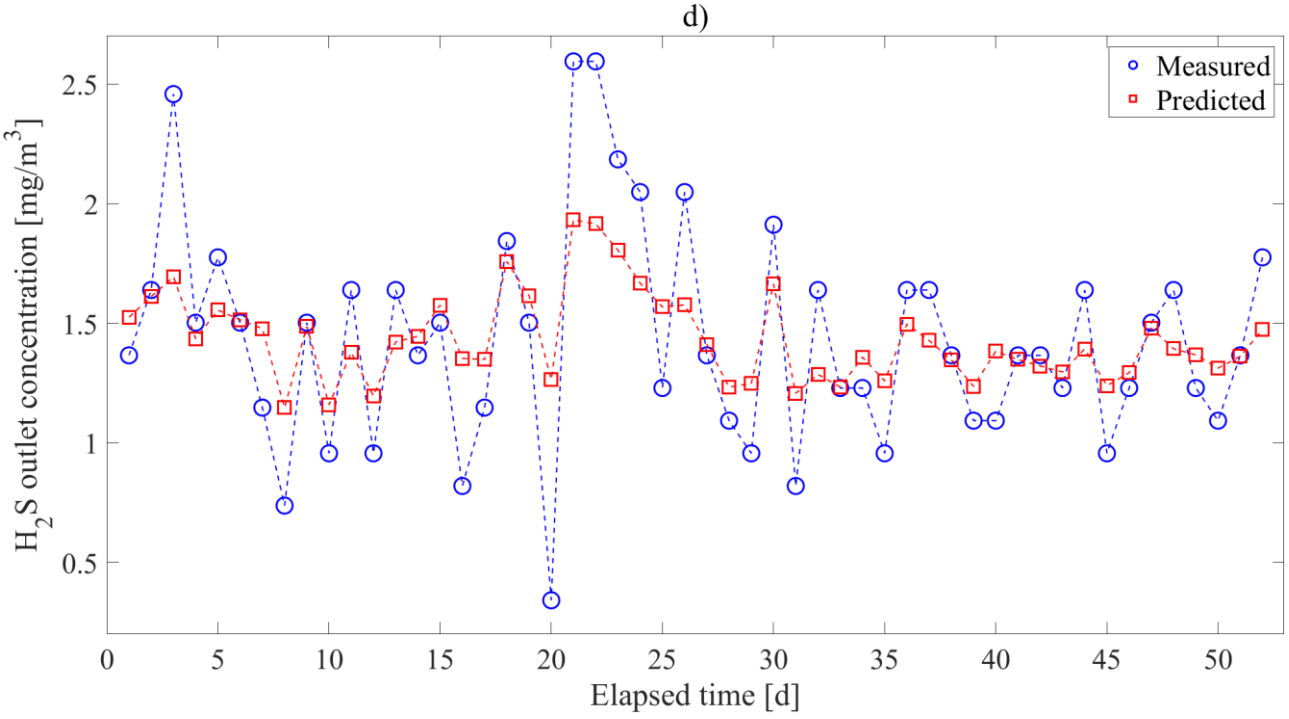

b)
Measured
Predicted
H_2S outlet concentration [mg/m^3]
Elapsed time [d]

918


c)
Measured
Predicted
H_2S outlet concentration [mg/m^3]
Elapsed time [d]

919

920


d)
Measured
Predicted
H_2S outlet concentration [mg/m^3]
Elapsed time [d]

921

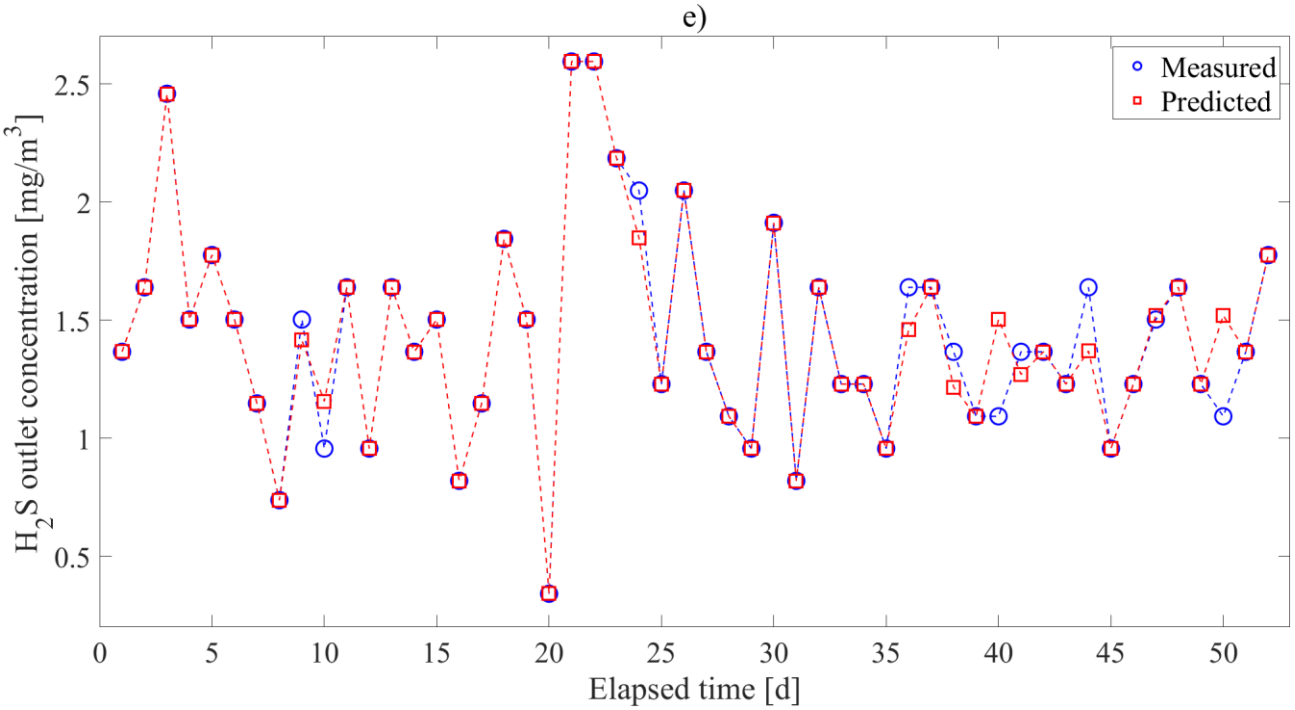
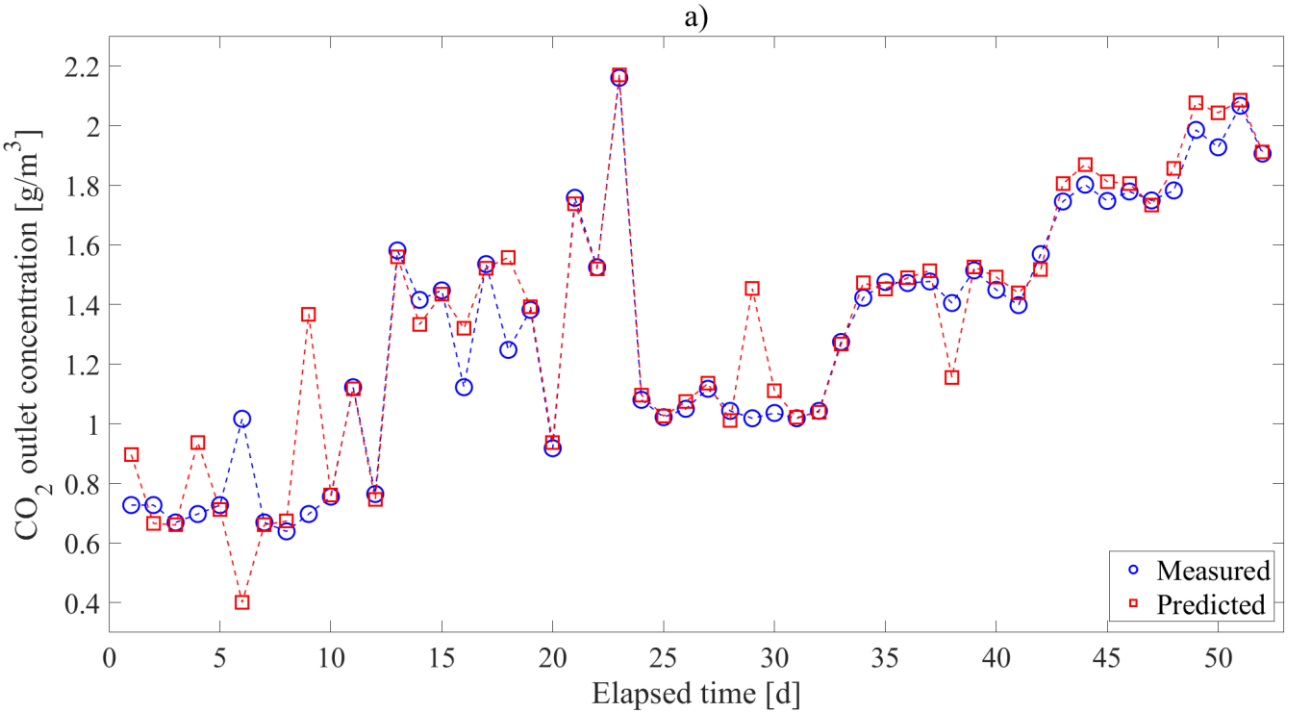
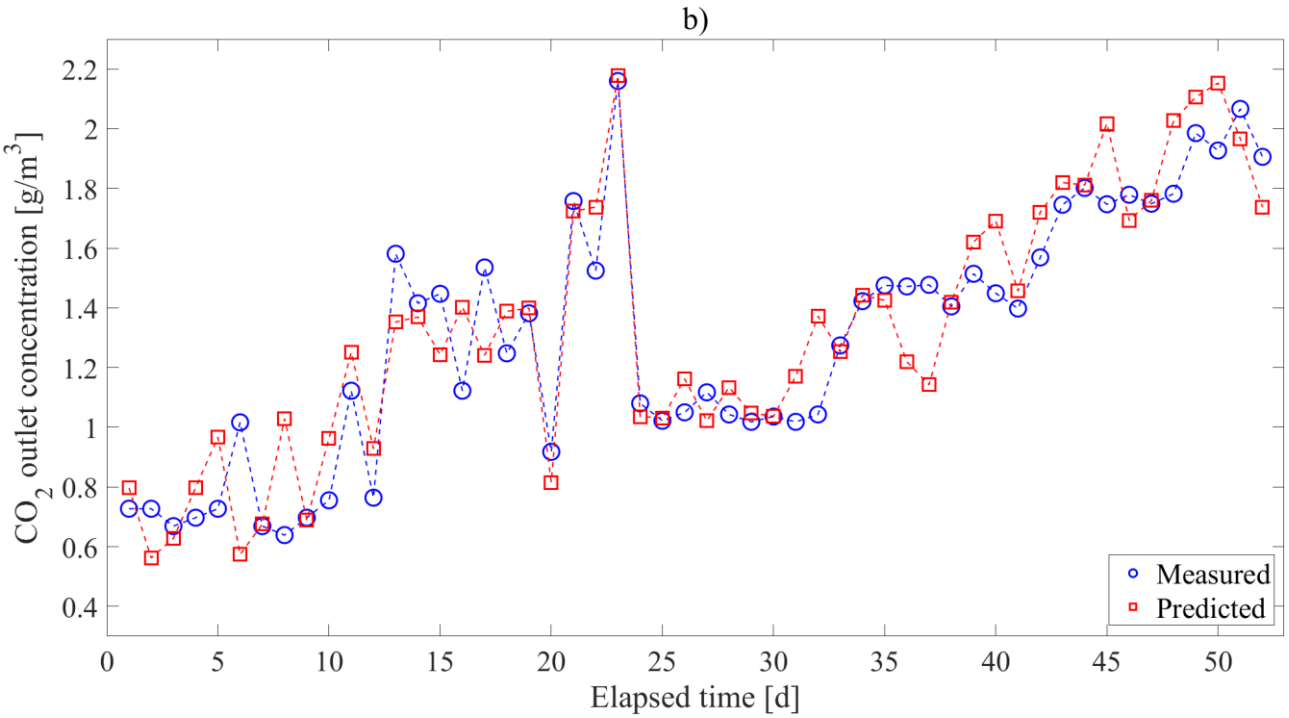


Figure 2. Predictions of H_2S outlet concentrations from the MBBR using (a) ANN, (b) multi-task ANN, (c) SVM, (d) RF, and (e) LSBoost models.

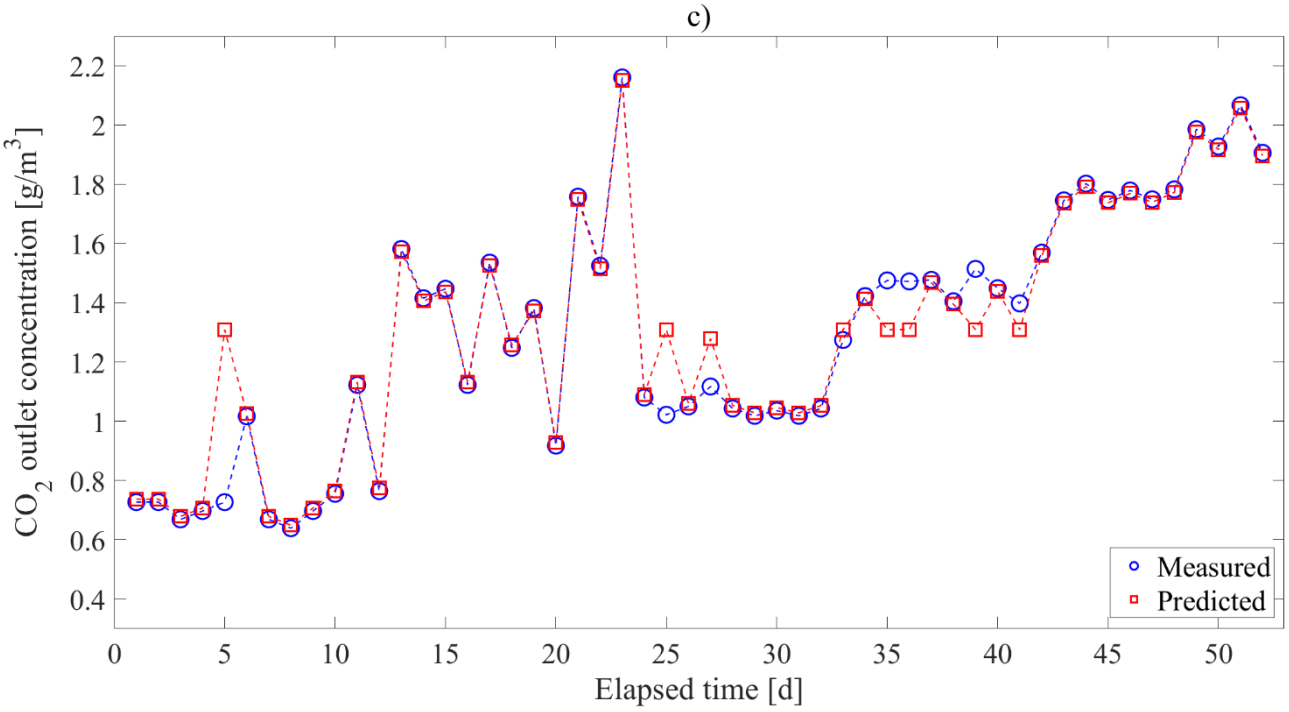

a)
CO_2 outlet concentration [g/m^3]
Elapsed time [d]
Measured
Predicted

926

927

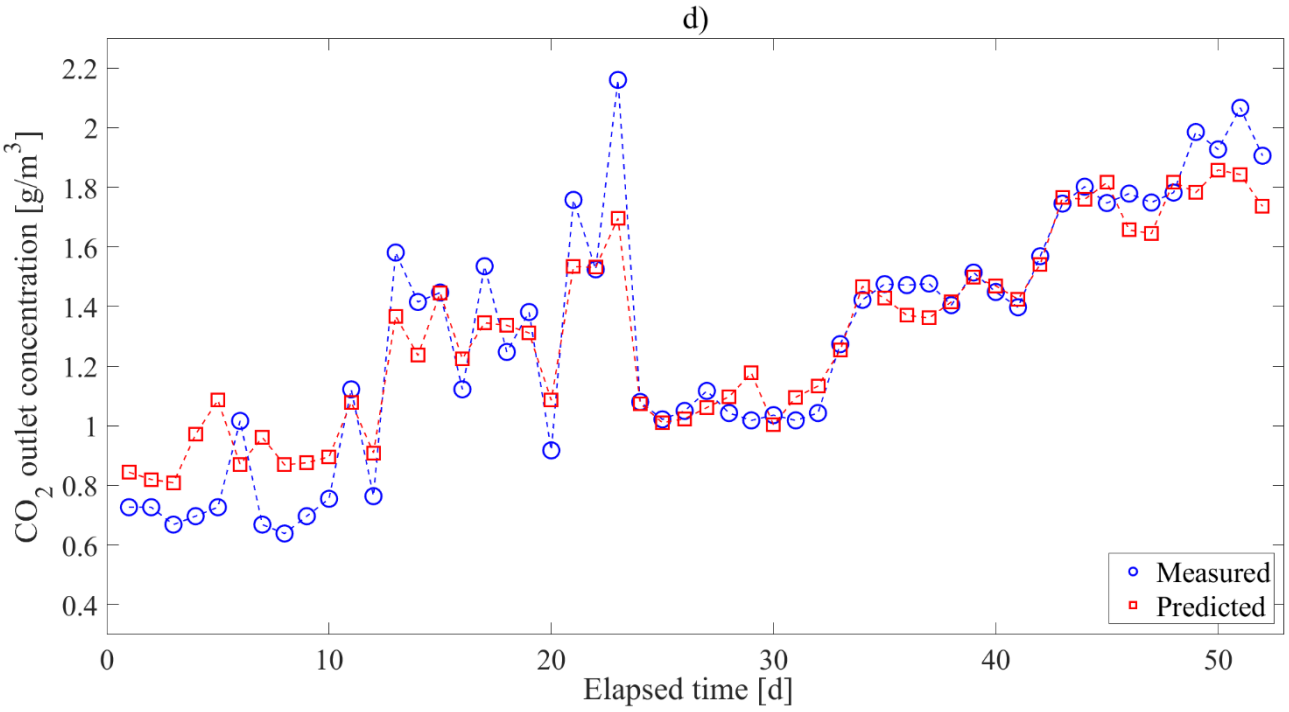

b)
CO_2 outlet concentration [g/m^3]
Elapsed time [d]
Measured
Predicted

928

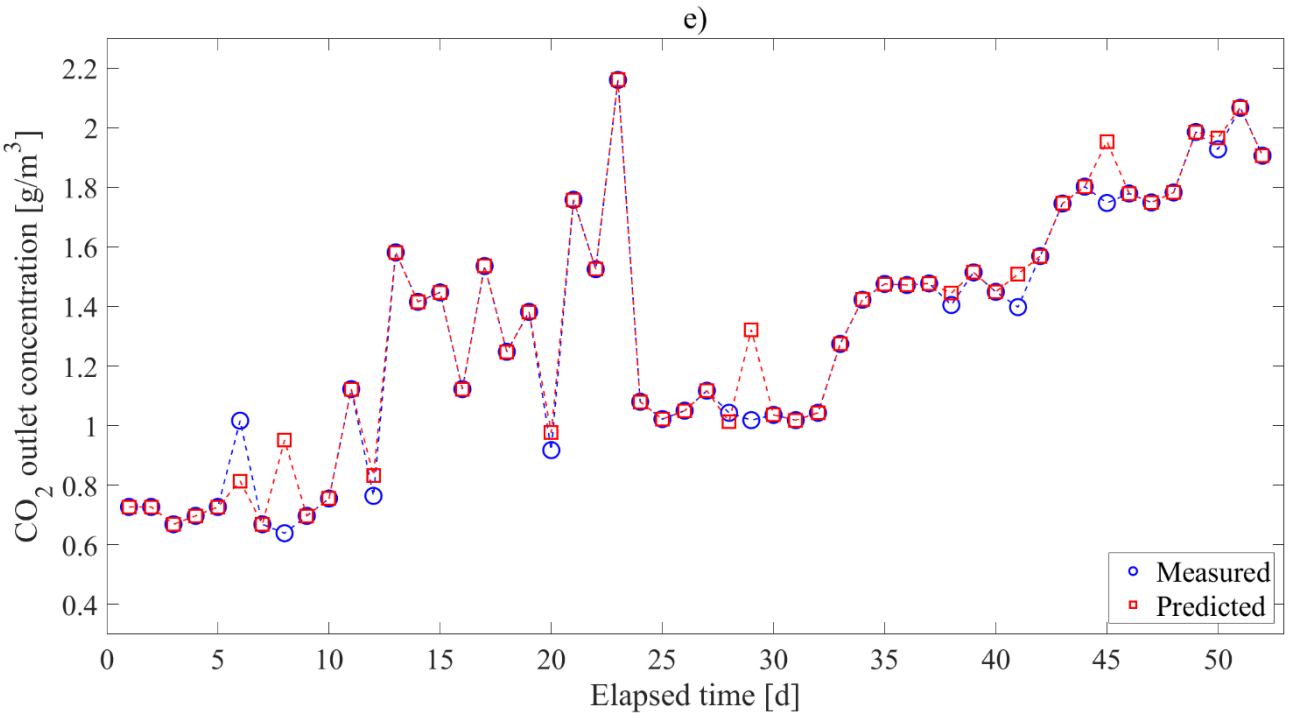

c)
CO_2 outlet concentration [g/m^3]
Elapsed time [d]
Measured
Predicted

929

930


d)
CO_2 outlet concentration [g/m^3]
Elapsed time [d]
Measured
Predicted

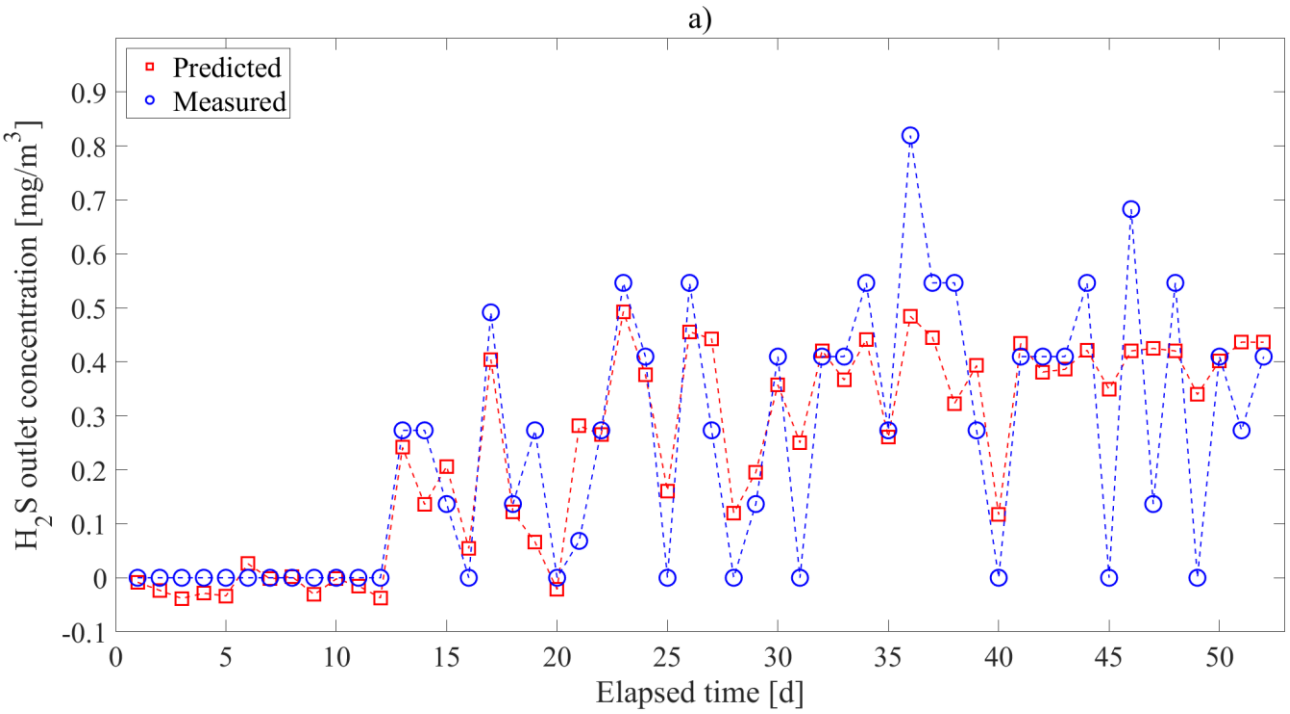
931



932

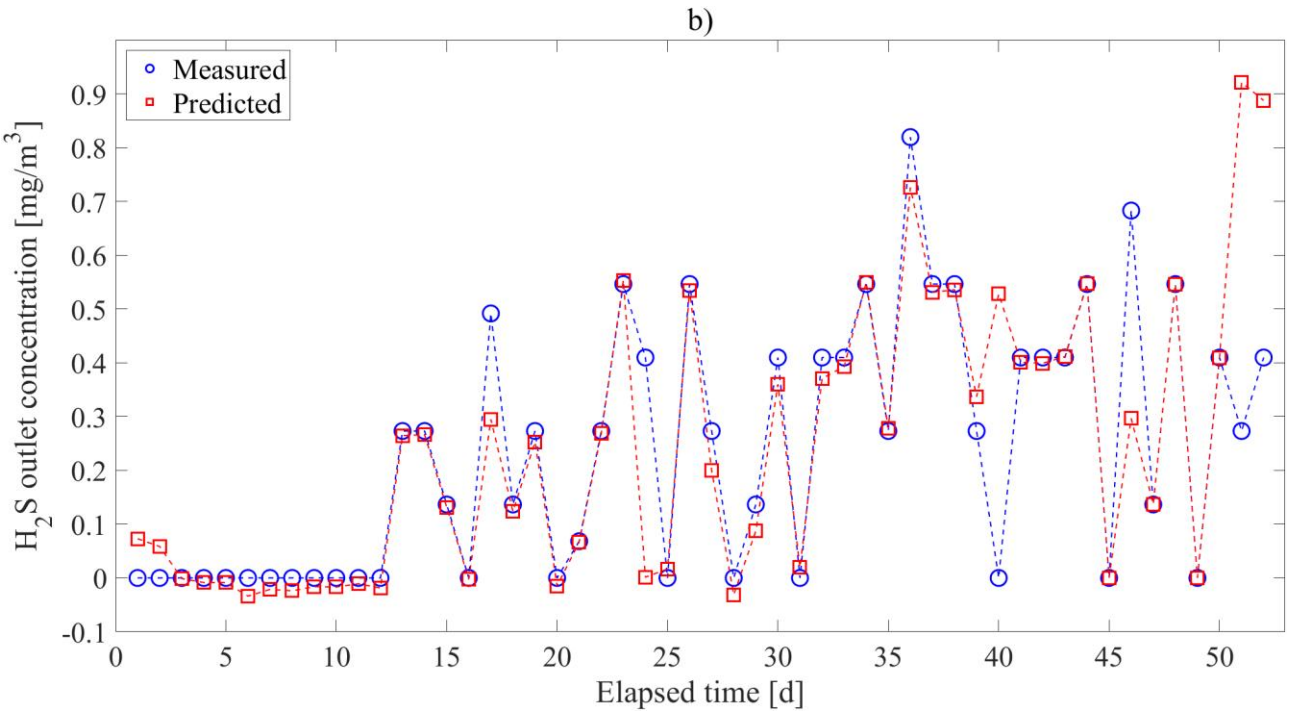
933 **Figure 3.** Predictions of CO_2 outlet concentrations from the MBBR using (a) ANN, (b) multi-task

934 ANN, (c) SVM, (d) RF, and (e) LSBoost models.

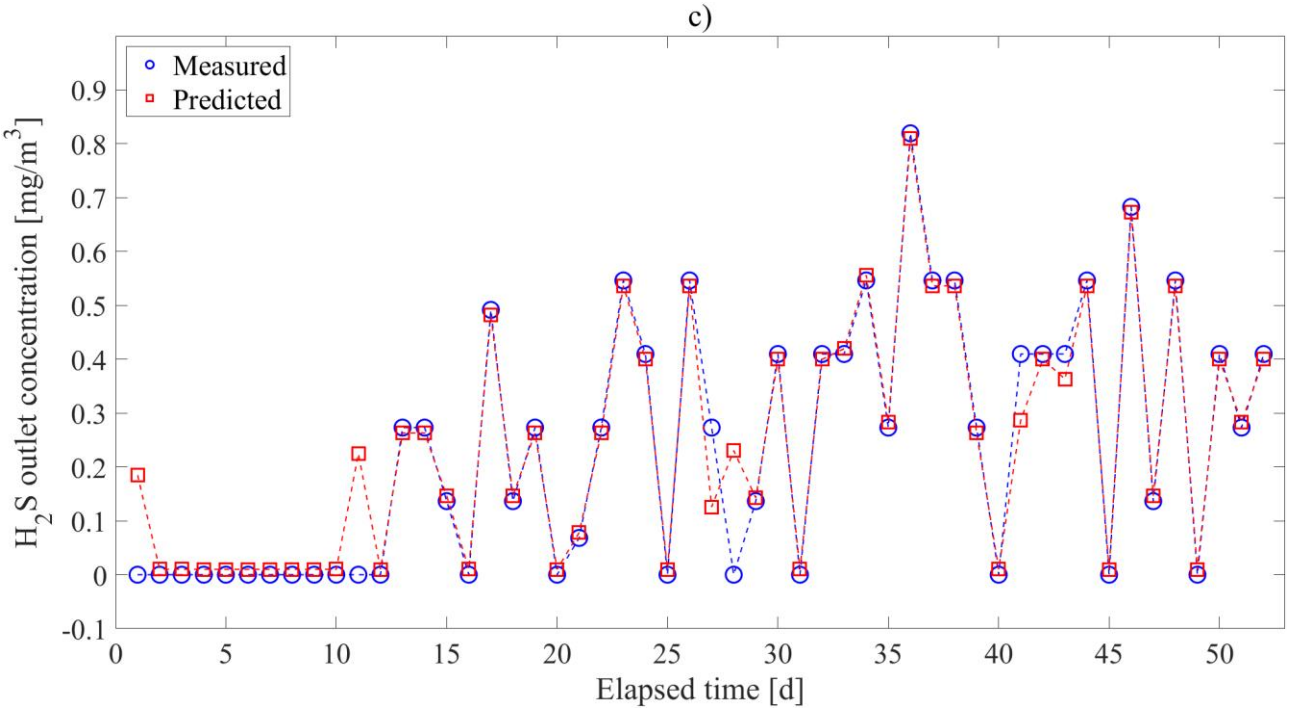

a)
Predicted
Measured
H2S outlet concentration [mg/m3]
Elapsed time [d]

935

936

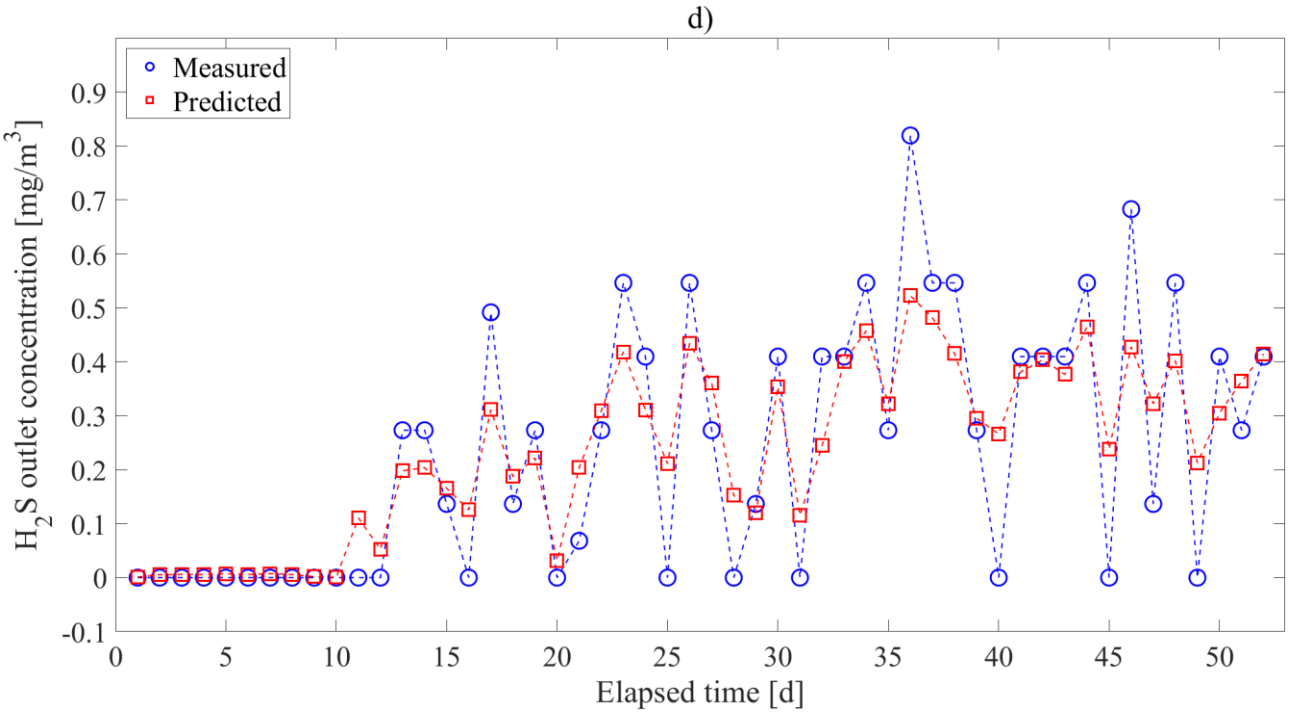

b)
Measured
Predicted
H2S outlet concentration [mg/m3]
Elapsed time [d]

937

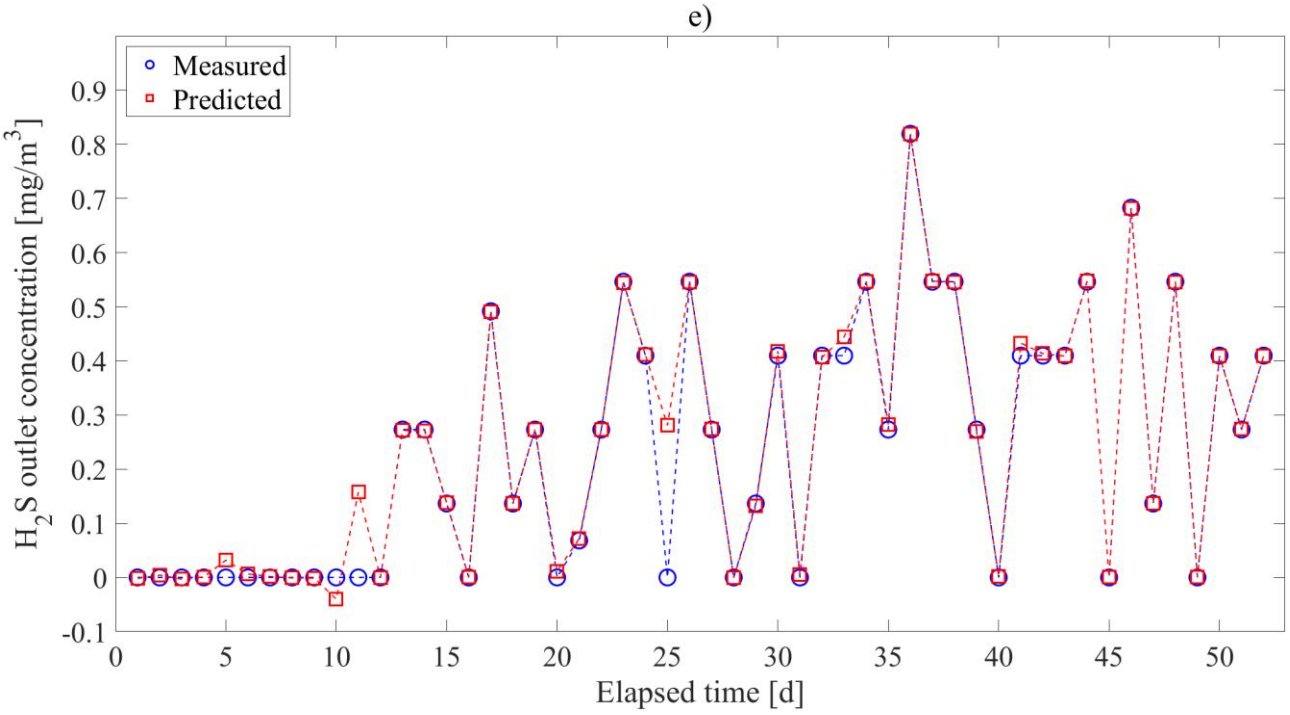

c)
Measured
Predicted
H2S outlet concentration [mg/m3]
Elapsed time [d]

938

939


d)
Measured
Predicted
H2S outlet concentration [mg/m3]
Elapsed time [d]

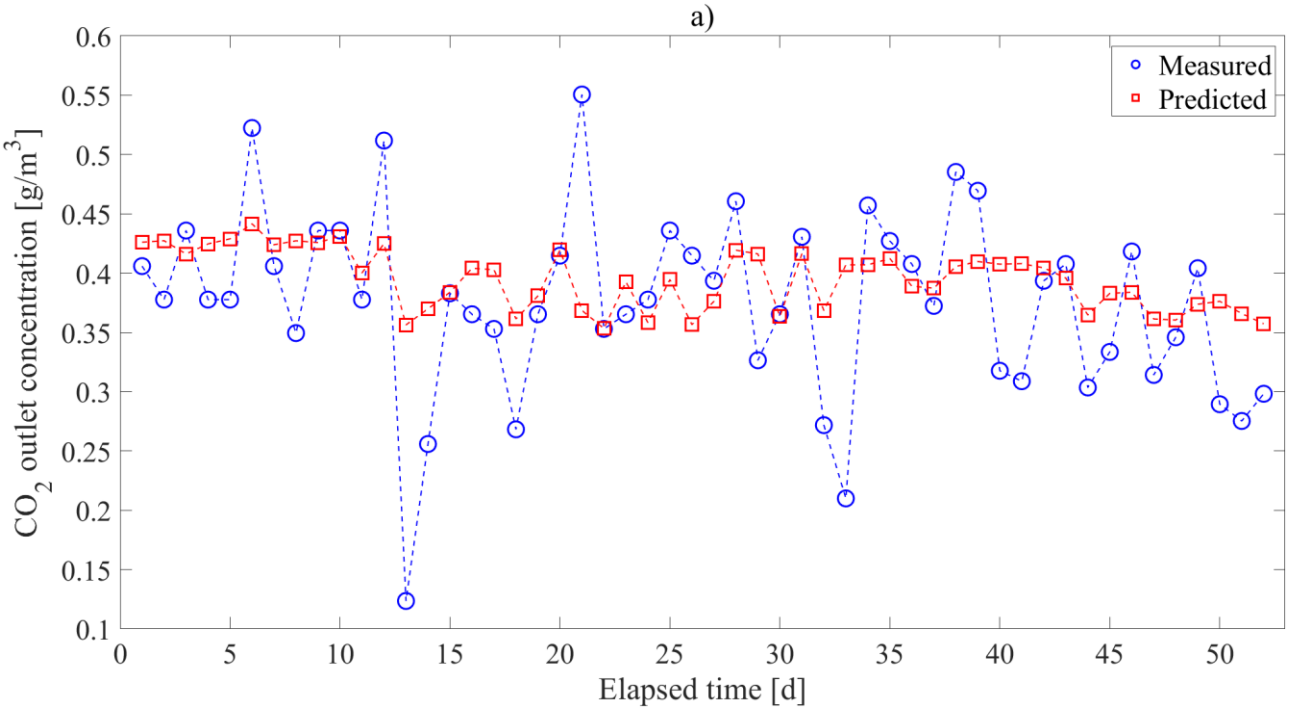
940



941

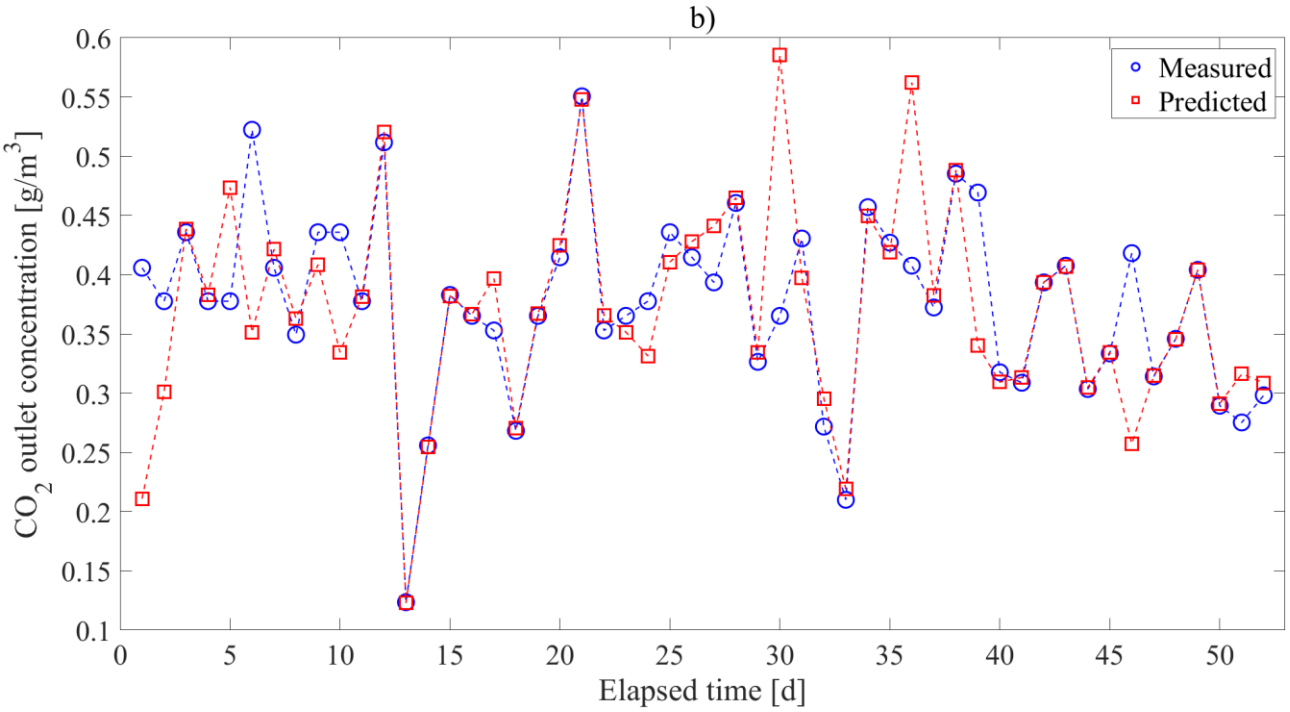
942 **Figure 4.** Predictions of H_2S outlet concentrations from the aPBR using (a) ANN, (b) multi-task

943 ANN, (c) SVM, (d) RF, and (e) LSBoost models.

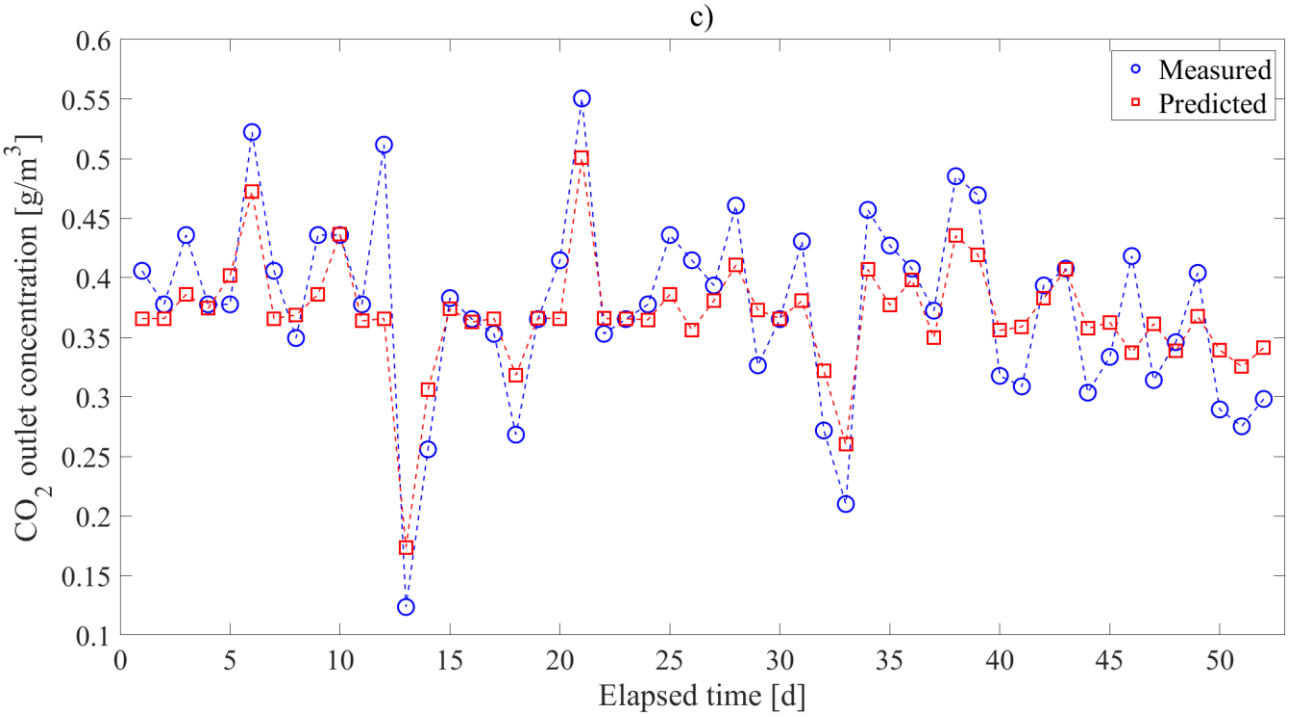

a)
Measured
Predicted
CO_2 outlet concentration [g/m^3]
Elapsed time [d]
0.6
0.55
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0
5
10
15
20
25
30
35
40
45
50

944

945

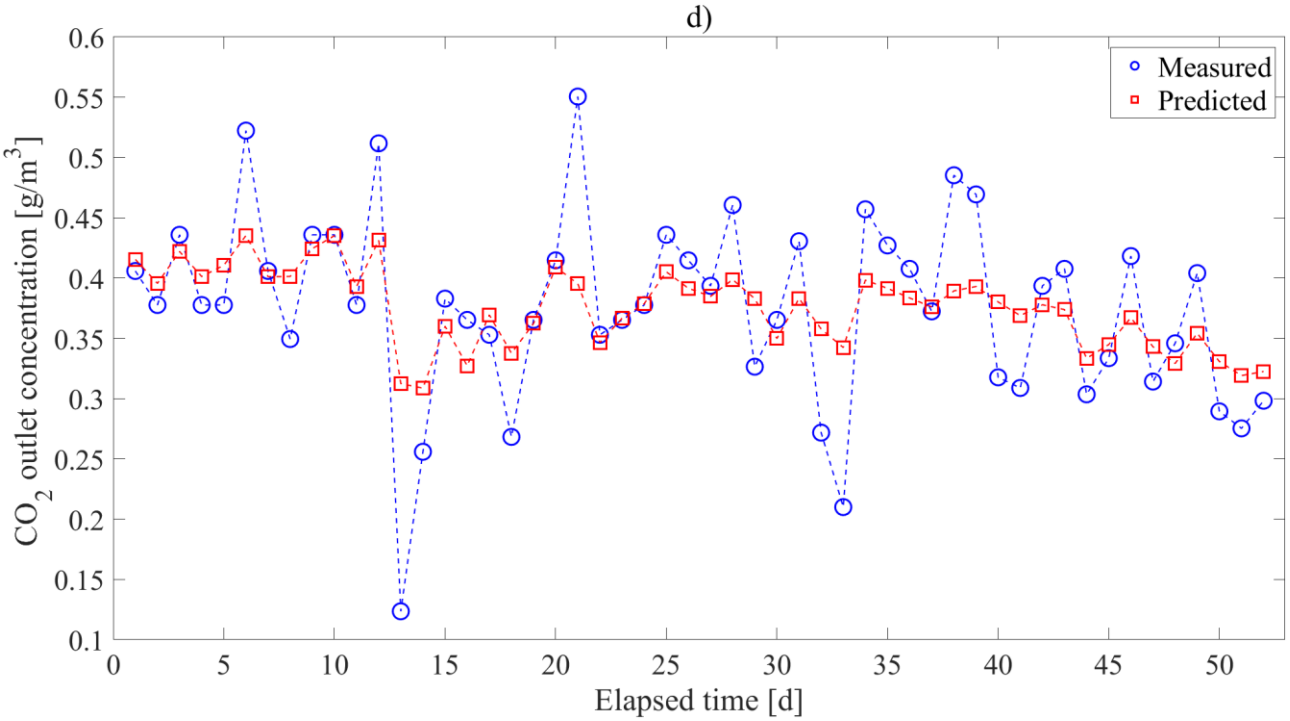

b)
Measured
Predicted
CO_2 outlet concentration [g/m^3]
Elapsed time [d]
0.6
0.55
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0
5
10
15
20
25
30
35
40
45
50

946

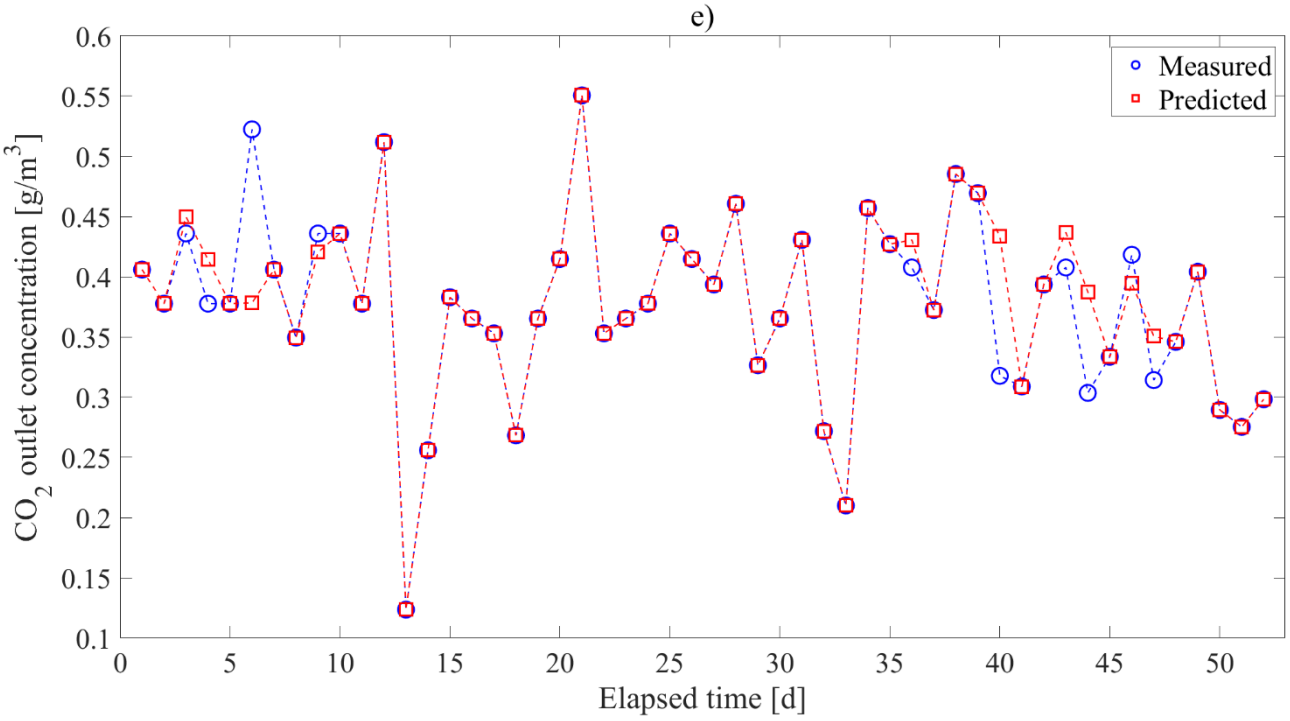

c)
Measured
Predicted
CO_2 outlet concentration [g/m^3]
Elapsed time [d]
0.6
0.55
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0
5
10
15
20
25
30
35
40
45
50

947

948


d)
Measured
Predicted
CO_2 outlet concentration [g/m^3]
Elapsed time [d]
0.6
0.55
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0
5
10
15
20
25
30
35
40
45
50

949



950

951 **Figure 5.** Predictions of CO_2 outlet concentrations from the aPBR using (a) ANN, (b) multi-task

952 ANN, (c) SVM, (d) RF, and (e) LSBoost models.

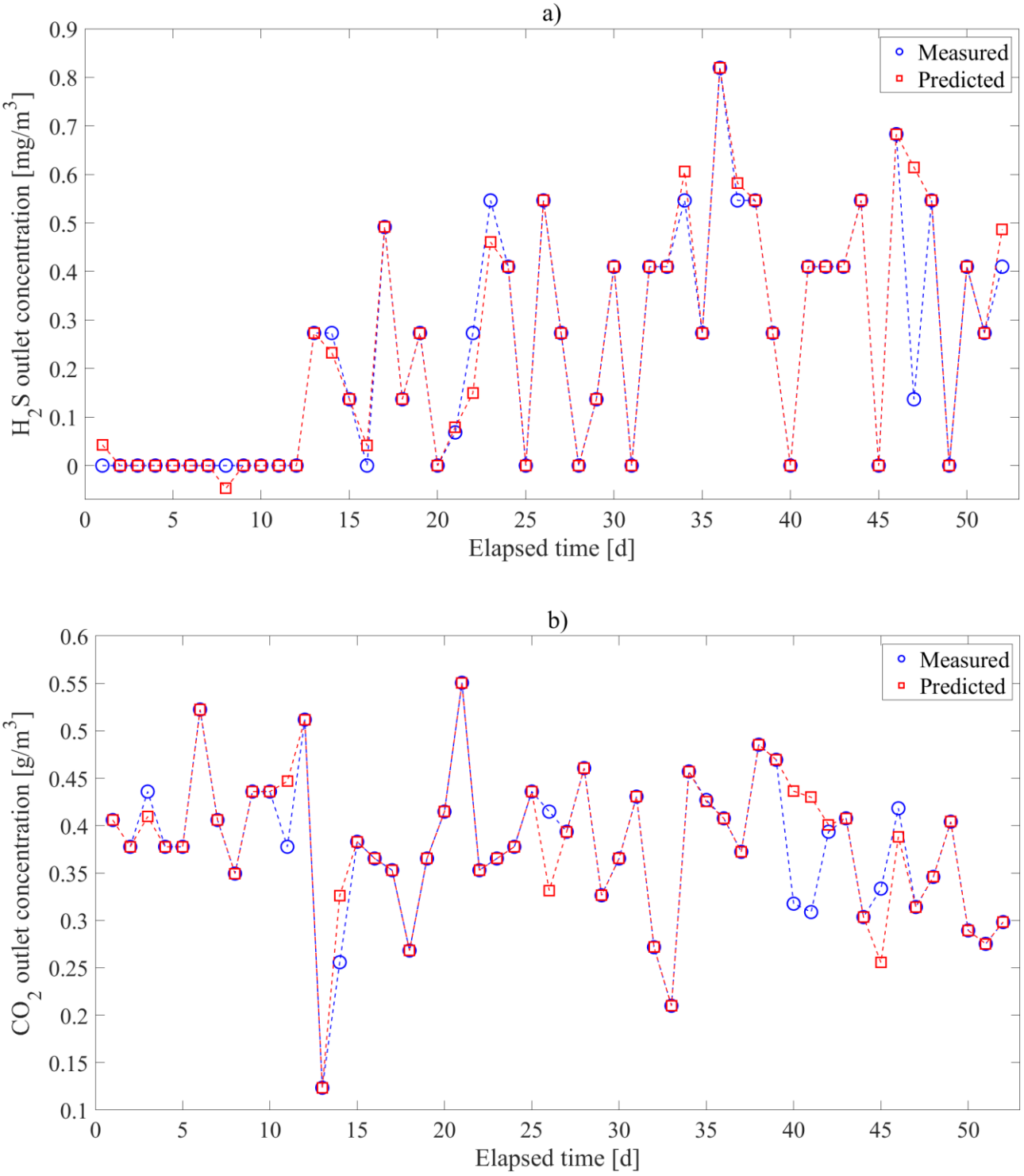


Figure 6. LSBoost+LSBoost approach for predicting (a) H_2S and (b) CO_2 outlet concentrations from the MBBR+aPBR integrated system.

957 **Table**

958 **Table 1.** Statistical parameters for the MBBR and aPBR.

Parameter	Measurement units	MBBR			aPBR		
		Mean	**St. Dev.**	**Cv**	**Mean**	**St. Dev.**	**Cv**
Temperature	°C	21.96	2.19	0.10	24.08	1.86	0.08
pH	-	6.37	0.27	0.04	7.00	0.88	0.13
Saturated O_2	%	80.85	7.09	0.09	89.73	10.18	0.11
Dissolved O_2	mg/L	6.97	0.69	0.10	7.72	1.07	0.14
TSS	g/L	1.25	0.17	0.14	1.41	0.33	0.24
H_2S inlet	mg/m^3	216.48	60.58	0.28	1.43	0.45	0.32
CO_2 inlet	g/m^3	0.59	0.07	0.11	1.30	0.42	0.32
H_2S outlet	mg/m^3	1.43	0.45	0.32	0.24	0.23	0.98
CO_2 outlet	g/m^3	1.30	0.42	0.32	0.38	0.08	0.21

959

Declaration of interests

☒The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: