Water Science & Technology

Water Science & Technology Vol 00 No 0, 1 doi: 10.2166/wst.2026.231

Data-driven modelling of N_2O production in wastewater processes using neural ordinary differential equations

Xiangjun Huang [a,*], Alireza Mousavi[b], Kyriakos Kandris[c] and Evina Katsou[c]

[a] Department of Civil and Environmental Engineering, Brunel University London, London, UK
[b] Department of Mechanical and Aerospace Engineering, Brunel University London, London, UK
[c] Department of Civil and Environmental Engineering, Imperial College London, London, UK
*Corresponding author. E-mail: xiangjun.huang@brunel.ac.uk; huang_xiangjun@hotmail.com

XH, 0000-0001-9020-3490

ABSTRACT

Modelling nitrous oxide (N_2O) production in wastewater treatment processes presents greater challenges than for other components, owing to its multiple production pathways and pronounced spatiotemporal variations. This study proposes a novel data-driven approach employing neural ordinary differential equations (NODEs) to capture the intrinsic dynamics of N_2O production in typical activated sludge processes. The NODE models are trained directly on state trajectory data, which incorporate continuous influent variations and operational adjustments as external forcings to the system dynamics. To address these external influences, we extend standard training procedures. In addition, a normalisation technique and an incremental strategy are introduced to enhance the computational efficiency of NODE implementation in stiff wastewater systems. This methodology is validated using simulated data from the benchmark simulation model no. 1 (BSM1) plant, adapted to integrate the activated sludge model for greenhouse gases no. 1 (ASMG1). Results demonstrate the efficacy of NODE-based approach in accurately capturing the complex dynamics governing N_2O production, highlighting its potential for controlling and mitigating greenhouse gases emissions in wastewater treatment.

Key words: activated sludge, data-driven, modelling, neural ordinary differential equations, nitrous oxide

HIGHLIGHTS

- Captured the underlying dynamics of typical activated sludge processes, focusing on N_2O production, using neural ordinary differential equation (NODE) models.
- Developed a normalisation method for efficient training of stiff NODE models.
- Extended NODE training algorithms to incorporate exogenous inputs.

1. INTRODUCTION

Nitrous oxide (N_2O) emissions from wastewater treatment plants (WWTPs) pose a significant environmental threat. Accurate modelling of N_2O production is crucial for understanding, predicting, and ultimately mitigating these emissions in WWTPs.

Conventional biokinetic models for N_2O employ systems of ordinary differential equations (ODEs) to describe biochemical interactions among biomass and substrates, often structured using a Gujer matrix (Spérandio *et al.* 2022). These dynamics are typically expressed as:

$$\frac{dY(t)}{dt} = f(Y(t), \varphi) \tag{1}$$

where t denotes time, and vector $Y(t)$ represents temporal trajectories of the system's component states, specifically their evolving concentrations. These components are carefully defined based on the understanding of the underlying mechanisms, such as active heterotrophic biomass X_{BH}, readily biodegradable substrate S_S, and dissolved nitrous oxide S_{N_2O}. The function f is central, defining the component reaction kinetics – the relationship between component concentrations and their

corresponding rates of change over time (mathematically the derivatives of concentrations with respect to time t), and ϕ encapsulates stoichiometric and kinetic coefficients that modulate these relationships.

Current biokinetic models incorporate three widely recognised pathways of N_2O production (Figure 1): (1) hydroxylamine oxidation; (2) nitrifier denitrification; (3) heterotrophic denitrification. While abiotic reaction pathway is still under debate (Soler-Jofra *et al.* 2016; Su *et al.* 2019), well-calibrated models can estimate site-specific production and inform mitigation strategies (Ni *et al.* 2015; Liu *et al.* 2016; Lv *et al.* 2022). However, development of biokinetic models requires a complete understanding of the underlying mechanisms and judicious simplification and abstraction, and application of these models necessitates calibration of numerous parameters (Abulimiti *et al.* 2022), a process hindered by inherent non-linearity and parameter uncertainty (Belia *et al.* 2009; Khalil *et al.* 2024). These challenges limit the accuracy, reliability and adaptability of purely mechanistic models.

Data-driven modelling has emerged as a promising alternative to complex mechanistic models for quantifying N_2O emissions from WWTPs, particularly with advances in machine learning (Khalil *et al.* 2023). Initial research primarily utilised unsupervised learning and classification models to analyse emission patterns and identify contributing features. The field has since progressed to employ a diverse set of supervised regression models for quantitative prediction, including k-nearest neighbours, decision trees, ensemble methods like AdaBoost and random forests, and deep neural networks, with studies demonstrating high predictive performance on full-scale data (Porro *et al.* 2022; Khalil *et al.* 2024). However, a key limitation of traditional machine learning methods is their static, input-output nature, which often lacks a direct representation of the underlying biokinetic processes governing N_2O dynamics. To transcend this limitation, this study introduces a novel approach using neural ordinary differential equations (NODEs) (Chen *et al.* 2018). Different from conventional practice, NODEs integrate data-driven learning with the first-principles structure of mechanistic models, offering improved generalisation and a natural bridge between process knowledge and operational data. In addition, unlike traditional machine learning methods constrained by fixed time-steps, NODEs excel at dealing with irregular data frequently encountered in wastewater measurement (Kidger *et al.* 2020). The combination of this flexibility with the simulated mechanistic structure, makes NODEs a well-suited tool for modelling wastewater systems with such real-world complexities. This study is to demonstrate that this alternative approach to plant-wide simulations is both feasible and reliable.

In our method, NODEs retain the structural form of mechanistic ODEs but replace manually defined kinetic equations f by a neural network NN with parameters θ, as shown in Equation (2).

$$\frac{dY(t)}{dt} = NN(Y(t), \theta). \tag{2}$$

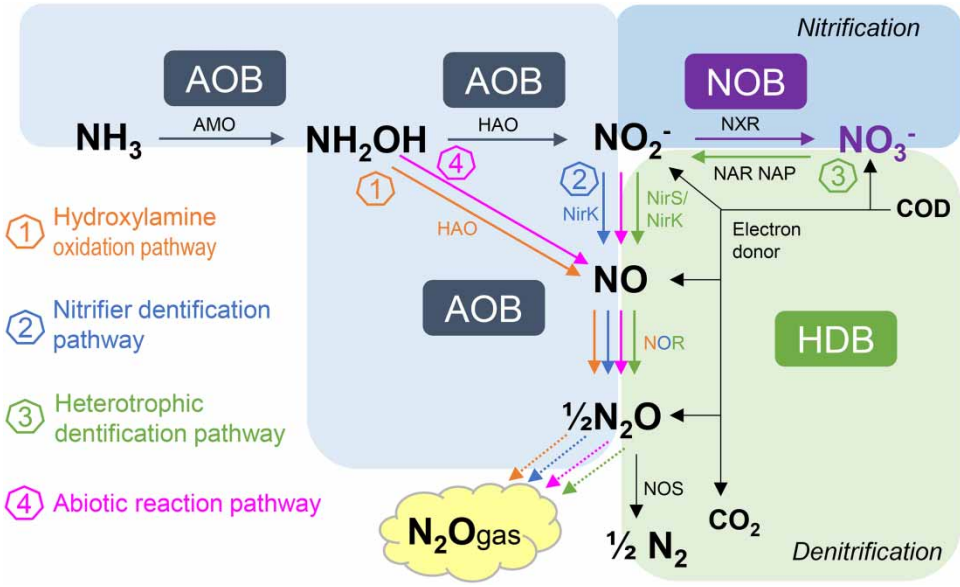

Figure 1 | Possible pathways of N_2O production from biological wastewater treatment processes.

The key distinction lies in how f is defined: mechanistic models rely on manually crafted formulas requiring significant domain expertise, whereas NODEs learn this process dynamics from monitoring data. This reduces dependency on hypothesised assumptions, enabling practical adaptation to observed real-world system behaviour.

However, training NODEs presents two primary challenges. First, monitoring data from WWTPs often contain signals from external forcings, originating from sources such as influent variability and operational parameter adjustments. Mathematically, this corresponds to solving NODEs with exogenous excitement (Böttcher & Asikis 2022). To address this, we extended the training algorithm that disentangles intrinsic dynamics from superimposed observations which include external influences. Second, the inherent stiffness in wastewater ODEs, which necessitates small step sizes for numerical solver stability, is exacerbated in NODEs, risking training instability (Kim *et al.* 2021). We propose a paired normalisation technique and a two-phase incremental training strategy to improve computational efficiency and robustness.

The proposed methodology is validated using the benchmark simulation model no. 1 (BSM1) (Alex *et al.* 2008), integrated with the activated sludge model for greenhouse gases no. 1 (ASMG1) (Guo & Vanrolleghem 2014). Results demonstrate the efficacy of NODEs in capturing minute-scale N_2O dynamics, highlighting their potential for real-world predictive modelling and emission mitigation.

2. METHODS

For initial methodological validation, synthetic data, generated following the BSM1 framework, were used to train and evaluate NODE models. This allowed us to leverage controllable ground truth for initial testing and analysis, circumventing the inherent challenges of real-world data, such as limited metadata and unknown underlying values. This crucial step precedes the application of the model to empirical observations.

The experimental framework employed the BSM1, a widely adopted platform in wastewater treatment research, to simulate component concentration trajectories. To capture N_2O production dynamics, the ASMG1 was substituted for the standard activated sludge model no. 1 (ASM1) in BSM1. All simulations were conducted using MATLAB® (The MathWorks Inc. 2024) as the primary computational environment.

Root mean squared error (RMSE) was selected for performance assessment. However, RMSE's sensitivity to data scale and outlier influence complicates cross-dataset comparisons, particularly for datasets of differing magnitudes. To mitigate these limitations, results were compared with the same length, and analyses focused on component-wise and reactor-specific trend rather than cumulative absolute values only.

2.1. Configuration of ASMG1-based BSM1 plant

The BSM1 plant comprises five serially connected activated sludge reactor tanks (termed 'R1–R5' for brevity) followed by a secondary clarifier. The first two tanks operate under anoxic conditions, while the remaining three function aerobically. This configuration mimics a conventional anoxic/oxic (A/O) process, integrating nitrification and denitrification for nitrogen removal. While retaining the BSM1's standard hydraulic and operational parameters, we substituted the original ASM1 with the ASMG1 to explicitly simulate greenhouse gas dynamics, primarily N_2O production. The plant's overview, Simulink® diagram, control parameter settings are detailed in Supplementary material, Appendix A.

The BSM1 framework provides 14-day influent datasets for three weather scenarios: dry, rain, and storm. The first 7 days are consistent across all scenarios, representing baseline stable conditions. Scenarios diverge in the second week: the dry scenario maintains baseline conditions, while the rain scenario introduces prolonged rainfall and the storm scenario features two superimposed storm events. These perturbations occur specifically between days 8 and 12, with their effects gradually diminishing as conditions return to baseline from day 12 to day 14. For comprehensive specifications – including reactor volumes, flow rates, and biomass kinetics – readers are directed to the BSM1 technical report (Alex *et al.* 2008).

2.2. Nitrous oxide production simulation

The ASMG1 model extends conventional activated sludge frameworks by integrating the dynamics of greenhouse gas components, particularly N_2O. It mechanistically describes N_2O production by ammonia-oxidising bacteria (AOB) and consumption by heterotrophic de-nitrifiers, incorporating temperature and pH dependencies while strategically excluding poorly quantifiable intermediates (e.g., hydroxylamine) to maintain computational tractability through 18 components and 15 kinetic equations.

Given that the model has been progressively refined through successive research rather than being defined statically, this work utilises the most recent formulation, synthesising foundational contributions from three key studies:

- Hiatt & Leslie Grady (2008)**:** Formalised N_2O production via a four-step denitrification pathway in ASMN model, elucidating the dual role of free ammonia and free nitrous acid as substrates and inhibitors in N_2O generation.
- Mampaey *et al.* (2013)**:** Incorporated a AOB denitrification pathway, enhancing mechanistic resolution of N_2O fluxes under varying redox conditions.
- Guo & Vanrolleghem (2014)**:** Recalibrated dissolved oxygen (DO) inhibition kinetic terms using Haldane-type equations, experimentally validating improved predictions of DO-mediated N_2O suppression.

We further revised the Gujer kinetic matrix and rate equations from prior publications to improve consistency, and rationalised coefficients to enhance continuity checks. Please see Supplementary material, Appendix B for complete formulations. Stoichiometric and kinetic parameters were initialised using default values from Flores-Alsina *et al.* (2014).

To ensure compatibility with the ASMG1 model structure, modifications were made to the BSM1 influent dataset:

- **Component substitutions:** Specifically, X_{BA} (total autotrophic biomass in ASM1) was replaced with X_{BA1} (AOB-specific biomass in ASMG1), and S_{NO} (combined nitrate/nitrite in ASM1) was substituted by S_{NO_3} (nitrate-only in ASMG1).
- **Component additions**: The dataset was further augmented to include six additional components: S_{NO_2} (nitrite), S_{NO} (nitric oxide), S_{N_2O} (nitrous oxide), S_{N_2} (dissolved nitrogen), X_{BA_2} (nitrite-oxidizing bacteria), and temperature. Initial influent concentrations for S_{NO_2}, S_{NO}, S_{N_2O}, S_{N_2}, X_{BA_2} were set to zero with temperature constant at 15 °C, and pH at 7.0.

Extended 14-day influent datasets (see Supplementary material, Appendix C) under dry, rain, and storm weather conditions were simulated in the modified BSM1-ASMG1 framework, producing five sets of incoming flow and component concentration trajectory data for each reactor at 15-min sampling intervals. While this paper focuses on N_2O dynamics, results for all components are provided in Supplementary material, Appendices E–G.

2.3. NODE with exogenous influences

WWTPs operate subject to continuous variations in influent loading and periodic operational adjustments. The temporal evolution of process state variables (i.e., component concentrations) is observed, and their rates of change are derived from this trajectory data. These derived rates represent a superposition of two distinct drivers:

1. **Intrinsic biochemical dynamics**: Endogenous interactions among biomass and substrates, governed by microbial kinetics and equilibrium-driven mechanisms (Henze *et al.* 2000). These dynamics, represented by kinetic rate functions (e.g., f in Equation (1)) in mechanistic models, are generalisable across systems and constitute our primary modelling target.
2. **Exogenous inputs and controls**: External perturbations from time-varying influent loads and operational control actions (e.g., aeration modulation, external carbon dosing, activated sludge recycling). These inputs and controls obscure the underlying intrinsic system behaviour.

Mathematically, the total rate of change of the component concentrations, $dY(t)/dt$, which is computed by the ODE solver during training, is expressed as the sum of the intrinsic biochemical reaction rates, $r_{\text{intrinsic}}$, and the exogenous rates, $r_{\text{exogenous}}$:

$$\frac{dY(t)}{dt} = r_{intrinsic}(Y(t)) + r_{exogenous}(t) \tag{3}$$

where (t) is the state vector of component concentrations, $r_{\text{intrinsic}}(Y(t))$ represents the biochemical reaction kinetics (dependent on the instantaneous state Y under the assumption of a continuous stirred-tank reactor), and $r_{\text{exogenous}}(t)$ captures the rate contributions from exogenous inputs and controls.

Here, $dY(t)/dt$ can be derived from the temporal trajectories of the observed component concentrations, while $r_{\text{exogenous}}$ is interpolated from operational data such as influent profile and control signals. During the training of the NODE model, this formulation allows us to isolate the intrinsic dynamics by rearranging the equation to $r_{\text{intrinsic}} = (dY(t)/dt) - r_{\text{exogenous}}$. This subtraction is performed at each ODE solver step, ensuring that the NODE is trained exclusively on the purified signal of the intrinsic system dynamics (see Figure 2). This methodology transforms the conventional BSM1 approach – where influent rates and reaction rates are added to model derivatives – by extracting intrinsic dynamics from composite data to yield purified training targets for the neural network. This ensures that only the intrinsic dynamics data train the NODE, a practice

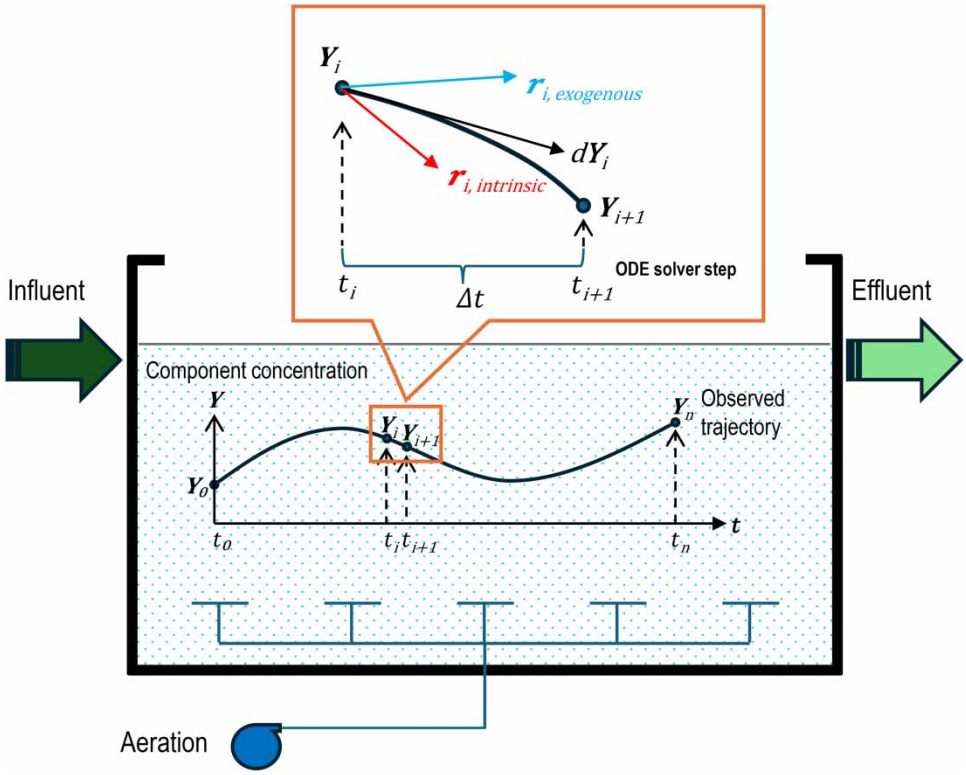

Figure 2 | Derivation of intrinsic biochemical reaction rates from observations including exogenous inputs and controls for NODE model training.

aligning with first-principles decomposition and upholding mathematical rigour by disentangling mechanistic drivers from superimposed observations (Alex *et al.* 2008).

2.4. Paired normalisation for stiff NODEs

Stiffness in dynamical systems arises from processes occurring across vastly disparate timescales (Hairer & Wanner 1996), a characteristic prevalent in wastewater treatment simulations. For example, heterotrophic biomass concentrations evolve over hours to days at 10^3 mg/L scale, while transient intermediates like N_2O fluctuate around 10^{-3} mg/L rapidly within minutes. In NODEs, this inherent stiffness is further amplified by neural network stochasticity – randomness in weight and bias initialisation, and gradient updates during training (Kim *et al.* 2021). Conventional stiff ODE solvers struggle with this dual challenge, leading to unstable training and poor convergence (Huang *et al.* 2025).

To address this, we integrate paired *input normalisation* and *output de-normalisation* layers into the NODE framework:

1. Input normalisation:

Input state variables (t) are normalised using z-score standardisation (Cabello-Solorzano *et al.* 2023):

$$Y_{normalised} = \frac{Y - \mu_Y}{\delta_Y} \tag{4}$$

where μ_Y (mean) and δ_Y (standard deviation) are computed component-wise from input state trajectories.

2. **Output de-normalisation:**

The neural network's output (normalised derivatives) is rescaled to original units:

$$\dot{Y} = \dot{Y}_{normalised} \cdot \delta_{\dot{Y}} + \mu_{\dot{Y}} \tag{5}$$

Here, the mean ($\mu_{\dot{Y}}$) and standard deviation ($\delta_{\dot{Y}}$) of the state derivatives $\dot{Y}$ implicitly represent plant-specific reaction rate ranges. Ideally, their values are informed by domain expertise or derived empirically from operational data. In the absence of such expertise or data, they can be calculated from trajectories of estimated state derivatives $\dot{Y}$ via finite differences, as shown in Equation (6):

$$\dot{Y} = \frac{1}{\Delta t}\,(y_2 - y_1,\ y_3 - y_2,\ \ldots,\ y_n - y_{n-1}) \tag{6}$$

This paired normalisation approach offers key advantages:

- **Preservation of dynamics**: Embedding normalisation pair inside the neural network maintains the functional relationships between states and their derivatives, ensuring kinetic rate functions can be learnt without distortion.
- **Numerical stability**: By constraining all components to comparable scales (typically near zero mean and unit variance), gradient propagation is stabilised, and solver instability due to magnitude disparities is mitigated.

Experimental validation (Huang *et al.* 2025) confirms this approach alleviates stiffness-induced training bottlenecks, stabilising convergence in wastewater NODE applications.

2.5. Incremental training strategy

Training NODEs is computationally intensive due to the repeated calls of ODE solver. This burden is particularly pronounced in stiff systems, where numerical solvers necessitate very small time-steps for stability. To further enhance computational efficiency, we propose a two-phase incremental training strategy:

1. **Collocation-based initialisation**: Collocation methods (Roesch *et al.* 2021) are employed to generate smoothed states-derivative pairs ($\boldsymbol{Y}$, $\dot{Y}$) through local polynomial regression. By training a neural network to approximate $\boldsymbol{Y} \rightarrow \dot{Y}$ without invoking an ODE solver, this phase circumvents stiffness-related instabilities and computational bottlenecks. The collocation step provides a near-optimal initialisation, positioning the network close to the loss landscape's global minimum.
2. **Direct NODE refinement**: The pre-trained network from collocation phase is fine-tuned using standard NODE training, which integrates the ODE solver into the computational graph. This phase incrementally refines model fidelity, leveraging the collocation-derived weights as a stabilising 'warm start.' By initialising near plausible solutions, direct training converges faster and avoids local minima induced by stiffness or poor initialisation.

Training hyperparameter settings, empirically determined through grid search, are catalogued in Supplementary material, Appendix D.

3. RESULTS AND DISCUSSION

3.1. Training convergence

By employing the proposed paired normalisation method and incremental training strategy, all models in our experiments demonstrated stable and convergent training behaviour. Figure 3 illustrates a representative training loss curve for the

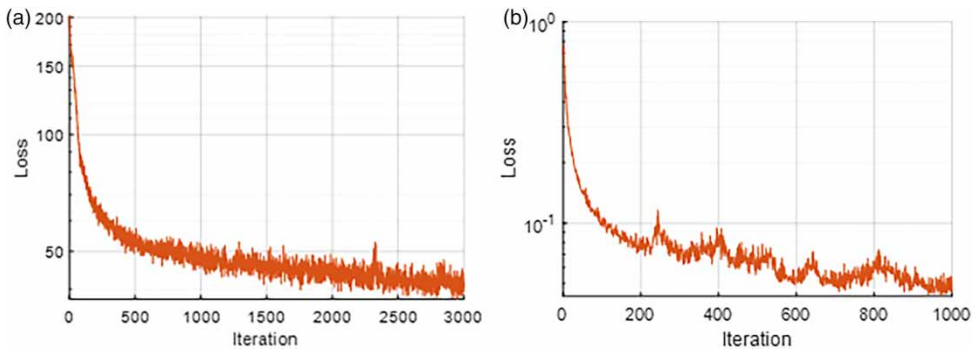

Figure 3 | Training loss curve: (a) collocation phase and (b) direct NODE training phase.

NODE model trained on the first 7 days (Days 0–6) of dry weather data. During both the collocation and direct NODE training phases, the mean absolute error (MAE) exhibits a monotonic decrease, indicating consistent and stable convergence.

3.2. Cross-scenario evaluation

Given the influent profile, days 8–12 represent a critical evaluation window, particularly for wet scenarios. To assess model generalisability across diverse weather patterns, a NODE model trained on 14 days of rain weather data – encompassing stable periods and rain events – was tested during the second week across dry and storm scenarios. Conversely, to evaluate extrapolation capability under unseen conditions, a separate NODE model was trained exclusively on the 14-day dry scenario dataset (containing no rainfall events) and tested on rain and storm scenarios (days 7–14), with emphasis on predictive accuracy during the hydraulically perturbed period (days 8–12).

To emulate real-world constraints where complete monitoring data may be unavailable, a NODE model with a reduced number of dimensions was also developed and tested. This is to explore the feasibility of modelling under partial system observability.

3.2.1. Rain weather data trained model

Figure 4 compares predicted and reference N_2O concentrations across all five reactors for the dry and storm scenarios (days 7–14), using a model trained solely on 14 days of rain scenario data. The model demonstrated robust cross-scenario predictive capability, with predictions closely tracking reference values and yielding low overall RMSEs (dry: 0.000377 mg N/L; storm: 0.000335 mg N/L). Per-reactor performance indicated distinct patterns influenced by location and process dynamics:

- **R1** proximate to the influent intake, exhibited the greatest sensitivity to flow variability. While capturing general trends under dry conditions, it showed the highest error (RMSE: 0.00072 mg N/L). Storm predictions achieved significantly higher accuracy (RMSE: 0.00028 mg N/L), precisely capturing two loading shocks between days 8 and 12 with an error approximately 2.6-fold lower than dry conditions.
- **R2** maintained consistently low prediction errors across both scenarios.

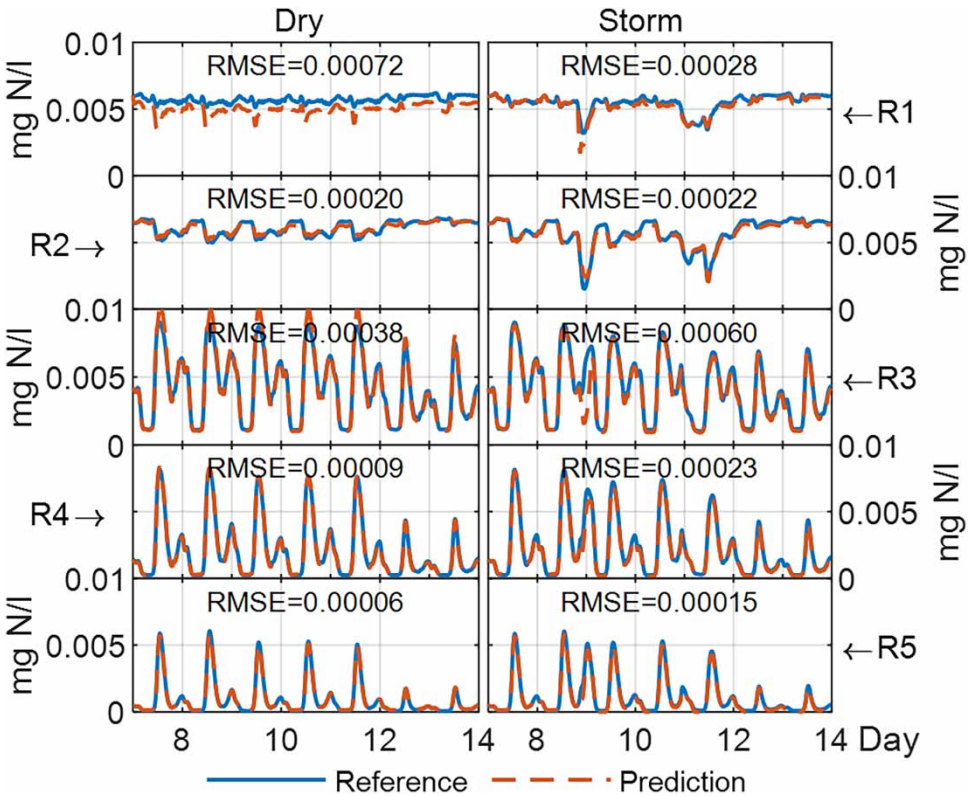

Figure 4 | Predicted vs. reference N_2O on days 7–14 in dry and storm scenarios (trained on days 0–14 of rain scenario).

- **R3**, an oxic transition zone, demonstrated heightened error during the storm scenario (RMSE: 0.00060 mg N/L), likely due to hydraulic transients intensifying anoxic–oxic transition dynamics. It maintained stable performance under dry conditions.
- **R4 and R5** exhibited the most stable and accurate predictions across both scenarios, attributed to the system approaching processing equilibrium downstream.

Notably, transition zones (R1 and R3) exhibited higher prediction errors compared to anoxic R2 or fully oxic R4–R5 zones under hydraulic perturbation. These findings reinforce that while hydraulic variations amplify prediction challenges in specific zones, the fundamental biokinetic mechanisms governing N_2O production retain their relative influence across the treatment line. The model's ability to maintain comparable overall accuracy between dry and storm scenarios, despite rain-only training, highlights its strong generalisation potential.

3.2.2. Dry weather data trained NODE model

Figure 5 quantifies the impact of hydraulic regime alignment between training (dry scenario, days 0–14) and evaluation (rain/storm scenarios, days 7–14) on N_2O prediction accuracy. Compared to the rain-trained model, the dry-trained model demonstrated declined predictive capability, evidenced by 5-fold higher overall RMSE values of 0.002347 mg N/L (rain) and 0.001953 mg N/L (storm). Reactor-specific analysis reveals three distinct response patterns:

- **Faltering prediction for upstream hydraulic shock**

  R1 and R2 exhibited significant accuracy degradation in prediction for wet conditions, particularly during hydraulic perturbations (days 8–12). R1 showed the most pronounced decline (rain: 0.0032 mg N/L; storm: 0.0022 mg N/L), with errors approximately 15% higher than R2 (rain: 0.00272 mg N/L; storm: 0.00192 mg N/L). This indicates limited generalisation of dry-trained dynamics to rain-induced hydraulic loading variations in upstream zones dominated by particulate hydrolysis and heterotrophic N_2O pathways.

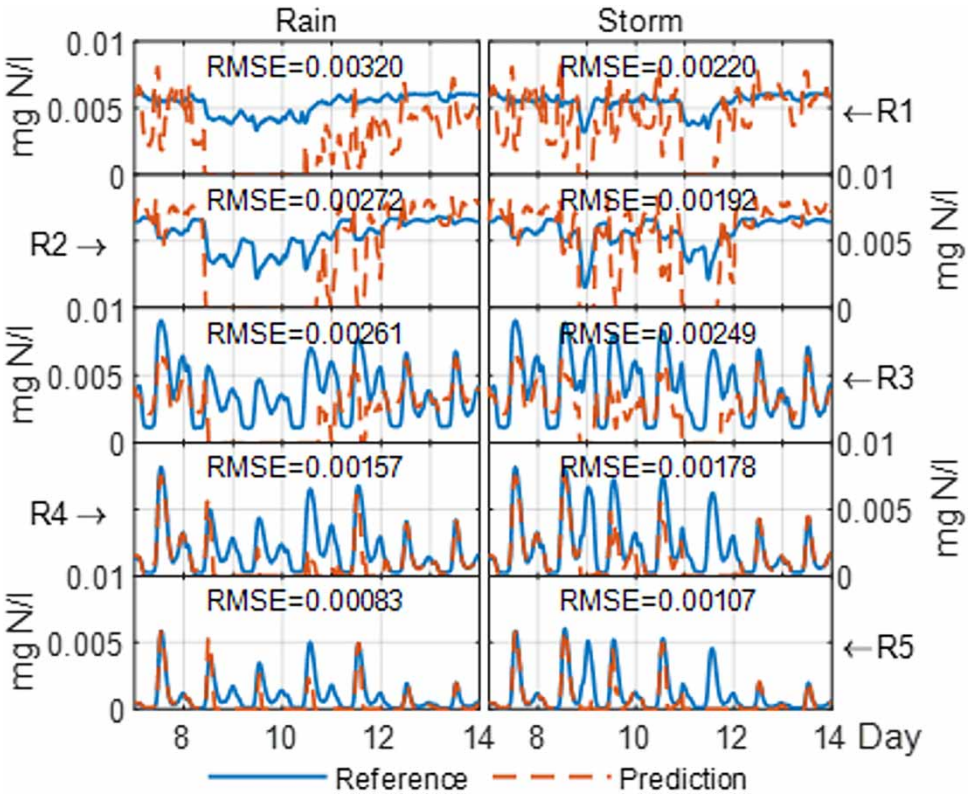

Figure 5 | Predicted vs. reference N_2O on days 7–14 in rain and storm scenarios (trained on days 0–14 of dry scenario).

• **Transition zone fluctuation**

R3 maintained high error levels across rain (RMSE: 0.00261 mg N/L) and storm (0.00249 mg N/L) scenarios, suggesting combined effects of hydraulic fluctuation and anoxic–oxic transition biokinetics. Prediction variability during days 8–12 was attenuated relative to R2, reflecting reduced hydraulic influence and greater dominance of intrinsic biological dynamics.

• **Downstream Attenuation**

R4 and R5 demonstrated progressively restored prediction stability (R4 rain: 0.00157 mg N/L, storm: 0.00178 mg N/L; R5 rain: 0.00083 mg N/L, storm: 0.00107 mg N/L), confirming serial reactor design dampens upstream perturbations. This physical buffering partially compensated for training data limitations in terminal reactors.

Collectively, these results highlight that while dry-exclusive training constrains upstream generalisation during wet scenarios, inherent process stability in transition and downstream zones mitigates overall performance degradation. The model exhibited limitations when predicting unseen hydraulic regimes, indicating a need for targeted wet-weather calibration for optimised upstream accuracy.

3.2.3. Analysis and implications

With adequate training, the NODE model demonstrated cross-scenario generalisation capacity, highlighting its utility for deployment in WWTPs. Crucially, the rain-trained model achieved superior cross-scenario performance relative to its dry-trained counterpart, attributable to the rain dataset's encapsulation of baseline dry weather patterns – a bidirectional generalisation not afforded by dry-trained models which lack wet condition exposures.

This performance asymmetry is mechanistically contextualised by Figure 6, which illustrates the joint distribution of four key components (S_s, X_{BA}, S_{NH_4}, S_{N_2O}) during the hydraulically active window (days 8–12). Dry conditions exhibit tightly clustered state-space trajectories within elevated concentration ranges (e.g., X_{BA} vs S_{NH_4}) representing stable process equilibria. Wet conditions (rain and storm) manifest expansive, heterogeneous distributions (e.g., S_S–X_{BA}, S_{NH_4} – S_{N_2O}), with broadened concentration envelopes, indicative of disrupted mass-transfer kinetics and community-level microbial stress under hydraulic perturbation.

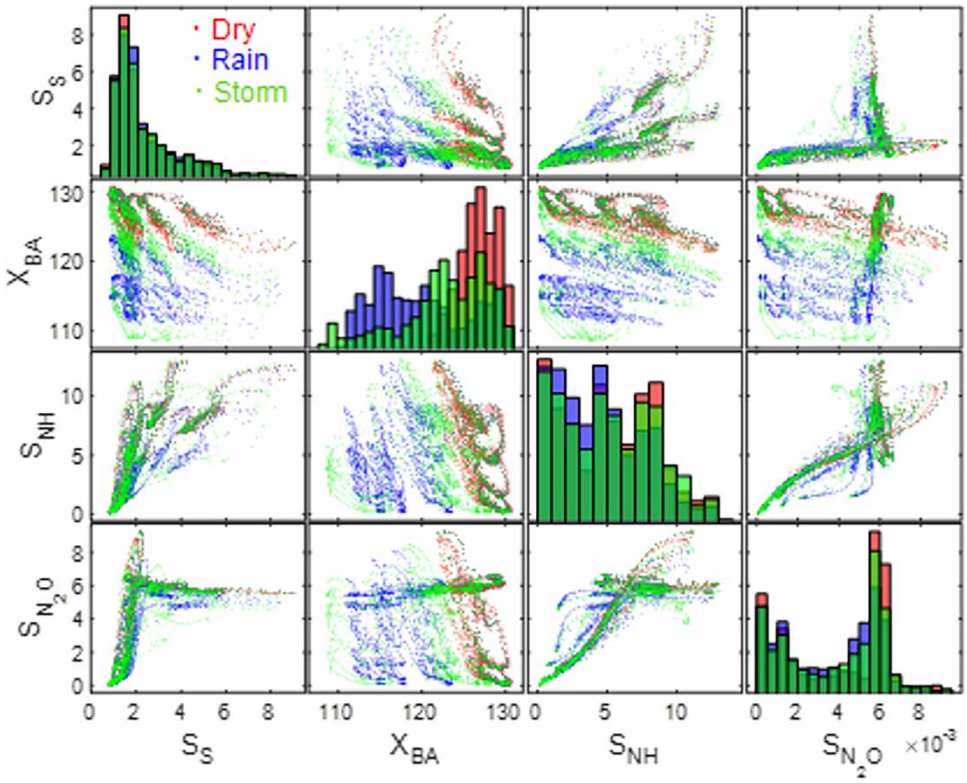

Figure 6 | Joint distribution of four components (days 8–12) in dry, rain and storm scenarios.

Wet-scenario performance degradation in dry-trained models arises not from structural limitations but from unseen hydraulic shock dynamics – specifically, rapid load dilution, substrate gradient destabilisation, and community functional resilience thresholds. This highlights a fundamental principle: training data diversity prioritise over volume as the primary determinant of model generalisability. Strategic integration of heterogeneous scenarios (dry + rain) during training would:

- Internalise hydraulic transient responses within constitutive model relationships
- Extend predictive capability to operationally adjacent regimes
- Enhance real-world applicability through covariate augmentation (e.g., rainfall intensity, flow variability)

These findings resonate with advancing wastewater digitalisation frameworks (Kurniawan *et al.* 2024), affirming that scenario-augmented training – encompassing the full spectrum of hydraulic regimes – is imperative for reliable NODE deployment in WWTPs. Practical implementation must integrate all-condition variability during training to ensure robust generalisation.

3.3. NODE model with reduced dimensions

3.3.1. Model simplification

In real-world WWTPs, comprehensive monitoring of all model components is often infeasible due to cost, sensor availability, or operational constraints. To address this, we explored a reduced-dimension NODE model by consolidating and excluding variables as follows:

1. Input consolidation:

- **Carbonaceous substrates:** Four components (soluble inert substrate S_I, soluble biodegradable substrate S_S, particulate inert substrate X_I, and particulate biodegradable substrate X_S) aggregated into total COD (T_{COD}).
- **Nitrogen species:** Ammonium (S_{NH_4}), soluble organic nitrogen (S_{ND}), and particulate organic nitrogen (X_{ND}) combined into total Kjeldahl nitrogen (TKN).

2. Retained critical variables:

DO (S_{O_2}), nitrate (S_{NO_2}), nitrite (S_{NO_3}), and ammonium (S_{NH_4})—key drivers of N_2O production (Pijuan & Zhao 2022)

3. Excluded variables:

Stable (e.g., biomass X_{BH}) or low-sensitivity parameters (e.g., alkalinity S_{ALK}).

This reduced-dimension model operates on seven input dimensions (vs. 18 in ASMG1), alleviating the burden on sensor configurations.

3.3.2. Predictive performance of reduced-dimension model

The reduced-dimension model was trained on dry weather data from days 0 to 6 and evaluated against the subsequent week (days 7–14). As illustrated in Figure 7, the model exhibited robust predictive performance across all five bioreactors, achieving remarkably low RMSE values despite dimensionality reduction.

Anoxic reactors R1 and R2 displayed relatively stable N_2O concentration profiles with minimal temporal variability, a pattern accurately captured by the model (RMSE: 0.00005 and 0.00004 mg N/L, respectively). In contrast, oxic reactors R3–R5 exhibited pronounced cyclical fluctuations in N_2O concentrations, reflecting influent load variability. Despite these complexities, the model effectively replicated both the amplitude and frequency of oscillations, achieving RMSE values of 0.00018, 0.00009, and 0.00008 mg/L for R3, R4, and R5, respectively. The aggregate RMSE of 0.000102 mg N/L further corroborates the model's capacity to maintain high predictive accuracy while retaining essential system dynamics through strategic parameter reduction.

3.3.3. Practical implications

The model's performance underscores three critical operational insights:

1. **Identification of critical drivers:** Retention of DO, nitrate, nitrite, and ammonium – key variables governing N_2O production kinetics – ensured preservation of prediction accuracy.
2. **Sensor feasibility:** Consolidation of inputs to total chemical oxygen demand (T_{COD}) and Kjeldahl nitrogen (TKN) aligns with standard WWTP monitoring practices, enabling deployment with accessible data.
3. **Balanced parameterisation:** Exclusion of slowly varying parameters (e.g., heterotrophic biomass, X_{BH}) had negligible impact under steady-state conditions. However, model recalibration may be warranted during periods of microbial community succession or process instability.

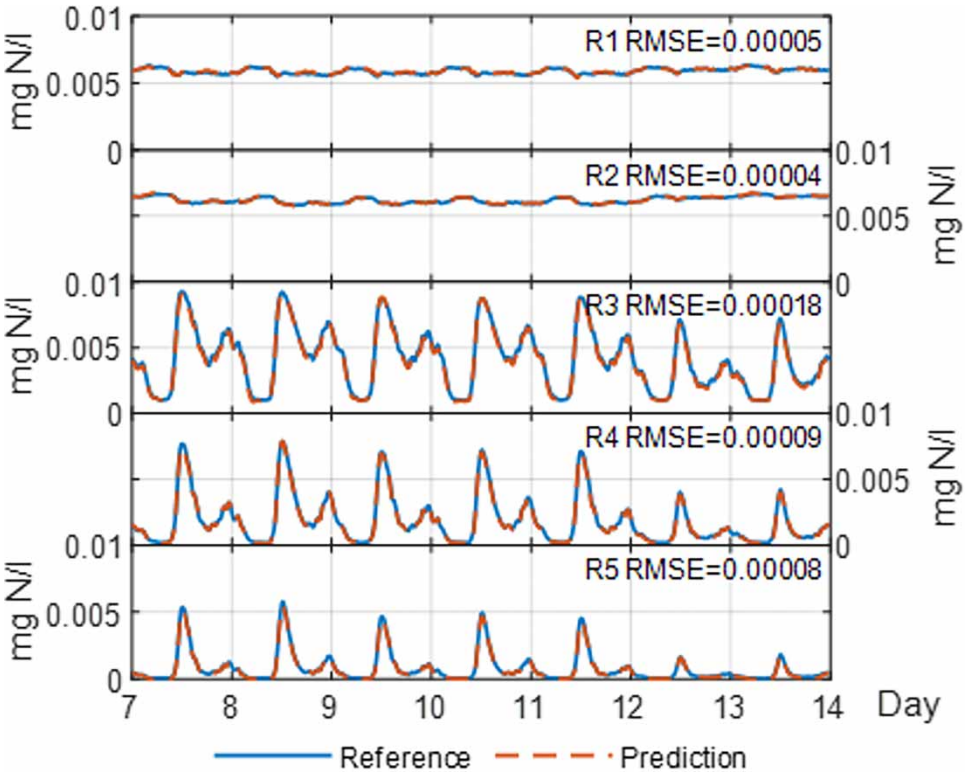

Figure 7 | Predicted vs. reference N_2O under dry weather conditions (days 7–14) for the reduced-dimension model trained on days 0–6.

This streamlined framework offers a pragmatic compromise between data insufficiency and model performance, demonstrating viability for N_2O monitoring in resource-constrained WWTPs. By prioritising critical biochemical drivers and leveraging routinely monitored parameters, the approach enhances operational feasibility without decreasing predictive accuracy.

4. CONCLUSION

This study establishes NODEs as a robust framework for data-driven modelling of N_2O dynamics in wastewater treatment, with four key advancements bridging methodological innovation and operational applicability:

1. **Stable convergence via paired normalisation and intrinsic dynamics separation**: Integrated paired normalisation/denormalisation layers ensured robust training convergence while maintaining learnt kinetic rate functions without distortion. By isolating intrinsic process dynamics from mixed observational data, this approach preserved physicochemical interpretability while mitigating gradient instability – a common challenge in differential equation-based architectures.
2. **High-fidelity prediction with scenario-diverse training:** When trained on datasets encompassing diverse weather conditions (dry, rain) and operational states (anoxic–oxic transitions, hydraulic peak and trough), the NODE framework consistently captured nonlinear N_2O dynamics, demonstrating resilience to process variability.
3. **Comprehensive operational regime coverage as prerequisite:** Model generalisability critically depends on training data that systematically encapsulate all key operational modes (e.g., diurnal cycles, stormwater inflows) and microbial regimes. Scenario-spanning training – particularly integrating dry and rainy conditions – proved essential to prevent context-specific underperformance, underscoring the necessity of holistic data curation for plant-wide deployment.
4. **Operational viability via reduced-dimension modelling:** By strategically retaining process-critical variables (e.g., ammonium, nitrite) while consolidating correlated inputs (e.g., T_{COD} and TKN), the reduced-dimension model maintained predictive fidelity with minimal infrastructure demands. This balance between parsimony and accuracy enables deployment in resource-limited settings, where high-frequency N_2O monitoring is otherwise constrained by sensor costs or availabilities.

To transition from simulation to real-world implementation, future work should prioritise mechanism-informed refinements and evaluation against long-term, full-scale datasets capturing seasonal microbial shifts. By harmonising data-driven adaptability with mechanistic interpretability, NODE-based approaches offer significant potential for advancing real-time N_2O mitigation, supporting sustainable wastewater management amid escalating climatic uncertainties.

ACKNOWLEDGEMENT

The work was supported by the CRONUS project (grant agreement ID: 101084405) funded by the European Union under Horizon Europe Research and Innovation Action scheme.

DATA AVAILABILITY STATEMENT

All relevant data are available from an online repository or repositories: https://github.com/Xiangjun-Huang/NODE_BSM1.git.

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

Abulimiti, A., Wang, X., Kang, J., Li, L., Wu, D., Li, Z., Piao, Y. & Ren, N. (2022) The trade-off between N_2O emission and energy saving through aeration control based on dynamic simulation of full-scale WWTP, *Water Research*, **223**, 118961. doi:10.1016/j.watres.2022.118961.

Alex, J., Benedetti, L., Copp, J., Gernaey, K. V., Jeppsson, U., Nopens, I., Pons, M., Steyer, J. & Vanrolleghem, P. A. (2008) *Benchmark Simulation Model no. 1 (BSM1)*. Lund, Sweden: Lund University. Available at: https://www.iea.lth.se/publications/Reports/LTH-IEA-7229.pdf.

Belia, E., Amerlinck, Y., Benedetti, L., Johnson, B., Sin, G., Vanrolleghem, P. A., Gernaey, K. V., Gillot, S., Neumann, M. B., Rieger, L., Shaw, A. & Villez, K. (2009) Wastewater treatment modelling: dealing with uncertainties, *Water Science and Technology*, **60** (8), 1929–1941. doi:10.2166/wst.2009.225.

Böttcher, L. & Asikis, T. (2022) Near-optimal control of dynamical systems with neural ordinary differential equations, *Machine Learning: Science and Technology*, **3** (4), 045004. doi:10.1088/2632-2153/ac92c3.

Cabello-Solorzano, K., Ortigosa de Araujo, I., Peña, M., Correia, L. & Tallón-Ballesteros, A. J. (2023) *The Impact of Data Normalization on the Accuracy of Machine Learning Algorithms: A Comparative Analysis*. Cham: Springer Nature Switzerland, p. 344.

Chen, R. T. Q., Rubanova, Y., Bettencourt, J. & Duvenaud, D. K. (2018) *Neural Ordinary Differential Equations*. Red Hook, NJ, USA: Curran Associates, Inc.

Flores-Alsina, X., Arnell, M., Amerlinck, Y., Corominas, L., Gernaey, K. V., Guo, L., Lindblom, E., Nopens, I., Porro, J., Shaw, A., Snip, L., Vanrolleghem, P. A. & Jeppsson, U. (2014) Balancing effluent quality, economic cost and greenhouse gas emissions during the evaluation of (plant-wide) control/operational strategies in WWTPs, *Science of The Total Environment*, **466–467**, 616–624. doi:10.1016/j.scitotenv.2013.07.046.

Guo, L. & Vanrolleghem, P. A. (2014) Calibration and validation of an activated sludge model for greenhouse gases No. 1 (ASMG1): prediction of temperature-dependent N_2O emission dynamics, *Bioprocess and Biosystems Engineering*, **37** (2), 151–163. doi:10.1007/s00449-013-0978-3.

Hairer, E. & Wanner, G. (1996) *Solving Ordinary Differential Equations II -Stiff and Differential Algebraic Problems*, 2nd edn. Berlin, Heidelberg, Germany: Springer.

Henze, M., Gujer, W., Mino, T. & Van Loosedrecht, M. (2000) *Activated Sludge Models ASM1, ASM2, ASM2d and ASM3*. London, UK: IWA Publishing.

Hiatt, W. C. & Leslie Grady, C. P. (2008) An updated process model for carbon oxidation, nitrification, and denitrification, *Water Environment Research*, **80** (11), 2145–2156. doi:10.2175/106143008x304776.

Huang, X., Kandris, K. & Katsou, E. (2025) Training stiff neural ordinary differential equations in data-driven wastewater process modelling, *Journal of Environmental Management*, **373**, 123870. doi:10.1016/j.jenvman.2024.123870.

Khalil, M., AlSayed, A., Liu, Y. & 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. doi:10.1016/j.watres.2023.120667.

Khalil, M., AlSayed, A., Elsayed, A., Sherif Zaghloul, M., Bell, K. Y., Al-Omari, A., Laqa Kakar, F., Houweling, D., Santoro, D., Porro, J. & Elbeshbishy, E. (2024) Advances in GHG emissions modelling for WRRFs: from state-of-the-art methods to full-scale applications, *Chemical Engineering Journal*, **494**, 153053. doi:10.1016/j.cej.2024.153053.

Kidger, P., Morrill, J., Foster, J. & Lyons, T. (2020) *Neural Controlled Differential Equations for Irregular Time Series*. *In*: *Advances in Neural Information Processing Systems 33 (NeurIPS 2020)*. San Diego, CA, USA: Neural Information Processing Systems Foundation.

Kim, S., Ji, W., Deng, S., Ma, Y. & Rackauckas, C. (2021) Stiff neural ordinary differential equations, *Chaos: An Interdisciplinary Journal of Nonlinear Science*, **31** (9), 093122.

Kurniawan, T. A., Mohyuddin, A., Casila, J. C. C., Sarangi, P. K., Al-Hazmi, H., Wibisono, Y., Kusworo, T. D., Khan, M. M. H. & Haddout, S. (2024) Digitalization for sustainable wastewater treatment: a way forward for promoting the UN SDG#6 'clean water and sanitation' towards carbon neutrality goals, *Discover Water*, **4** (1), 71. doi:10.1007/s43832-024-00134-5.

Liu, Y., Zhang, S., Qiu, L. & Zhang, Y. (2016) *Progress and Prospect of Research on N_2O Dynamics Model in Wastewater Biological Treatment Process*. In: *Proceedings of the 2015 4th International Conference on Sustainable Energy and Environmental Engineering*. Paris, France: Atlantis Press, pp. 464–467.

Lv, Y., Zhang, S., Xie, K., Liu, G., Qiu, L., Liu, Y. & Zhang, Y. (2022) Establishment of nitrous oxide (N_2O) dynamics model based on ASM3 model during biological nitrogen removal via nitrification, *Environmental Technology*, **43** (8), 1170–1180. doi:10.1080/09593330.2020.1822447.

Mampaey, K. E., Beuckels, B., Kampschreur, M. J., Kleerebezem, R., Van Loosdrecht, M. C. M. & Volcke, E. I. P. (2013) Modelling nitrous and nitric oxide emissions by autotrophic ammonia-oxidizing bacteria, *Environmental Technology (United Kingdom)*, **34** (12), 1555–1566. doi:10.1080/09593330.2012.758666.

Ni, B., Pan, Y., Van Den Akker, B., Ye, L. & Yuan, Z. (2015) Full-Scale modeling explaining large spatial variations of nitrous oxide fluxes in a step-feed plug-Flow wastewater treatment reactor, *Environmental Science & Technology*, **49** (15), 9176–9184. doi:10.1021/acs.est.5b02038.

Pijuan, M. & Zhao, Y. (2022) Full-scale source, mechanisms and factors affecting nitrous oxide emissions. In: Ye, L., Porro, J. & Nopens, I. (eds) *Quantification and Modelling of Fugitive Greenhouse Gas Emissions From Urban Water Systems*. London, UK: IWA Publishing.

Porro, J., Vasilaki, V., Bellandi, G. & Katsou, E. (2022) Knowledge-based and data-driven approaches for assessing greenhouse gas emissions from wastewater systems. In: Yev, L., Porro, J. & Nopens, I. (eds) *Quantification and Modelling of Fugitive Greenhouse Gas Emissions From Urban Water Systems*. London, UK: IWA Publishing.

Roesch, E., Rackauckas, C. & Stumpf, M. P. H. (2021) Collocation based training of neural ordinary differential equations, *Statistical Applications in Genetics and Molecular Biology*, **20** (2), 37–49. doi:10.1515/sagmb-2020-0025.

Soler-Jofra, A., Stevens, B., Hoekstra, M., Picioreanu, C., Sorokin, D., van Loosdrecht, M. C. M. & Pérez, J. (2016) Importance of abiotic hydroxylamine conversion on nitrous oxide emissions during nitritation of reject water, *Chemical Engineering Journal*, **287**, 720–726. doi:10.1016/j.cej.2015.11.073.

Spérandio, M., Lang, L., Sabba, F., Nerenberg, R., Vanrolleghem, P., Domingo-Félez, C., Smets, B. F., Duan, H., Ni, B. & Yuan, Z. (2022) Modelling N2O production and emissions. In: Ye, L., Porro, J. & Nopens, I. (eds) *Quantification and Modelling of Fugitive Greenhouse Gas Emissions From Urban Water Systems*. London, UK: IWA Publishing.

Su, Q., Domingo-Félez, C., Jensen, M. M. & Smets, B. F. (2019) Abiotic nitrous oxide (N2O) production is strongly pH dependent, but contributes little to overall N2O emissions in biological nitrogen removal systems, *Environmental Science & Technology*, **53** (7), 3508–3516. doi:10.1021/acs.est.8b06193.

The MathWorks Inc (2024) *MATLAB Version: 24.1.0.2628055 (R2024a) Update 4* Available at: https://www.mathworks.com.

First received 1 May 2025; accepted in revised form 17 December 2025. Available online 17 February 2026