Modelling N_2O and NH_4^+ in a Full-scale Wastewater Treatment Plant with NeuralODE

Francisco José Matos Nogueira Filho[1], **Michela Mulas**[1]

[1]Departamento de Engenharia de Teleinformática – Universidade Federal do Ceará (UFC)
Fortaleza – CE – Brazil

francisconog@alu.ufc.br, michela.mulas@ufc.br

Abstract. *Accurate prediction of nitrous oxide (N_2O) emissions from wastewater treatment plants is essential for mitigation efforts. Since N_2O production is linked to nitrification, forecasting ammonium (NH_4^+) helps capture precursor conditions. This study evaluates a continuous-time Neural Ordinary Differential Equation (NODE) model for predicting N_2O and NH_4^+ using full-scale data, comparing it with NARX and LSTM models. Preliminary results show that compared with discrete models, NODE captures short-term variability and explains the fraction of the observed variance in both nitrogen compounds, highlighting the potential of continuous-time neural models for short-term prediction for supporting advanced aeration control strategies.*

Resumo. *Previsão precisa das emissões de óxido nitroso (N_2O) em estações de tratamento de águas residuais é essencial para esforços de mitigação. Como sua produção está ligada à nitrificação, prever amônio (NH_4^+) ajuda capturar condições precursoras. Este estudo avalia um modelo de Equação Diferencial Ordinária Neural em tempo contínuo (NODE) para prever N_2O e NH_4^+ usando dados em escala real, comparando com modelos NARX e LSTM. Resultados preliminares mostram que, frente a modelos discretos, NODE captura a variabilidade de curto prazo e explica a fração da variância observada em N_2O e NH_4^+ , destacando potencial de modelos neurais em tempo contínuo para previsão de curto prazo para apoiar estratégias avançadas de controle de aeração.*

1. Introduction

While essential for public health, wastewater treatment plants (WWTPs) are also major sources of greenhouse gases. Nitrous oxide (N_2O) is of particular concern, with a 100-year global warming potential 273 times that of carbon dioxide [Masson-Delmotte and et. al 2021]. N_2O emissions from WWTPs increased by 44% between 1990 and 2014, contributing up to 26% of the water supply chain's total greenhouse gas footprint [Vasilaki et al. 2019]. Effective mitigation therefore depends on quantifying N_2O under dynamic operational conditions, enabling identification of high-emission events and the design of targeted control strategies.

Mechanistic models, such as the Activated Sludge Model 2d extended for N_2O (ASM2d-N_2O) [Massara et al. 2018], represent the complex biological pathways underlying N_2O production, but they require calibration of numerous kinetic parameters, often constrained by incomplete knowledge of the processes [Li et al. 2022, Hansen et al. 2024b]. Due to their computational complexity and the large number of

parameters, these models are challenging to implement for real-time control in full-scale WWTPs. As an alternative, data-driven machine learning approaches have emerged as promising tools for predicting N_2O emissions [Khalil et al. 2023, Szelag et al. 2023, Debel et al. 2025]. However, standard deep learning models operate at discrete time intervals and often struggle to generalize across highly non-stationary, time-varying biological regimes. These limitations highlight the need for adaptive modelling strategies capable of capturing continuous-time process dynamics, motivating the adoption of approaches such as Neural Ordinary Differential Equations (NODEs).

NODEs address the continuous nature of process dynamics by directly parameterising the hidden state derivative [Chen et al. 2018]. However, the complex biological reactions in wastewater treatment make the system highly sensitive to small changes, which results in complex hidden states in the NODE model that can be difficult to integrate numerically. Recent advancements mitigate these challenges through incremental training strategies, such as initialising the model with collocation methods to approximate derivatives directly from observational data [Huang et al. 2025], as well as applying regularisations to the model's dynamics and Jacobian, which enforce smooth, physically plausible trajectories and reduce integration drift [Finlay et al. 2020].

This study develops a NODE framework for predicting continuous N_2O emissions in full-scale WWTPs. To mitigate training instabilities, we combine the incremental collocation strategy of [Huang et al. 2025] with the vector field regularisations of [Finlay et al. 2020], and apply a continuous rolling-retraining scheme to account for system non-stationarity. The predictive performance and adaptive capabilities of this framework are compared with discrete Nonlinear AutoRegressive model with eXogenous inputs (NARX) and Long Short-Term Memory network (LSTM) baselines.

2. Material and Methods

2.1. Case Study and Dataset Description

Data for this study originates from an open-access dataset [Hansen et al. 2024a]. The data were collected from Avedøre WWTP in Copenhagen, Denmark, serving 350,000 population equivalents. Biological treatment is performed in four parallel lines with two carousel reactors each, and this work focuses on the activated sludge process (ASP) in the aeration tanks (Figure 1). Nitrification and denitrification are controlled via aeration using bubble diffusers, with operation based on ammonium, nitrate, and dissolved oxygen concentrations. After treatment, activated sludge is separated from the mixed liquor and recycled. Data were obtained from line 3, primarily from a single carousel reactor.

The dataset provides operational measurements from online sensors within the biological reactors, with a sample interval of 2 minutes. Dissolved N_2O (mg-N/L) is monitored using Clark-type electrochemical sensors (Unisense Environment, Denmark) with an accuracy of $\pm 5\%$. Additional measured variables include ammonium (NH_4^+, mg-N/L), nitrate (NO_3^-, mg-N/L), dissolved oxygen (DO, mg/L), temperature (T, °C), suspended solids (g/L), phosphate (PO_4, mg-P/L), as well as operational parameters such as inlet flow rate (Q_{IN}, m^3/h), airflow intensity (Q_{AIR}, Nm^3/h), and aeration valve positions (%). Although the full dataset spans July 2022 to July 2024, this study focuses on a subset from February to August 2023, selected based on data quality considerations.

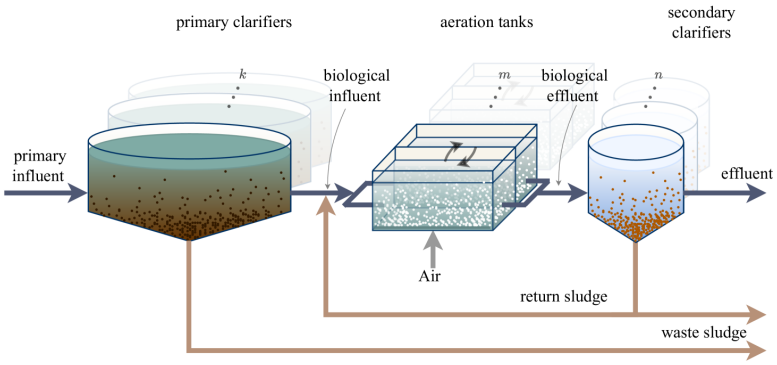

Figure 1. Layout of a wastewater treatment plant with an activated sludge process [Hansen et al. 2024c].

To ensure data quality and physical consistency for the data-driven models, a preprocessing pipeline is applied. The raw dataset is first aligned to establish temporal continuity and then resampled to 15-minute intervals to reduce dimensionality. Signal drift and sensor bias are mitigated by calibrating near-zero N_2O and DO measurements to zero, as suggested by [Hansen 2025]. Invalid values, including negative readings and N_2O measurements outside the sensor's physical range [0, 1.5] mg-N/L, are removed and imputed using a forward-fill strategy to preserve temporal causality.

2.2. Neural Ordinary Differential Equation

To model the continuous-time dynamics of key nitrogen compounds from discrete sensor measurements, we use a Neural Ordinary Differential Equation (NODE). The model tracks a 2-dimensional state vector representing reactor concentrations: $\mathbf{y}(t) = [N_2O(t), NH_4^+(t)]^T$. These variables are chosen because N_2O emissions are closely linked to nitrogenous nutrients and influenced by temperature variations, consistent with previous data-driven studies. The model is driven by an exogenous input vector $\mathbf{u}(t)$ containing $[Q_{IN}, Q_{AIR}, DO, T]$ alongside their explicit historical delays at lags of 1, 2, 4, and 8 steps (15-minute intervals). By embedding these specific lagged features, we evaluate whether modelling the continuous-time derivatives of key state variables enables the NODE framework to accurately learn dynamic input-output relationships, providing continuous-time predictions of N_2O and NH_4^+ concentrations that reflect the system's observed behaviour.

Unlike traditional discrete-time models which map inputs to outputs through a fixed sequence of layers, NODEs learn the continuous-time derivative of the system state directly using a neural network. The future trajectory is obtained by solving an initial value problem via a numerical ODE solver. In this study, the continuous-time dynamics of the system are modelled by a neural network f_θ, which predicts the rate of change of the state vector $\mathbf{y}(t)$ based on the current state and inputs $\mathbf{u}$:

$$\frac{d\mathbf{y}}{dt} = f(\mathbf{y}(t), \mathbf{u}(t), \theta) \tag{1}$$

By learning the parameters θ, the network captures how the state evolves over time. The full trajectory of $\mathbf{y}(t)$ is obtained by solving this initial value problem with a numerical ODE solver, which integrates the predicted derivatives continuously. A key

challenge in applying NODEs to real-world data is that numerical solvers evaluate derivatives at arbitrary, continuous time points, while the measured inputs ($\mathbf{u}(t)$) are only available at discrete intervals. To handle this, we use linear interpolation, which at each solver evaluation identifies the nearest recorded measurements, before and after the current time and computes a linear approximation of $\mathbf{u}(t)$. This yields a continuous, piecewise-linear representation of the input $\mathbf{u}(t)$, enabling continuous-time predictions during integration.

To preserve physical feature magnitudes and ensure stable gradient descent, the inputs are Z-score normalised before being passed through the Multi-Layer Perceptron (three 32-unit hidden layers with Sigmoid activations). The network's output is then denormalised to yield physical concentration derivatives $[\dot{N}_2O, \dot{N}H_4]^T$.

We utilize a two-phase training strategy to efficiently optimize the network. In Phase 1 (Collocation Pre-training), inspired by the incremental collocation strategy of [Huang et al. 2025], the network is initialised without the ODE solver. To achieve this, temporal gradients $[\dot{N}_2O, \dot{N}H_4]^T$ are first approximated from the discrete time-series using a second-order central difference scheme. The network is then trained via standard regression to map $[\mathbf{y}_k, \mathbf{u}_k]$ directly to the numerical gradients. This phase provides a fast and computationally inexpensive "warm start" for the network weights. In Phase 2 (End-to-End NODE Training), pre-trained weights are integrated via explicit 5th-order Runge Kutta solver (Dopri5) [Dormand and Prince 1980] over the 6-hour horizon to minimize the composite loss: $\mathcal{L} = \mathcal{L}_{\text{WMSE}} + \lambda_{\text{kin}} \int \|f_\theta\|^2 dt + \lambda_{\text{jac}} \int \|\nabla_y f_\theta\|_F^2 dt$. To prevent large NH_4^+ magnitudes from dominating, the Weighted Mean Squared Error ($\mathcal{L}_{\text{WMSE}}$) scales errors using inverse-mean-squared weights. The kinetic and Jacobian regularisations (λ_{kin} = λ_{jac} = 10^{-6}) penalize the neural vector field (f_θ) and its spatial gradients ($\nabla_y f_\theta$), promoting smooth and physically plausible dynamics.

To assess real-time forecasting performance under operational conditions, we simulate a rolling deployment in which a dedicated model is trained daily using the preceding 14 days of historical data. Normalisation statistics are dynamically updated for each window, allowing the model to continuously adapt to seasonal shifts and changing operational regimes before forecasting the next 24 steps (6 hours).

2.3. Nonlinear Autoregressive with Exogenous Inputs Model

The Nonlinear Autoregressive Network with Exogenous Inputs (NARX) serves as the discrete-time system identification baseline [Leontaritis and Billings 1985]. It approximates an unknown nonlinear function $f(\cdot)$ mapping past system states and external inputs to future outputs [Billings 2013]. As a stateless feed-forward neural network, it lacks the internal memory gating of an LSTM and the continuous integration of a NODE. Temporal dependencies are encoded by flattening a 3-hour history of states and the 6-hour trajectory of exogenous drivers into a 526-dimensional input vector. To avoid compounding errors from recursive forecasting, the network processes this input in a single forward pass to produce a 24-step multi-horizon forecast. Predictions are made in Z-score normalised space and optimised using standard mean squared error (MSE) loss.

2.4. Long Short-Term Memory

The Long Short-Term Memory (LSTM) network serves as the recurrent deep learning baseline. It captures temporal dependencies via a persistent cell state regulated by nonlinear forget, input, and output gates [Hochreiter and Schmidhuber 1997]. A single-layer

LSTM with 10 hidden units [Hansen et al. 2024b] processes a 3-hour historical sequence of concatenated system states and unlagged exogenous drivers. At each discrete step, the network updates the cell state and filters it through a hyperbolic tangent function and an output gate to compute the hidden state, which serves as a fixed-size summary of the historical context. To forecast the subsequent 6-hour horizon, this summary is concatenated with the scheduled future exogenous inputs and projected forward via a linear decoder.

Operating strictly in discrete time, the LSTM treats future inputs as a fixed conditioning signals rather than continuous-time trajectories. Predictions are made in Z-score normalised space and optimised using a standard MSE loss, without the vector-field regularisations used in continuous ODE. For fair comparison, the LSTM shares the same data preprocessing and 14-day rolling retraining strategy as the NODE framework.

2.5. Experimental Setup and Evaluation Metrics

All models were trained using the AdamW optimiser [Loshchilov and Hutter 2019], which combines adaptive moment estimation with decoupled weight decay to ensure stable and efficient training. To evaluate seasonal robustness and generalisation, the 14-day walk-forward validation scheme was applied across three distinct one-week test periods representing different operational regimes: winter (February), spring (May), and summer (August).

To ensure physical plausibility, a Rectified Linear Unit (ReLU) transformation was applied to all model forecasts during the evaluation phase to strictly enforce non-negativity. Model performance over the 6-hour forecast horizon was then evaluated using multiple metrics. Root Mean Square Error (RMSE) measures the magnitude of prediction errors, while the coefficient of determination (R^2) quantifies how well the model explains the variance in the observed data. Goodness of Fit (GoF) reports the fraction of observed dynamics captured by the model, and Symmetric Mean Absolute Percentage Error (SMAPE) provides a normalised measure of relative forecast error, reducing bias from large values. Predictive reliability was further assessed using the Sign Test, which tracks the model's ability to correctly predict the direction of concentration changes. Together, these metrics capture both the accuracy and reliability of the forecasts.

3. Results and Discussion

3.1. Predictive Accuracy and Seasonal Generalisation

To evaluate the efficacy of the continuous-time framework, the NODE was compared with the discrete-time NARX and LSTM architectures over a 6-hour prediction horizon. The comparative considered focused on absolute accuracy, variance capture, directional reliability, and seasonal robustness. Table 1 summarises the performance metrics over the 6-hour prediction horizon, while Table 2 presents the seasonal analysis.

In modelling N_2O dynamics, the NODE achieved the lowest overall RMSE (0.1750 mg-N/L) and the highest explained variance (R^2 = 0.7051). Over the 6-hour prediction horizon, both the NARX and LSTM exhibited negative GoF scores (-0.2412 and -0.1576, respectively). A negative GoF indicates that the multi-step predictions performed worse than a historical mean forecast, suggesting susceptibility to long-term error accumulation or drift. The NODE maintained a positive GoF (0.0549). For NH_4^+ predictions, the single-shot NARX model achieved a slightly lower absolute error than

the NODE (RMSE 1.645 vs. 1.707 mg-N/L). However, the NARX exhibited low variance capture (R^2 = 0.1380), indicating flatter, mean-reverting predictions. In contrast, the NODE captured a higher proportion of the actual system variance (R^2 = 0.3328).

Directional reliability was assessed using the Sign Test, revealing that model performance differed depending on the chemical species. For NH_4^+, the NODE yielded the highest directional accuracy at the 6-hour horizon (74.79%), outperforming the NARX (69.75%) and the LSTM, which did not reliably capture the system's directionality (48.55%, worse than random). Conversely, predicting the directional shifts of N_2O proved slightly more favourable for the discrete-time architectures, though all models performed comparably at this extended horizon. The NARX achieved the highest accuracy at 65.27%, closely followed by the LSTM (64.34%) and the NODE (63.38%). Consequently, while the continuous formulation provides the most reliable basis for tracking NH_4^+ trajectories, the single-shot NARX exhibits a marginal advantage in anticipating overarching N_2O trends.

To assess temperature-driven generalisation, the models were evaluated across three distinct operational regimes. During the winter period (Feb 01-07), low absolute N_2O emissions yielded minimal errors across all models (RMSE 0.019-0.022 mg-N/L). While all architectures struggled to isolate signal from noise, the NODE maintained the highest NH_4^+ variance capture (R^2 =0.354). In spring (May 01-07) the data show larger peak events and higher variability; here, the NODE achieved an N_2O RMSE of 0.268 mg-N/L (R^2 =0.573) compared to the LSTM's 0.323 mg-N/L (R^2 =0.391). During the summer week (Aug 01-07), the LSTM outperformed the NODE on NH_4^+ error (RMSE 1.086 vs. 1.285 mg-N/L). However, while the discrete baselines exhibited a trade-off in accuracy between the two chemical species depending on the season, the NODE consistently provided balanced predictions across the evaluated temperature spectrum.

Table 1. Global performance metrics over the 6-hour prediction horizon.

Metric	Variable	NODE	LSTM	NARX
RMSE (mg-N/L)	N_2O	**0.1750**	0.2151	0.2068
	NH_4^+	1.7072	1.7179	**1.6452**
R^2	N_2O	**0.7051**	0.5662	0.5010
	NH_4^+	**0.3328**	0.2792	0.1380
SMAPE (%)	N_2O	**89.96**	102.89	117.66
	NH_4^+	**38.69**	41.31	42.02
GoF	N_2O	**0.0549**	-0.1576	-0.2412
	NH_4^+	0.4089	0.4039	**0.4450**
Sign Test (%)	N_2O	63.38	64.34	**65.27**
	NH_4^+	**74.79**	48.55	69.75

3.2. Horizon Analysis and Error Degradation

For aeration control, forecast accuracy across varying time horizons is critical; short-term predictions (15-60 minutes) dictate immediate aeration actions, while extended horizons (3-6 hours) inform trajectory planning. Analysis of performance degradation over a 6-hour window reveals distinct behavioural differences between the continuous- and

Table 2. Seasonal analysis across varying operational regimes. Performance is reported for a 6-hour forecast horizon.

Season	Ammonium (NH_4^+)						Nitrous Oxide (N_2O)					
	RMSE (mg-N/L)			R^2			RMSE (mg-N/L)			R^2		
	NODE	NARX	LSTM	**NODE**	NARX	LSTM	**NODE**	NARX	LSTM	**NODE**	NARX	LSTM
Winter (Feb)	1.671	**1.639**	1.903	**0.354**	0.163	0.309	0.020	**0.019**	0.022	**0.187**	0.081	0.001
Spring (May)	2.117	**1.779**	2.154	0.150	**0.152**	0.115	**0.268**	0.290	0.323	**0.573**	0.472	0.391
Summer (Aug)	1.285	1.212	**1.086**	0.431	0.349	**0.507**	**0.088**	0.092	0.113	0.431	**0.486**	0.276

discrete-time architectures. The NODE demonstrates short-term precision by seamlessly integrating forward from the initial state. At the 30-minute mark, the NODE captures the immediate local gradients required for receding-horizon control, achieving an N_2O R^2 of 0.9757 and a directional Sign Test of 75.98%, alongside an NH_4^+ RMSE of 0.4545 mg-N/L. Conversely, the discrete-time NARX and LSTM architectures exhibit a flatter error profile, projecting a generalised trajectory from the very first time step. At 30 minutes, the NARX produces a lower N_2O Sign Test of 62.26% and an N_2O RMSE of 0.1747 mg-N/L, nearly identical to its 6-hour RMSE of 0.2068 mg-N/L. This indicates that discrete architectures tend to output persistent average values rather than capturing rapid local deltas, rendering them less precise for immediate short-term control actuation.

As the prediction horizon extends to 6 hours, the NODE naturally degrades due to the accumulation of integration errors, with its NH_4^+ RMSE increasing from 0.8397 (at 1 hour) to 1.7072 mg-N/L. However, despite this absolute magnitude drift, the continuous formulation retains better structural integrity over long horizons. At 6 hours, the NODE maintains an N_2O R^2 of 0.7051, compared to the NARX (0.5010) and LSTM (0.5662), indicating that it preserves the general shape and timing of concentration peaks. Crucially for constraint avoidance, the NODE's directional reliability for NH_4^+ remains fairly stable over time, degrading only slightly from 78.56% at 1 hour to 74.79% at 6 hours. By comparison, the NARX reaches 69.75%, while the recurrent LSTM operates below random chance (48.55%) across all measured horizons. However, for N_2O directionality, the discrete models actually improve as the horizon extends, with the NARX reaching 65.27% at 6 hours, demonstrating that discrete mappings can capture generalised long-term trends even when failing to capture local gradients.

To visually contextualise these degradation metrics, Figure 2 presents the 6-hour time-series forecasts. For conciseness, only the Summer period is displayed, though it effectively illustrates the behavioural trends observed across all test regimes. Visual inspection corroborates the statistical superiority of the continuous formulation during dynamic shifts. The NODE demonstrates distinctly higher responsiveness to environmental changes, successfully tracking the underlying trajectories and transient NH_4^+ peaks that the discrete baselines either delay or flatten into historical averages.

3.3. Operational Implications and Future Directions

The comparative analysis provides clear insights for integrating data-driven models into WWTP control architectures. The NODE's short-term precision and directional reliability make it highly suitable for receding-horizon forecasting. Specifically, its ability to accurately capture 15- to 60-minute local NH_4^+ gradients provides a robust foundation for real-time aeration adjustments, enabling operators to mitigate effluent constraint viola-

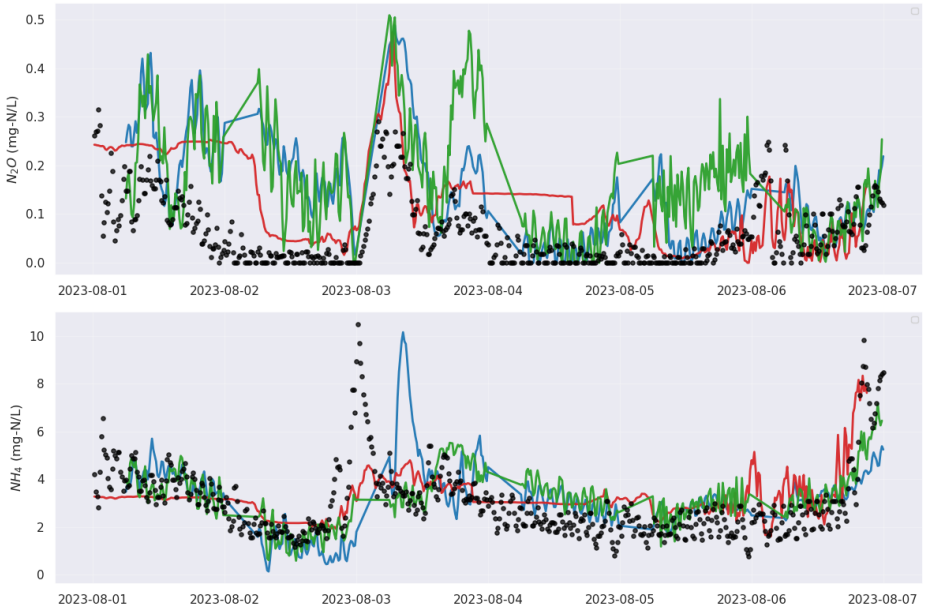

Figure 2. Time-series comparison of the 6-hour forecast for the NODE (blue line), NARX (red line), and LSTM (green line) against historical observations (black dots) for N_2O (top) and NH_4^+ (bottom) during the Summer period.

tions during transient peak events.

However, identified vulnerabilities - namely, long-term integration drift - highlight areas for architectural refinement. While this study successfully employed a post-evaluation ReLU filter to pragmatically enforce non-negative concentrations, future iterations should embed these physical constraints intrinsically during training. This can be achieved by applying activation bounds directly to the decoder output or by integrating Physics-Informed Neural Network (PINN) loss penalties that strictly enforce mass-balance conservation and zero-lower bounds. Furthermore, because the discrete NARX demonstrated a slight advantage in tracking overarching N_2O trends, future research could explore hybrid strategies that combine the NODE for short-term precision with discrete models for long-term trajectory estimation. To ensure broader generalisation, this continuous-time framework must be evaluated across diverse full-scale WWTPs and expanded to encompass multi-gas emission dynamics (e.g., simultaneous N_2O and CO_2). Finally, deploying the NODE within a closed-loop control strategy remains the critical next step to validate its real-world energy and mitigation potential.

4. Conclusion

This study evaluated a continuous-time NODE against discrete architectures (NARX, LSTM) for modelling N_2O and NH_4^+ dynamics in a full-scale wastewater treatment plant. Results demonstrate that standard discrete models struggle to capture short-term local gradients, frequently defaulting to flattened average trajectories. In contrast, the continuous-time NODE achieved superior short-term precision and captured significantly more system variance. Furthermore, the NODE maintained robust directional reliability for NH_4^+ over extended 6-hour horizons, supporting informed operational decision-making. Although the discrete models proved slightly more adept at tracking overarching long-term

N_2O trends, and all architectures require intrinsic physical constraints to fully mitigate long-term integration drift, the continuous-time framework demonstrates clear advantages for immediate forecasting. Ultimately, the NODE is best positioned not as a standalone solution, but as a highly effective short-term prediction component within broader, multi-horizon control strategies for wastewater treatment.

Data and Code Availability

The dataset, data preprocessing scripts, model training source code, and pre-trained weights used in this study are available from the corresponding author upon reasonable request.

References

Billings, S. A. (2013). *Nonlinear system identification: NARMAX methods in the time, frequency, and spatio-temporal domains*. John Wiley & Sons.

Chen, R. T., Rubanova, Y., Bettencourt, J., and Duvenaud, D. K. (2018). Neural ordinary differential equations. *Advances in neural information processing systems*, 31.

Debel, L., Stentoft, P. A., Stokholm-Bjerregaard, M. A., Ortiz-Arroyo, D., and Durdevic, P. (2025). Grey-box model of N_2O and NH_4^+ for predictive control in the activated sludge process. *Journal of Water Process Engineering*, 72:107503.

Dormand, J. R. and Prince, P. J. (1980). A family of embedded runge-kutta formulae. *Journal of computational and applied mathematics*, 6(1):19–26.

Finlay, C., Jacobsen, J.-H., Nurbekyan, L., and Oberman, A. (2020). How to train your neural ODE: The world of jacobian and kinetic regularization. In *International conference on machine learning*, pages 3154–3164. PMLR.

Hansen, L. D. (2025). *Data-driven Modeling and Predictive Control for N_2O Mitigation in Wastewater Treatment Processes*. PhD thesis, Aalborg University, Denmark.

Hansen, L. D., Rani, A., Stokholm-Bjerregaard, M. A., Stentoft, P. A., Ortiz Arroyo, D., and Durdevic, P. (2024a). Time series dataset for modeling and forecasting of N_2O in wastewater treatment. arXiv preprint and associated dataset. Version 4 of the dataset available at Mendeley Data.

Hansen, L. D., Stentoft, P. A., Ortiz-Arroyo, D., and Durdevic, P. (2024b). Black-box modelling of non-stationary N_2O dynamics in a full-scale wastewater treatment plant. *IFAC-PapersOnLine*, 58(28):714–719. The 4th Modeling, Estimation, and Control Conference 2024.

Hansen, L. D., Stentoft, P. A., Ortiz-Arroyo, D., and Durdevic, P. (2024c). Exploring data quality and seasonal variations of N_2O in wastewater treatment: a modeling perspective. *Water Practice & Technology*, 19(3):1016–1031.

Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8):1735–1780.

Huang, X., Kandris, K., and Katsou, E. (2025). Training stiff neural ordinary differential equations in data-driven wastewater process modelling. *Journal of Environmental Management*, 373:123870.

Khalil, M., AlSayed, A., Liu, Y., and Vanrolleghem, P. A. (2023). Machine learning for modeling N_2O emissions from wastewater treatment plants: Aligning model performance, complexity, and interpretability. *Water Research*, 245:120667.

Leontaritis, I. J. and Billings, S. A. (1985). Input-output parametric models for non-linear systems part i: deterministic non-linear systems. *International journal of control*, 41(2):303–328.

Li, K., Duan, H., Liu, L., Qiu, R., van den Akker, B., Ni, B.-J., Chen, T., Yin, H., Yuan, Z., and Ye, L. (2022). An integrated first principal and deep learning approach for modeling nitrous oxide emissions from wastewater treatment plants. *Environmental Science & Technology*, 56(4):2816–2826.

Loshchilov, I. and Hutter, F. (2019). Decoupled weight decay regularization.

Massara, T. M., Solís, B., Guisasola, A., Katsou, E., and Baeza, J. A. (2018). Development of an ASM2d-N_2O model to describe nitrous oxide emissions in municipal WWTPs under dynamic conditions. *Chemical Engineering Journal*, 335:185–196.

Masson-Delmotte, V. and et. al, editors (2021). *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.

Szelag, B., Zaborowska, E., and Makinia, J. (2023). An algorithm for selecting a machine learning method for predicting nitrous oxide emissions in municipal wastewater treatment plants. *Journal of Water Process Engineering*, 54:103939.

Vasilaki, V., Massara, T., Stanchev, P., Fatone, F., and Katsou, E. (2019). A decade of nitrous oxide (N_2O) monitoring in full-scale wastewater treatment processes: A critical review. *Water Research*, 161:392–412.