

# Nitrifier activity as a cross-facility screening indicator of N<sub>2</sub>O emission potential in full-scale wastewater treatment facilities

Seongbong Heo<sup>a</sup>, Hyeongju Park<sup>a,b</sup>, Chamsol Kwon<sup>a</sup>, Young Mo Kim<sup>a,b,\*</sup>

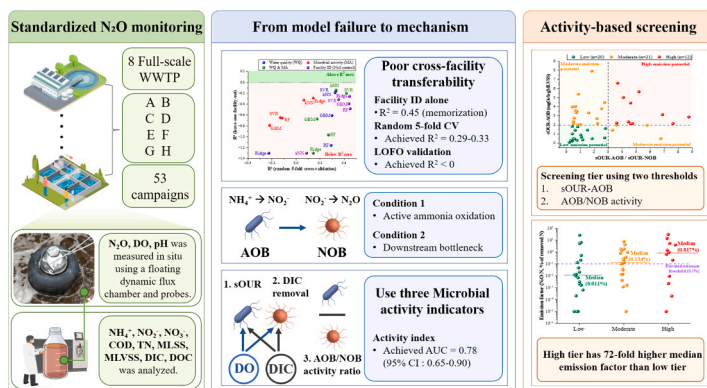
<sup>a</sup> Department of Civil and Environmental Engineering, Hanyang University, Seongdong-gu, Seoul, 04763, Republic of Korea

<sup>b</sup> Developing a Team Response Using Digital Construction to Mitigate Disasters Related to Climate Change (BK21 FOUR), Hanyang University, Seongdong-gu, Seoul, 04763, Republic of Korea

## HIGHLIGHTS

- N<sub>2</sub>O emission factors varied by four orders of magnitude across eight full-scale sites.
- Machine-learning models did not transfer to facilities excluded from training.
- Facility identity alone outperformed water-quality models under random splitting.
- Activity-based indicators identified elevated N<sub>2</sub>O emission states (AUC 0.78).
- Two screening thresholds separated tiers differing 72-fold in median emission.

## GRAPHICAL ABSTRACT



## ARTICLE INFO

### Keywords:

Aerobic treatment  
Emission factor  
Machine learning  
Nitrous oxide  
Respirometry  
Wastewater treatment

## ABSTRACT

Nitrous oxide (N<sub>2</sub>O) dominates the process carbon footprint of biological nitrogen removal, but direct monitoring remains impractical. N<sub>2</sub>O emissions were measured at seven municipal wastewater treatment plants and one landfill leachate facility over 53 campaigns using a common flux-chamber protocol, together with water quality and the respirometric activities of ammonia- and nitrite-oxidizing bacteria (sOUR-AOB, sOUR-NOB). Quantified removal-based emission factors ranged from 0.0002% to 31.5% of removed nitrogen and showed little relationship with process configuration. The municipal median on an influent-nitrogen basis was approximately 160-fold lower than the IPCC 2019 default. Machine-learning models reached R<sup>2</sup> up to 0.33 under random cross-validation but failed across all 15 algorithm-feature-set combinations under leave-one-facility-out validation. Facility identity alone reached R<sup>2</sup> = 0.45, indicating facility-specific rather than transferable relationships. In contrast, an activity index based on sOUR-AOB, the AOB/NOB activity ratio, and inorganic carbon removal identified elevated-emission states (AUC 0.78), and the three-indicator block remained significant after facility adjustment ( $p = 0.006$ ). Thresholds of sOUR-AOB  $\geq 2.0$  mg O<sub>2</sub>/g MLVSS•h and an AOB/NOB ratio  $\geq 3.0$  separated tiers with a 72-fold difference in median emission factor. Measuring the top 25% of campaigns captured 42% of emissions, supporting respirometric screening to prioritize N<sub>2</sub>O measurement and mitigation.

\* Corresponding author at: Department of Civil and Environmental Engineering, Hanyang University, Seongdong-gu, Seoul 04763, Republic of Korea.

E-mail address: [youngmo@hanyang.ac.kr](mailto:youngmo@hanyang.ac.kr) (Y.M. Kim).

<https://doi.org/10.1016/j.dwt.2026.101944>

Received 27 August 2026; Received in revised form 11 September 2026; Accepted 13 September 2026

Available online 14 September 2026

1944-3986/© 2026 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY-NC license (<http://creativecommons.org/licenses/by-nc/4.0/>).

## 1. Introduction

Nitrous oxide ( $\text{N}_2\text{O}$ ) is a greenhouse gas with a 100-year global warming potential approximately 273 times that of  $\text{CO}_2$  and is a major contributor to the direct process carbon footprint of biological nitrogen removal [1,2]. As greenhouse-gas accounting expands across the water sector, utilities and national inventory compilers are increasingly required to quantify  $\text{N}_2\text{O}$  emissions from wastewater treatment plants (WWTPs). The IPCC 2019 Refinement assigns a Tier 1 emission factor of 1.6% of influent nitrogen to centralized aerobic wastewater treatment, with a reported range of 0.016–4.5% [3]. Direct  $\text{N}_2\text{O}$  quantification at WWTPs is resource-intensive because it requires gas-phase analyzers, flux chambers or off-gas capture, and repeated monitoring to capture strong temporal variability [4,5]. Consequently, only a small fraction of WWTPs worldwide have been monitored, and most inventories still rely on default emission factors [6,7].

$\text{N}_2\text{O}$  formation during biological nitrogen removal mainly occurs through hydroxylamine oxidation, nitrifier denitrification, and heterotrophic denitrification [1,8–10]. Incomplete hydroxylamine oxidation can produce  $\text{N}_2\text{O}$  during rapid ammonia oxidation, whereas nitrifier denitrification is promoted by low dissolved oxygen (DO) and nitrite accumulation. During heterotrophic denitrification,  $\text{N}_2\text{O}$  is an obligate intermediate, and its reduction to  $\text{N}_2$  can be limited by insufficient electron donors, low pH, oxygen intrusion, and inhibition by nitrite or free nitrous acid [11,12]. These conditions can therefore increase  $\text{N}_2\text{O}$  production while limiting its further reduction to  $\text{N}_2$ .

Elevated  $\text{N}_2\text{O}$  emissions generally require two conditions. First, active ammonia oxidation generates reactive nitrogen intermediates. Second, a downstream bottleneck limits their complete conversion to nitrate or  $\text{N}_2$ . Respirometric measurements can directly assess these two conditions. The specific oxygen uptake rates of ammonia- and nitrite-oxidizing bacteria (sOUR-AOB and sOUR-NOB) quantify the potential activities of the two nitrifying guilds under excess-substrate conditions, while the sOUR-AOB/sOUR-NOB ratio indicates the balance between nitrite-production and nitrite-oxidation capacity [13]. These assays require only conventional laboratory respirometry and are more directly related to nitrification activity than bulk water-quality concentrations. However, they have rarely been evaluated as indicators of  $\text{N}_2\text{O}$  emission potential in full-scale surveys [5,14].

Full-scale studies have shown large differences in  $\text{N}_2\text{O}$  emissions among facilities and over time. Multi-plant surveys in the United States and Australia reported approximately two orders of magnitude variation in emission factors among facilities with similar process configurations [13, 15]. Long-term monitoring has also shown strong seasonal and diurnal variation within individual plants [4]. Country-scale monitoring further demonstrated that a small number of facilities and operating periods can account for a large share of total emissions [16]. Reviews have identified this variability, inconsistent operational drivers, and differences in measurement protocols as major barriers to comparison across studies [5]. Standardized multi-facility surveys remain limited, particularly for high-strength nitrogen streams. Landfill leachate treatment systems, for example, have reported  $\text{N}_2\text{O}$  emission factors of 8.9–11.9% of removed nitrogen [17], but they are rarely evaluated alongside municipal WWTPs using the same measurement protocol.

Machine learning (ML) has been investigated as an alternative to costly direct  $\text{N}_2\text{O}$  monitoring. Random forest and support vector machine models have reported  $R^2$  values of 0.7–0.9 within individual plants [18], while recurrent and long short-term memory models have reported within-plant accuracies above 0.95 for dynamic  $\text{N}_2\text{O}$  emission time series [19]. ML has also been applied to gap filling, soft sensing, aeration control, inventory estimation, and hybrid mechanistic-data-driven modeling [20–22]. However, most studies have focused on individual facilities and evaluated model performance using data from the same plant for both model development and validation [23,24]. Such validation demonstrates within-facility prediction but does not establish whether a model can be transferred to a facility that was not represented in the training data.

Three practical limitations therefore remain. First, direct  $\text{N}_2\text{O}$  monitoring requires specialized analyzers, flux chambers or off-gas collection systems, and substantial field effort. Most WWTPs remain unmeasured, and inventories therefore depend on default emission factors that may not represent individual facilities [6]. Second, relationships between  $\text{N}_2\text{O}$  emissions and routine variables such as DO, ammonium, and nitrite vary among facilities. This limits the use of fixed water-quality thresholds across different treatment systems. Third, the cross-facility transferability of ML models remains largely untested. Within-plant validation can overestimate model generalizability because data from the same facility are included in both model development and evaluation. If measured  $\text{N}_2\text{O}$  data are required to recalibrate a model at every new facility, ML cannot replace direct monitoring where monitoring data are unavailable. A practical method for screening and ranking  $\text{N}_2\text{O}$  emission potential across unmeasured facilities is therefore still needed.

This study combined standardized multi-facility  $\text{N}_2\text{O}$  measurements, nitrifier activity assays, and explicit cross-facility model validation.  $\text{N}_2\text{O}$  emissions were measured during 53 monthly campaigns at seven full-scale municipal WWTPs representing five process configurations and one landfill leachate treatment facility, using the same flux-chamber protocol, analytical platform, and field team, with sOUR-AOB and sOUR-NOB measured in parallel. Machine-learning models based on routine water-quality variables were tested on unseen facilities using leave-one-facility-out validation, with facility-identity models evaluated under random cross-validation as a benchmark for facility-specific effects. The aim was to quantify  $\text{N}_2\text{O}$  emission factors across the eight facilities and compare them with the IPCC default, determine whether ML models generalize to unseen facilities, and evaluate activity-based indicators derived from sOUR-AOB, the AOB/NOB activity ratio, and dissolved inorganic carbon removal for targeted  $\text{N}_2\text{O}$  screening, measurement, and mitigation.

## 2. Materials and methods

### 2.1. Full-scale WWTP $\text{N}_2\text{O}$ monitoring

Eight full-scale facilities located in two provinces (G Province and J Province) of the Republic of Korea were monitored. The facilities spanned three orders of magnitude in treatment capacity and represented the major nitrogen-removal configurations operated in the region (Table 1). Seven were municipal WWTPs, including three anaerobic-anoxic-oxic (A2O) plants (WWTP-A, 600,000  $\text{m}^3/\text{d}$ ; WWTP-B, 60,000  $\text{m}^3/\text{d}$ ; WWTP-G, 11,000  $\text{m}^3/\text{d}$ ), one modified Ludzack-Ettinger (MLE) train (WWTP-C, 60,000  $\text{m}^3/\text{d}$ ), one membrane bioreactor (MBR; WWTP-D), one oxidation ditch (WWTP-E), and one sequencing batch reactor (SBR; WWTP-F). WWTP-C was located at the same site as WWTP-B and received the same influent, allowing direct comparison of the two process configurations under the same influent and climatic conditions. Hydraulic retention times ranged from 4.0 to 32.6 h, while solids retention times ranged from 4.4 to 30 d at facilities where sludge age was actively controlled (Table 1).

At continuous-flow plants, measurements were conducted separately in the anaerobic, anoxic, and aerobic zones. At the SBR, the aeration period of each 140–160 min cycle was classified as aerobic and the settling period as anoxic. Within each cycle, the sequence comprised pre-mixing (5 min), aeration (60 min), post-mixing (5 min), settling (40 min), and decant (30 min) in the 140 min configuration, and in the 160 min configuration the aeration period equaled or exceeded the settling period; the pre-mixing, post-mixing, and decant periods were not assigned to either condition. The chamber was positioned at the center of the tank, and measurements were averaged over two to three cycles. The cycle-averaged fluxes of the aeration and settling periods were weighted by the corresponding phase durations within the cycle and applied to the tank surface area when the facility-level emission was computed with Eq. (2). The eighth facility, WWTP-H, treated high-strength landfill leachate using an MLE-type process. Influent  $\text{NH}_4^+\text{-N}$  ranged from 120 to 1090 mg N/L, with

nitrogen loads comparable to those of the largest municipal WWTP. The process was housed underground, and mixed-liquor temperature remained approximately 19–34 °C throughout the year. Measurements were conducted at one anoxic and two aerobic locations along the treatment line. Throughout this study, a campaign refers to one monthly monitoring event at one facility, comprising the chamber measurements at all monitored zones together with the accompanying water-quality sampling and respirometric assays; each campaign yields one observation. Monthly monitoring comprised 4–13 campaigns per facility and yielded 53 campaigns. The facility characteristics and monitoring coverage are summarized in Table 1.

## 2.2. N<sub>2</sub>O flux measurement and emission factors

Gaseous N<sub>2</sub>O was measured using a dynamic flux chamber designed according to US EPA recommendations [25,26]. The hemispherical chamber had a volume of 30–40 L, a diameter of 0.41 m, and an enclosed water-surface area of 0.132 m<sup>2</sup> (Fig. 1b). The chamber floated on the reactor surface with its lower rim submerged and was anchored on both sides. Ballast weights were added when internal pressure caused gas loss at the water seal (Fig. 1c). Nitrogen sweep gas was supplied at 600 cm<sup>3</sup>/min. Chamber off-gas passed through inert PTFE tubing and 0.2 µm cartridge filters to a T320 N<sub>2</sub>O analyzer (Teledyne API, USA; sample flow rate, 770 cm<sup>3</sup>/min), which continuously recorded N<sub>2</sub>O concentration, exhaust-gas temperature, and flow rate (Fig. 1a). Measurements began after the N<sub>2</sub>O signal stabilized and continued for 180 min at the continuous-flow facilities; at the SBR, measurements continued over two to three complete cycles (Section 2.1). Sweep-gas flow and gas temperature were recorded before and after each measurement to constrain sampling uncertainty.

Gas flow from the enclosed water surface was quantified using a tracer method following US EPA protocols [27]. Helium was injected into the chamber at a constant rate of 100 cm<sup>3</sup>/min. The mixed exhaust gas was collected in foil-lined Tedlar bags, and the helium concentration was determined by gas chromatography with thermal conductivity detection (GC-TCD) using argon as the carrier gas. These measurements were used to determine the total chamber gas flow (Fig. 1a). The areal N<sub>2</sub>O flux from each zone was calculated as

$$F = (C \times Q) / A_{fc} \quad (\text{eq1})$$

where F is the N<sub>2</sub>O flux (mass/area•time), C is the N<sub>2</sub>O concentration in the chamber off-gas, Q is the total gas flow through the chamber, and A<sub>fc</sub> is the water-surface area enclosed by the chamber. Emissions from individual zones were scaled to their corresponding reactor surface areas and summed to estimate facility-level emissions:

$$E = \sum (F_i \times A_i), \quad i = 1 \dots n \quad (\text{eq2})$$

where E is the facility-level N<sub>2</sub>O emission (kg N<sub>2</sub>O-N/d) and A<sub>i</sub> is the surface area of zone i. Emission factors were calculated on influent-load and nitrogen-removal bases:

$$EF_{\text{load}} = 100 \times E / L_N, \quad EF_{\text{removal}} = 100 \times E / R_N \quad (\text{eq3})$$

where L<sub>N</sub> and R<sub>N</sub> are the daily influent nitrogen load and nitrogen removal (kg N/d), respectively. The load-based emission factor follows the IPCC Tier 1 convention and was used for comparison with the default emission factor. The removal-based emission factor accounts for differences in nitrogen removal and was used for cross-facility comparisons. The basis is stated explicitly in every figure and table in which an emission factor appears.

## 2.3. Water quality analysis

During each campaign, samples were collected from the influent, effluent, and each bioreactor zone three times per day and analyzed according to Standard Methods [28]. Nitrogen species (NH<sub>4</sub><sup>+</sup>-N, NO<sub>2</sub><sup>-</sup>-N, and NO<sub>3</sub><sup>-</sup>-N) were quantified by ion chromatography; soluble COD and TN were measured using standard colorimetric procedures; and mixed liquor suspended solids (MLSS) and mixed liquor volatile suspended solids (MLVSS) were determined gravimetrically. Dissolved inorganic carbon (DIC) and dissolved organic carbon (DOC) were measured using a TOC analyzer. DIC removal across the bioreactor was calculated as the relative decrease from influent to effluent and used as an integrative measure of realized autotrophic conversion and associated alkalinity demand [29]. DO, pH, and temperature were measured in situ using portable probes. Aeration rates were obtained from plant records as daily averages.

## 2.4. Respirometric activity assays

The specific oxygen uptake rates of AOB and NOB, measured in batch assays under excess-substrate conditions, were used to quantify the potential activities of the two nitrifying guilds [30]; these rates represent capacity rather than in situ nitrification rates. Mixed liquor collected from the aerobic zone was analyzed by batch respirometry using the selective-inhibition method of Ginestet et al. [30]. Total oxygen uptake was first measured under excess-substrate conditions. Sodium azide was then added to inhibit nitrite oxidation, followed by allylthiourea at 86 µM (10 mg/L) to inhibit ammonia oxidation. Guild-specific oxygen uptake rates were calculated from the differences between successive measurements and normalized to MLVSS (mg O<sub>2</sub>/g MLVSS•h). The sOUR-AOB/sOUR-NOB ratio, hereafter referred to as the AOB/NOB activity ratio, was used to indicate the balance between nitrite-production and nitrite-oxidation capacity [31].

## 2.5. Data analysis

All analyses were performed in Python 3.10 using scikit-learn, statsmodels, and SciPy.

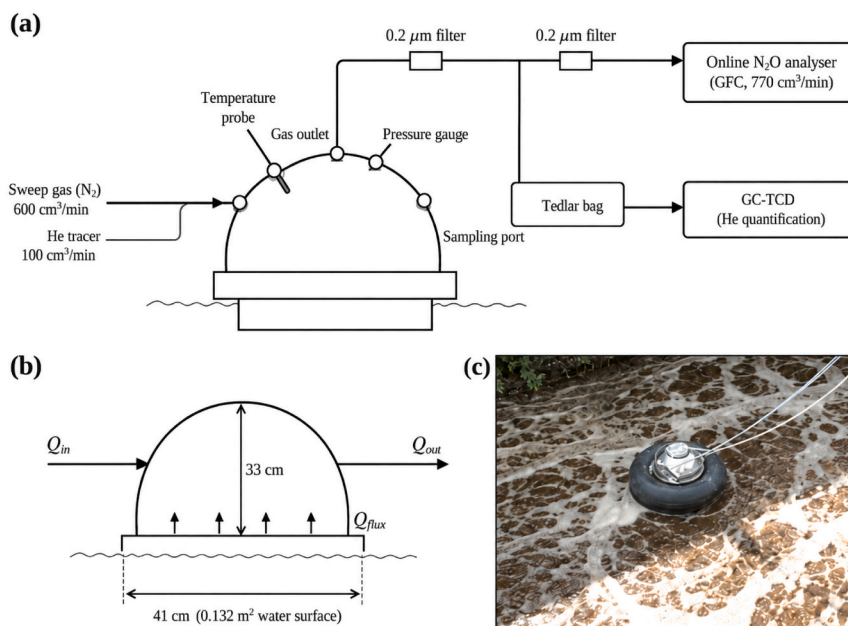
### 2.5.1. Dataset preparation

All analyses were based on the 53 monitoring campaigns. Each campaign included 14 routine water-quality and loading variables, four activity-related variables (sOUR-AOB, sOUR-NOB, the AOB/NOB

**Table 1**  
Characteristics of the monitored facilities.

| Code   | Process         | Capacity (t/d) | Aeration (m <sup>3</sup> /h) | HRT (h) | SRT (d) | Number of monitoring campaigns (n) |
|--------|-----------------|----------------|------------------------------|---------|---------|------------------------------------|
| WWTP-A | A2O             | 600,000        | 3333                         | 6.0     | N/A     | 5                                  |
| WWTP-B | A2O             | 60,000         | 1134                         | 7.9     | 12.1    | 8                                  |
| WWTP-C | MLE             | 60,000         | 1470                         | 11.2    | 18.1    | 6                                  |
| WWTP-D | MBR             | 6000           | 3600                         | 4.0     | 30.0    | 5                                  |
| WWTP-E | Oxidation ditch | 11,000         | N/A                          | 24.0    | N/A     | 4                                  |
| WWTP-F | SBR             | 4500           | 2358                         | 32.6    | N/A     | 8                                  |
| WWTP-G | A2O             | 11,000         | 720                          | 5.8     | 4.4     | 4                                  |
| WWTP-H | MLE             | 500            | 918                          | 4.8     | N/A     | 13                                 |

N/A: not applicable to the process configuration or not available from facility operating records; SRT is reported only for facilities where sludge age was actively controlled



**Fig. 1.** Dynamic flux chamber used in all campaigns. (a) Gas handling and analysis system, (b) chamber dimensions and gas flux across the enclosed water surface, and (c) field deployment in an aerated zone.

activity ratio, and DIC removal), gaseous  $\text{N}_2\text{O}$  concentration, and two emission factors. Emission factors were  $\log_{10}$ -transformed, and values below the analyzer detection limit were assigned a value of  $10^{-5}\%$ . This substitution serves only to place the five censored campaigns on the logarithmic scale. Substitutes of  $10^{-4}\%$  and  $10^{-6}\%$  remain below the lowest quantified removal-based value, so every ranking is preserved: the area under the receiver operating characteristic curve (AUC), the Spearman correlations, and the tier classification are unchanged, the quantified ranges are unaffected, and the only facility medians that move are those in which a censored campaign enters the median, namely the medians of WWTP-E (two of four campaigns censored) and, at  $10^{-4}\%$ , the load-based median of WWTP-A, each by less than 0.0001 %age points. The regression and mixed-effects estimates, which use the substituted values directly, varied slightly with the substitute, but no conclusion changed: cross-facility transfer failed for every substitute, and the activity indicators remained jointly significant after accounting for facility identity. One  $\text{NH}_4^+\text{-N}$  value at WWTP-G exceeded the total nitrogen concentration of the same sample and was excluded as a recording error. An elevated-emission state was defined as a removal-based emission factor  $\geq 0.1\%$ . This threshold approximates the median removal-based emission factor across the 53 campaigns (0.105%; Table S1), classifying 27 campaigns as elevated and yielding two groups of similar size for AUC estimation. It follows the median-based approach used for the operational sOUR-AOB and DIC removal cutoffs (Section 2.5.3) and was not optimized to maximize discrimination. For context, at the median municipal nitrogen-removal efficiency of 57%, this threshold corresponds to approximately 0.06% of influent nitrogen, within the IPCC 2019 reported range (0.016–4.5%) but well below the Tier 1 default of 1.6% [3]. The threshold therefore serves as a dataset-specific classification criterion. Sensitivity to this choice was assessed using alternative thresholds of 0.05, 0.2, and 0.5% (Table S6). Cohort-level medians were calculated over all campaigns, with non-detect campaigns at the floor value. For comparison with the IPCC Tier 1 default, the municipal median is the median of the 40 campaign-level load-based emission factors; the median of the seven facility-level medians and the nitrogen-load-weighted aggregate emission factor (the sum of facility-level emissions divided by the sum of influent nitrogen loads) are reported as alternative estimates (Table S1). These aggregates summarize the monitored campaigns and are not

annual inventory estimates. In the main text, emission factors are reported to two significant figures, to three where a value enters a stated ratio, and to one below 0.001% where the measurement precision does not support more digits; ratios were calculated from unrounded values, and the [supplementary tables](#) retain additional digits for comparison and calculation.

#### 2.5.2. Machine-learning transferability experiment

Cross-facility model transferability was evaluated for prediction of the  $\log_{10}$ -transformed removal-based emission factor using five algorithms, four feature sets, and two validation schemes. The algorithms were 1) random forest, 2) gradient boosting, 3) support vector regression, 4) k-nearest neighbors, and 5) ridge regression. These algorithms were selected to span distinct inductive biases, including ensemble tree-based (random forest, gradient boosting), kernel-based (support vector regression), instance-based (k-nearest neighbors), and penalized linear (ridge regression) approaches, so that a consistent outcome across all five would indicate a property of the data rather than of a particular model. All algorithms were implemented in scikit-learn 1.3.2 with default hyperparameters except where noted (Table S4): random forest (500 trees, fixed before the analysis, unrestricted depth, all features considered at each split, minimum leaf size of 1), gradient boosting (100 trees, learning rate 0.1, maximum depth 3, squared-error loss), k-nearest neighbors ( $k = 5$ , uniform weights, Euclidean distance on standardized features), and ridge regression ( $\alpha = 1.0$ ). For support vector regression, a radial basis function kernel with  $C = 10$ ,  $\epsilon = 0.1$ , and  $\gamma = \text{'scale'}$  was used;  $C$  was fixed before the analysis, and leave-one-facility-out  $R^2$  remained negative for  $C$  between 0.1 and 100 (Table S4). Hyperparameters were not tuned, so that the two validation schemes were compared at identical model settings.

The feature sets included water quality only, activity-related variables only, both feature groups, and facility-identity dummy variables only. Model performance was evaluated using repeated random 5-fold cross-validation, averaged over five repeats, and leave-one-facility-out (LOFO) validation, in which all observations from one facility were withheld for testing. LOFO validation reproduces the intended application of the models at facilities without prior  $\text{N}_2\text{O}$  measurements, whereas random splitting allows observations from the same facility to appear in both the training and test sets. The facility-identity feature set contained no physicochemical information and was evaluated under random cross-

validation only, as a benchmark for facility-specific effects; under LOFO validation a held-out facility has no corresponding dummy variable in the training set, so this feature set cannot inform prediction; the corresponding LOFO entries in Table S4 are reported for completeness and are not among the 15 algorithm-feature-set combinations evaluated for transferability. Continuous predictors were standardized to zero mean and unit variance and missing values were imputed with the median, both fitted within each training partition and applied to the corresponding test partition.

### 2.5.3. Indicator screening and statistical analysis

The screening performance of training-free indicators was evaluated using the AUC for identifying elevated-emission states, with 95% confidence intervals obtained from 2000 bootstrap resamples. The activity index combined sOUR-AOB, the AOB/NOB activity ratio, and DIC removal. Each indicator was ranked in ascending order across all 53 campaigns, so that higher values, corresponding to greater potential ammonia-oxidizing activity, a larger ratio of nitrite-production to nitrite-oxidation capacity, and greater realized autotrophic conversion, received higher ranks (1–53). Tied values, which occurred only for DIC removal, were assigned the mean of the ranks they would otherwise occupy.

The activity index for campaign  $j$  was calculated as the unweighted sum of its three component ranks:

$$AI_j = r_{1j} + r_{2j} + r_{3j} \quad (\text{eq4})$$

where  $AI_j$  is the activity index for campaign  $j$ , and  $r_{1j}$ ,  $r_{2j}$ , and  $r_{3j}$  are its ranks for sOUR-AOB, the AOB/NOB activity ratio, and DIC removal, respectively. The ranks were not normalized, giving a possible index range of 3–159 across the 53 campaigns. No weights were fitted, and all three indicators had complete data. The subsequent analyses depended only on the ordering of the index, so rescaling the summed index would not affect the results. The two-indicator composite (Table 3) was calculated using the same approach, summing the ranks for sOUR-AOB and the AOB/NOB activity ratio.

Ranking across the full dataset incorporates information from the facility being evaluated. To assess the influence of this information, the index was also recalculated using a leave-one-facility-out (LOFO) procedure. For each held-out campaign, each indicator was converted to a percentile relative to the campaigns from the remaining seven facilities, defined as the proportion of reference values less than or equal to the held-out value. The three percentiles were then summed without weighting to obtain the LOFO activity index (Table S9).

Two thresholds were used for screening without reference N<sub>2</sub>O measurements. First, sOUR-AOB  $\geq 2.0$  mg O<sub>2</sub>/g MLVSS•h was used to indicate sufficient potential ammonia-oxidizing activity. Second, an AOB/NOB activity ratio  $\geq 3.0$  was used to indicate a potential nitrite-oxidation bottleneck. This threshold represents stoichiometric parity and is therefore independent of the present dataset (Section 3.4). Campaigns meeting both criteria were classified as high emission potential, those meeting one criterion as moderate, and those meeting neither criterion as low. A three-criterion score additionally including DIC removal  $\geq 35\%$  was also evaluated. The sOUR-AOB and DIC removal cutoffs were operational thresholds rather than mechanistic constants. They correspond to the cohort medians of 2.11 mg O<sub>2</sub>/g MLVSS•h and 34.3%, rounded to practical operational cutoffs, and were not optimized for discrimination. In comparison, cutoffs maximizing the Youden index were 1.25 mg O<sub>2</sub>/g MLVSS•h for sOUR-AOB and 30% for DIC removal, resulting in only a marginally higher AUC (0.76 versus 0.73; Table S10). Prospective validation at independent facilities is therefore required before these operational thresholds are applied more broadly.

Linear mixed-effects models with facility as a random intercept were used to test whether the activity indicators explained emission variability beyond facility identity. Continuous predictors in the mixed-effects models were standardized to zero mean and unit variance, and model improvement was assessed using likelihood-ratio tests. Emission capture was evaluated by ranking campaigns according to each indicator and

calculating the cumulative estimated N<sub>2</sub>O emission. The absolute N<sub>2</sub>O emission for each campaign was taken as the facility-level emission calculated using Eq. (2) (kg N<sub>2</sub>O-N/d). For indicators with tied values, capture curves were averaged over 600 random tie orderings.

## 3. Results and discussion

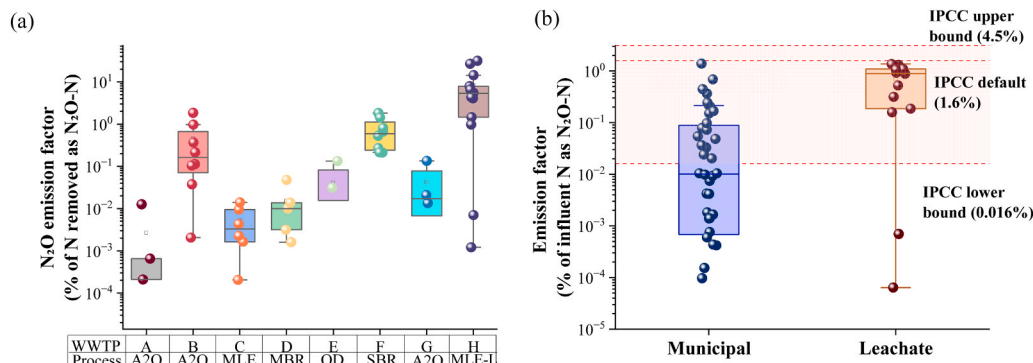
### 3.1. N<sub>2</sub>O emission factors across eight full-scale facilities

Removal-based emission factors differed by more than four orders of magnitude across the eight facilities despite the use of the same flux-chamber protocol (Fig. 2a, Table S1). Facility-level median values ranged from 0.0002% of removed nitrogen at WWTP-A to 5.3% at the leachate facility (WWTP-H), with campaign-level emission factors differing significantly between the leachate facility and the municipal cohort (Mann-Whitney  $U$  test,  $p < 0.001$ ). The seven municipal facilities alone ranged from 0.0002% to 0.59%. Similar variability has been reported in previous full-scale surveys [4,13,15], although differences in measurement methods may have contributed to the reported ranges. In this study, all facilities were monitored using the same protocol, analytical platform, and field team. The large differences therefore more likely reflect facility-specific operating conditions than differences in measurement methods.

Nitrogen-removal performance provides context for the observed differences in N<sub>2</sub>O emissions. Median nitrogen-removal efficiencies were 66–70% at the three largest municipal plants, approximately 20–42% at the smaller facilities, and 13% at the leachate facility. Volumetric nitrogen loading rates ranged from 0.004 to 0.091 kg N/m<sup>3</sup>•d, and specific nitrogen-removal rates ranged from 1.1 to 4.9 g N/kg MLVSS•d (Tables S2 and S3). In the pooled dataset (all 53 campaigns from both municipal and leachate), the removal-based emission factor was inversely correlated with nitrogen-removal efficiency (Spearman  $\rho = -0.39$ ,  $p = 0.004$ ; Fig. S1). However, this inverse relationship was not observed within the municipal cohort or the leachate facility alone. It therefore arises from the contrast between the two cohorts, in which the leachate facility combined the lowest removal efficiency with the highest emission factor, rather than from a within-facility relationship between removal performance and emission. Therefore, removal efficiency distinguishes high- from low-emitting facilities but does not track emission within a facility over time.

Process configuration showed little relationship with N<sub>2</sub>O emissions. The three A2O facilities had median emission factors of 0.0002, 0.164, and 0.017%, spanning approximately three orders of magnitude within the same process category. A direct comparison was possible between the A2O train (WWTP-B) and the MLE train (WWTP-C), which are operated in parallel at the same site (the second municipal WWTP of G Province). The two trains received the same influent at the same site, yet their median emission factors differed by approximately 50-fold (0.164 versus 0.0033%). The shared influent, climate, and measurement conditions make differences in operating state a more likely explanation than process configuration alone [32]. These results indicate that fixed emission factors based solely on process type may not represent facility-level emissions and agree with previous studies showing substantial temporal variability within individual plants [4,16].

Comparison with the IPCC default further shows the limitation of applying a single emission factor across facilities (Fig. 2b). On an influent-nitrogen basis, the median of the 40 municipal campaigns was 0.010% (range of quantified values, 0.0001–1.39%; five campaigns below detection), the median of the seven facility-level medians was 0.004%, and the nitrogen-load-weighted aggregate emission factor was 0.065% (Table S1). The campaign median was approximately 160-fold lower than the IPCC 2019 default of 1.6% and slightly below the lower bound of the reported range of 0.016–4.5% [3]. Applying the Tier 1 default to the aggregate nitrogen load of this cohort would overestimate its emissions by approximately 25-fold; WWTP-B alone accounted for 86% of the aggregate emission with 15% of the nitrogen load. However, high-emission events still occurred. The monthly removal-based emission factor at WWTP-B reached 1.8%, while WWTP-H showed persistently high emissions. A



**Fig. 2.** N<sub>2</sub>O emission factors across eight facilities. (a) Removal-based emission factors by facility and process configuration. (b) Load-based emission factors compared with the IPCC 2019 default (1.6%) and reported range (0.016–4.5%). Boxes show campaign-level emission factors (municipal,  $n = 40$  campaigns at seven facilities; leachate,  $n = 13$  campaigns); non-detect campaigns ( $n = 5$ ) are plotted at 10<sup>-5</sup>%. The number of campaigns per facility was 5, 8, 6, 5, 4, 8, 4, and 13 for WWTP-A to WWTP-H, respectively (53 in total; Table 1).

small number of facilities and operating periods therefore accounted for a large share of total emissions. Accurate inventories require identification of these high-emission facilities and operating states rather than reliance on a single generic emission factor.

Temporal emission patterns also differed between the municipal and leachate systems (Fig. S2). Gaseous N<sub>2</sub>O concentrations at the municipal facilities were episodic. Within-facility coefficients of variation for gaseous N<sub>2</sub>O ranged from 1.1 to 2.1 at six of the seven municipal facilities, the exception being WWTP-F (0.59), which emitted persistently (Section 3.2). At WWTP-B, the N<sub>2</sub>O concentration varied from 0.6 to 1101 ppmv between consecutive summer campaigns. In contrast, gaseous N<sub>2</sub>O concentrations at the leachate facility were chronically elevated. N<sub>2</sub>O concentrations exceeded 400 ppmv in 10 of 13 campaigns, with a coefficient of variation of 0.57. This pattern coincided with conditions unfavorable for complete nitrogen conversion, including influent COD/N below 2.2, aerobic-zone DO mostly below 1.5 mg O<sub>2</sub>/L, and sustained nitrite accumulation of 7.5–167 mg N/L (Section 3.2). These different temporal patterns have practical implications for monitoring. Episodic emissions may be missed by infrequent campaign-based measurements, whereas chronically elevated systems can be identified with less frequent monitoring [5]. This distinction supports the screening-based prioritization evaluated in Section 3.4.

### 3.2. Context-dependent drivers and two emission regimes

Relationships between routine operating parameters and N<sub>2</sub>O emissions differed between the municipal and leachate systems. The two groups differed significantly in pH, temperature, DO, ammonium, nitrite, COD/N, the AOB/NOB activity ratio, and nitrogen-removal efficiency (Mann-Whitney  $U$  test,  $p < 0.05$ ; Table 2, Fig. S3). Individual facilities also showed distinct operating profiles (Figs. 3 and 4). Fig. 3 presents the measured operating variables underlying this comparison; the gaseous N<sub>2</sub>O concentration in Fig. 3a is the measured quantity from which the emission factors in Fig. 2 are derived (Section 2.2). These differences were reflected in the correlation analysis (Fig. 5a). In the pooled dataset, nitrite showed the strongest positive correlation with the log emission factor (Spearman  $\rho = 0.54$ ,  $p < 0.001$ ), followed by DIC removal ( $\rho = 0.46$ ). However, nitrite was below the detection limit in most municipal campaigns, showing that the pooled correlation was driven mainly by the leachate facility. DO showed no correlation in the pooled dataset ( $\rho = 0.02$ ) but was positively correlated with emissions in the municipal facilities ( $\rho = 0.51$ ,  $p = 0.001$ ) and showed no clear relationship at the leachate facility (Fig. 5b). NH<sub>4</sub><sup>+</sup>-N and nitrogen loading also showed opposite correlation directions between the pooled dataset and the municipal facilities. In contrast, activity-related variables, particularly DIC removal, maintained the same direction across the different data subsets (Fig. 5c). These results show that the same bulk water-quality variable can

represent different operating states among facilities. Fixed concentration thresholds may therefore not be directly transferable across treatment systems.

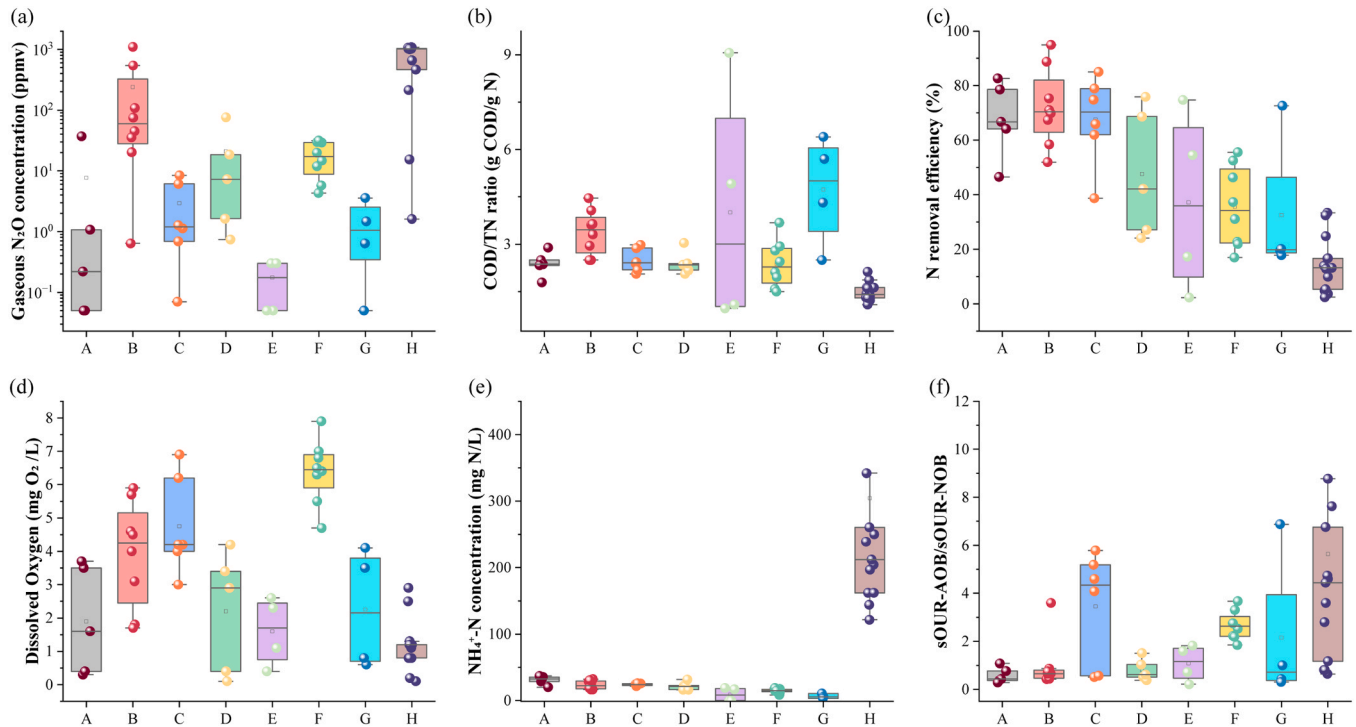
N<sub>2</sub>O emissions at the municipal facilities were episodic and occurred under different operating conditions [33]. At WWTP-B, the two highest-emission campaigns occurred under contrasting conditions. One campaign had a high temperature (29 °C) and high DO (4.5 mg/L), conditions consistent with rapid nitrification and enhanced gas stripping [34]. In contrast, another campaign had low DO (1.7 mg/L) and depleted nitrate, consistent with nitrifier denitrification under oxygen limitation [35]. WWTP-F, operated as an SBR, showed another pattern. Emission factors remained elevated throughout the year (0.21–1.83% of removed nitrogen) despite consistently high DO (4.7–7.9 mg/L), while the AOB/NOB activity ratio remained between 1.8 and 3.7. This pattern indicates relatively low potential NOB activity compared with potential AOB activity and is consistent with a recurring nitrite-oxidation bottleneck during SBR operation. These observations help explain the positive relationship between DO and N<sub>2</sub>O emissions in the municipal facilities. Aeration is associated with both nitrifying activity and gas stripping, so high DO can occur together with elevated N<sub>2</sub>O emissions without directly causing them. A single DO threshold derived from low-DO laboratory studies may therefore not be applicable across full-scale facilities [1].

The leachate facility showed a different pattern, with chronically high emissions under persistently unfavorable operating conditions. Influent pH was approximately 8.0 but decreased to about 4.3 along the treatment line, suggesting insufficient alkalinity recovery during denitrification to

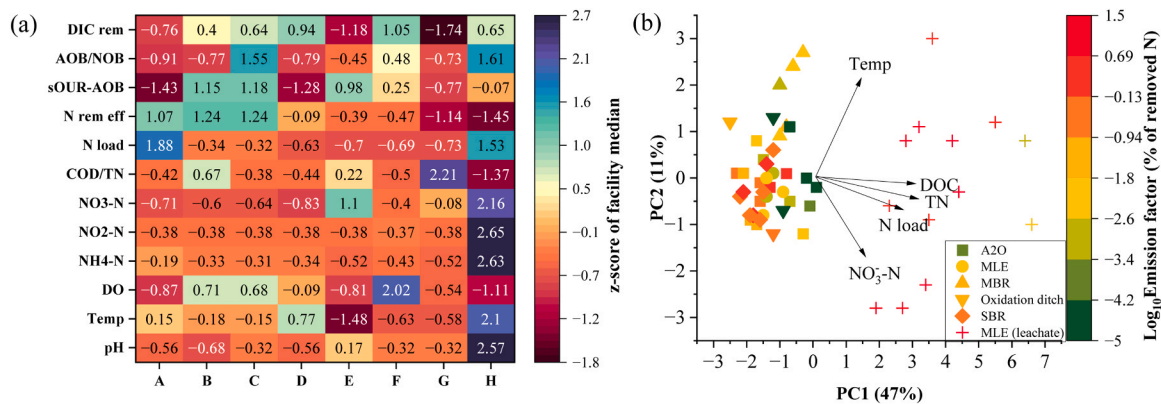
**Table 2**

Operating conditions and activity indicators of the municipal and leachate cohorts, median (range) of campaigns.

| Parameter                                | Municipal ( $n = 40$ ) | Leachate ( $n = 13$ ) |
|------------------------------------------|------------------------|-----------------------|
| pH                                       | 6.9 (5.7–7.4)          | 8.1 (5.7–8.7)         |
| Temperature (°C)                         | 19.7 (9.5–30.5)        | 29.0 (18.6–33.5)      |
| DO (mg/L)                                | 3.9 (0.1–7.9)          | 1.2 (0.1–2.9)         |
| MLSS (mg/L)                              | 3462 (1110–12980)      | 4200 (2040–11320)     |
| DIC (mg-C/L)                             | 47 (27–68)             | 147 (24–487)          |
| DOC (mg-C/L)                             | 50 (19–74)             | 248 (158–346)         |
| NH <sub>4</sub> <sup>+</sup> -N (mg-N/L) | 19.0 (0.0–37.2)        | 238.8 (121.5–1093.0)  |
| NO <sub>2</sub> -N (mg-N/L)              | 0.0 (0.0–1.8)          | 24.4 (0.0–166.6)      |
| NO <sub>3</sub> -N (mg-N/L)              | 0.3 (0.0–9.1)          | 3.8 (1.9–146.5)       |
| TN (mg/L)                                | 33.5 (4.0–56.9)        | 417.0 (312.0–718.4)   |
| COD/N (-)                                | 2.50 (0.97–9.06)       | 1.41 (1.09–2.13)      |
| N loading (kg-N/d)                       | 104 (11–1892)          | 1363 (1043–2401)      |
| N removal efficiency (%)                 | 57.0 (2.3–94.9)        | 13.2 (2.4–33.4)       |
| sOUR-AOB (mg O <sub>2</sub> /g MLVSS•h)  | 2.14 (0.07–7.92)       | 2.01 (0.35–17.87)     |
| sOUR-NOB (mg O <sub>2</sub> /g MLVSS•h)  | 1.03 (0.21–9.14)       | 0.53 (0.07–2.49)      |
| sOUR-AOB/NOB (-)                         | 0.93 (0.21–6.88)       | 4.44 (0.64–26.51)     |
| DIC removal (%)                          | 34.0 (2.6–69.4)        | 37.0 (0.0–100.0)      |
| Gaseous N <sub>2</sub> O (ppmv)          | 5.0 (0.0–1100.6)       | 1017.2 (1.6–1080.1)   |



**Fig. 3.** Facility-level distributions of measured operating variables across 53 campaigns: (a) gaseous  $\text{N}_2\text{O}$  concentration, (b) COD/TN ratio, (c) nitrogen removal efficiency, (d) dissolved oxygen, (e)  $\text{NH}_4^+\text{-N}$ , and (f) AOB/NOB activity ratio. Panel (a) shows the measured off-gas concentration, which is distinct from the emission factors in Fig. 2 derived from it through Eqs. (1) to (3). The number of campaigns per facility was 5, 8, 6, 5, 4, 8, 4, and 13 for WWTP-A to WWTP-H, respectively (53 in total; Table 1).



**Fig. 4.** Facility operating fingerprints. (a) Standardized facility medians and (b) PCA of 53 campaigns based on water-quality variables. Color indicates  $\log_{10}$  emission factor and arrows show the five strongest loadings; PC1 and PC2 explain 58% of total variance. The number of campaigns per facility was 5, 8, 6, 5, 4, 8, 4, and 13 for WWTP-A to WWTP-H, respectively (53 in total; Table 1).

buffer nitrification. As a result, the nitrification stage operated near pH 5 rather than the typical optimal range of 7–8. COD/N remained below 2.2 throughout the monitoring period, below the stoichiometric requirement for complete denitrification, and much of the influent organic carbon was recalcitrant. Low pH, low DO, and carbon limitation together created unfavorable conditions for NOB relative to AOB [36]. Nitrite accumulated to 7.5–167 mg N/L, while ammonium removal remained low at a median of 13%. Continued ammonia oxidation under this downstream limitation coincided with  $\text{N}_2\text{O}$  concentrations of 465–1080 ppmv in 10 of 13 campaigns, without seasonal dependence. The median removal-based emission factor was 5.3%, which was of the same order as the 8.9–11.9% previously reported for full-scale leachate treatment [17]. These results indicate that high-strength, carbon-limited systems should be considered separately from conventional municipal treatment when evaluating  $\text{N}_2\text{O}$  emission potential.

These observations support a two-condition framework for elevated  $\text{N}_2\text{O}$  emissions. First, ammonia oxidation must remain active to generate reactive nitrogen intermediates. Second, a downstream bottleneck must limit nitrite oxidation or denitrification. The dataset shows that neither condition alone was sufficient. At the leachate facility, sOUR-AOB decreased to 0.35 and 0.50  $\text{mg O}_2/\text{g MLVSS}\cdot\text{h}$  in two consecutive campaigns, the lowest values recorded at that facility, and gaseous  $\text{N}_2\text{O}$  decreased to 1.6 and 15.4 ppmv even though the carbon-limited downstream bottleneck remained. In the following campaign at the same facility, sOUR-AOB recovered to 17.87  $\text{mg O}_2/\text{g MLVSS}\cdot\text{h}$  and gaseous  $\text{N}_2\text{O}$  returned to 1025 ppmv. The gaseous  $\text{N}_2\text{O}$  concentration therefore tracked potential ammonia-oxidizing activity while the bottleneck was unchanged, and the removal-based emission factor showed the same pattern, falling to 0.01% or below in the two low-activity campaigns and returning to 1.45% upon recovery. WWTP-A represented the opposite condition,

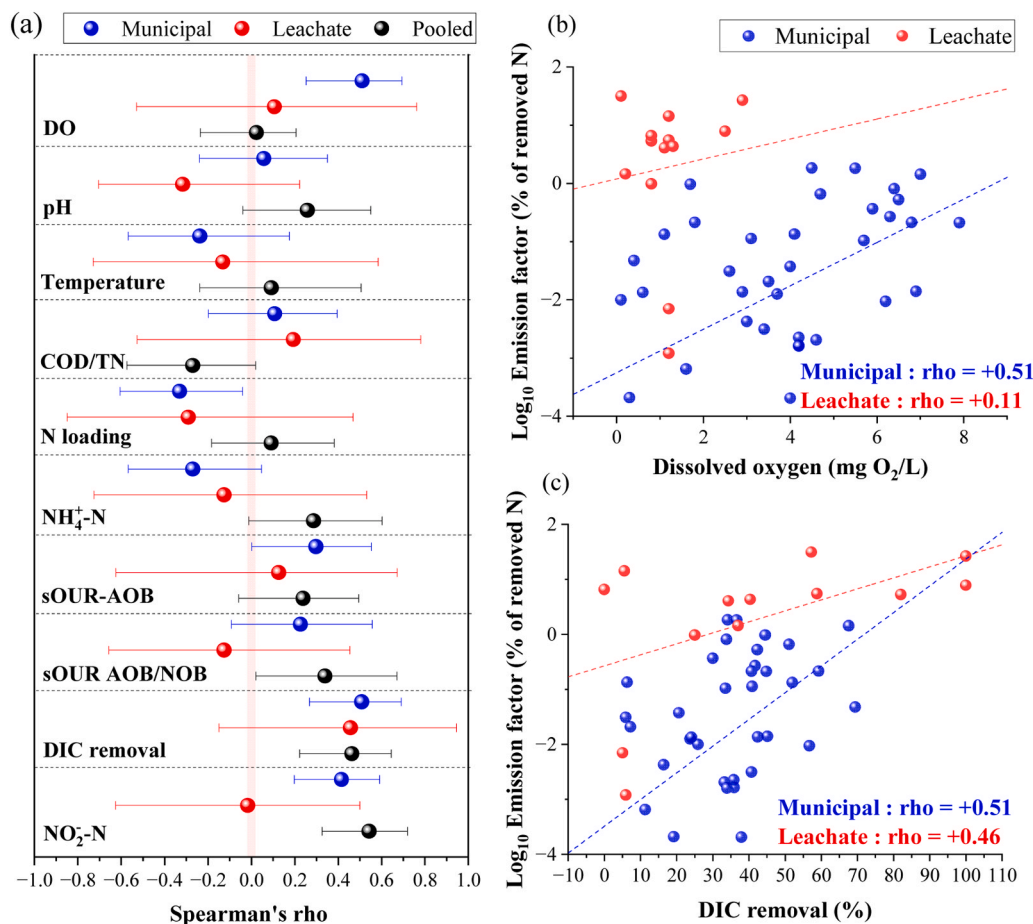


Fig. 5. Regime-dependent associations with log<sub>10</sub> emission factor. (a) Spearman rho with 95% bootstrap confidence intervals for pooled and cohort-specific data. Relationships with (b) dissolved oxygen and (c) DIC removal are shown with within-cohort linear trends.

with low potential nitrifying activity, no evident downstream bottleneck, and a median emission factor of only 0.0002%. When all 53 campaigns were classified using the two thresholds evaluated in Section 3.4, the median emission factor was 72-fold higher in the high-potential tier than in the low-potential tier (0.817 versus 0.011%; Fig. 7). This two-condition framework formed the basis for the activity indicators evaluated in the following sections.

### 3.3. Trained models do not transfer across facilities

The transferability analysis showed a consistent pattern across models (Fig. 6, Table S4). Under random 5-fold cross-validation, four of the five algorithms trained on the 14 water-quality variables achieved  $R^2$  values of 0.29–0.33 for the log-transformed emission factor, comparable to values reported in previous campaign-scale studies [18,37]. In contrast, all 15 combinations of algorithm and physicochemical or activity feature set produced negative  $R^2$  values under LOFO validation (Fig. 6a). Facility-identity models are shown for reference only, since a held-out facility has no corresponding dummy variable in the training set. A negative  $R^2$  indicates that model predictions performed worse than a constant mean-value prediction. For the water-quality random forest,  $R^2$  decreased from 0.33 in the representative cross-validation split shown in Figs. 6b to –1.16 under LOFO validation (Fig. 6c). Its average over repeated cross-validation was 0.29 (Table S4). The consistent failure across algorithms indicates that poor transferability was not specific to a particular model [23].

The facility-identity benchmark helped explain the higher performance under random cross-validation. Models using only facility dummy variables, without any physicochemical information, achieved

$R^2$  values of up to 0.45 and outperformed all models based on water-quality variables. This result shows that random splitting allowed the models to exploit facility-specific patterns. Routine water-quality variables effectively served as facility fingerprints, enabling the models to distinguish among plants rather than learn a common relationship between operating conditions and  $\text{N}_2\text{O}$  emissions [5,38]. The sign reversals observed in Section 3.2 support this interpretation. Relationships between bulk water-quality variables and  $\text{N}_2\text{O}$  emissions differed among operating regimes, so patterns learned at one facility did not necessarily apply to another [5,10].

Although quantitative prediction failed under LOFO validation, the linear model retained some rank-order information (Spearman rho = 0.55), even though its  $R^2$  fell below –100. Higher- and lower-emitting states could therefore still be partially distinguished despite poor quantitative prediction. This result suggests that cross-facility approaches may be more suitable for ranking or threshold-based screening than for direct prediction of emission factors [7]. This concept formed the basis of the screening approach evaluated in Section 3.4.

These results also affect how single-plant ML studies should be interpreted.  $R^2$  values near 0.9 obtained from random splits of data from a single facility demonstrate strong within-facility prediction but do not establish transferability to unmeasured plants [18,19,38]. ML studies intended for application beyond the training facility should therefore include facility-out validation in addition to random splits, since random splits assess within-facility prediction and period-out splits assess temporal extrapolation within the same facility, neither of which addresses generalization to a new facility. Random-split performance should be interpreted as within-facility predictive performance rather than evidence of cross-facility transferability.

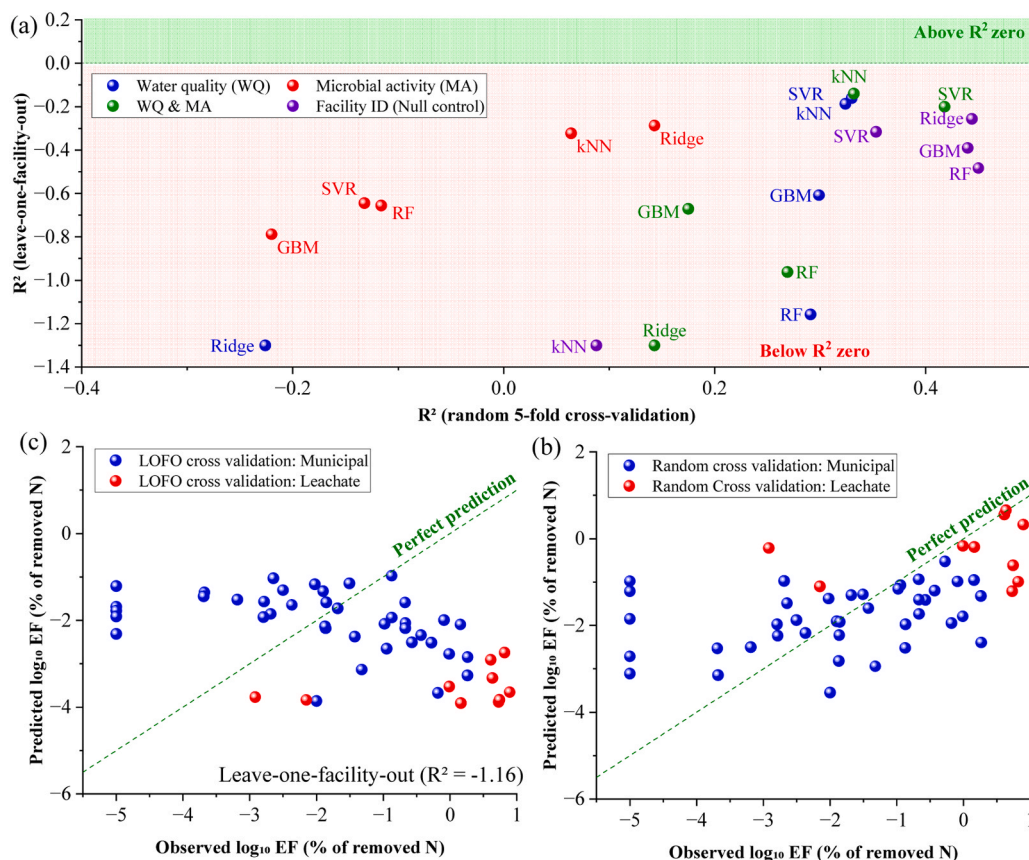


Fig. 6. Model transferability. (a)  $R^2$  under random 5-fold and leave-one-facility-out cross-validation for five algorithms and four feature sets. (b, c) Observed versus predicted  $\log_{10}$  emission factors for the water-quality random forest under random ( $R^2 = 0.33$ ) and facility-out validation ( $R^2 = -1.16$ ).

### 3.4. Activity-based indicators for emission screening

The poor cross-facility performance of the trained models motivated evaluation of fixed activity-based indicators. Three indicators were selected from the two-condition framework. sOUR-AOB represented potential ammonia-oxidizing activity [11], the AOB/NOB activity ratio represented a potential nitrite-oxidation bottleneck [9,36], and DIC removal represented realized autotrophic conversion and associated alkalinity demand. The rank-based activity index combining these three indicators identified elevated-emission states with an AUC of 0.78 (95% CI, 0.65–0.90), while the individual indicators achieved AUC values of 0.68–0.75 (Table 3).

Conventional variables showed weaker or system-dependent performance. The 95% confidence intervals for DO (AUC 0.58), temperature (0.52),  $\text{NH}_4^+\text{-N}$  (0.52), and nitrogen loading (0.49) all included 0.50. Nitrite achieved an AUC of 0.72, but it was below the detection limit in most municipal campaigns, so its screening performance was driven mainly by the leachate facility. The activity index remained stable across different definitions of elevated emission, with AUC values of 0.75–0.78 for thresholds ranging from 0.05% to 0.5% (Table S6). The index was also recalculated using a leave-one-facility-out ranking scheme, in which each campaign was scored against the distribution of the remaining facilities only. This analysis produced an AUC of 0.79 (95% CI, 0.65–0.90), showing that discrimination did not depend on including the evaluated facility in the reference distribution (Table S9).

Mixed-effects analysis showed that the activity indicators provided information beyond facility identity. With facility included as a random intercept, the three indicators jointly improved model likelihood relative to the facility-only model (likelihood-ratio chi-square = 12.4,  $df = 3$ ,  $p = 0.006$ ; Table S5). DIC removal was the strongest individual predictor (+0.50 log units per SD,  $p = 0.003$ ), whereas the coefficients for sOUR-

AOB and the AOB/NOB activity ratio were not individually significant ( $p = 0.10$  and  $0.13$ ; Table S5). The three-indicator block improved model fit, with DIC removal showing the strongest individual association. This block-level test does not establish that the summed activity index itself is

Table 3

Discrimination of elevated-emission states (removal-based EF  $\geq 0.1\%$ ) by fixed indicators: AUC with 95% bootstrap confidence interval and Spearman correlation with the  $\log_{10}$  emission factor ( $n = 53$ ).

| Indicator                         | Family       | AUC (95% CI)     | Spearman rho      |
|-----------------------------------|--------------|------------------|-------------------|
| Composite (3 activity indicators) | Composite    | 0.78 (0.65–0.90) | 0.46 <sup>a</sup> |
| DIC removal                       | Activity     | 0.75 (0.61–0.88) | 0.46 <sup>a</sup> |
| $\text{NO}_2\text{-N}$            | Conventional | 0.72 (0.62–0.82) | 0.54 <sup>a</sup> |
| Composite (2 activity indicators) | Composite    | 0.72 (0.57–0.86) | 0.33 <sup>a</sup> |
| sOUR-AOB                          | Activity     | 0.69 (0.53–0.83) | 0.24              |
| sOUR-AOB/NOB                      | Activity     | 0.68 (0.53–0.82) | 0.34 <sup>a</sup> |
| DO                                | Conventional | 0.58 (0.42–0.74) | 0.02              |
| Temperature                       | Conventional | 0.52 (0.36–0.68) | 0.09              |
| $\text{NH}_4^+\text{-N}$          | Conventional | 0.52 (0.35–0.68) | 0.26              |
| N loading                         | Conventional | 0.49 (0.32–0.65) | 0.09              |
| COD/N                             | Conventional | 0.37 (0.22–0.53) | -0.27             |

<sup>a</sup>  $p < 0.05$ .

significant after facility adjustment. DO and nitrite together did not improve the facility-only model ( $p = 0.39$ ). After excluding the leachate facility, the activity index retained an AUC of 0.76 (95% CI, 0.61–0.90) across the 40 municipal campaigns. DO also showed good discrimination within the municipal facilities (AUC 0.79), but its positive relationship there was absent at the leachate facility [34,35]. A variable whose relationship with  $N_2O$  changes among systems is difficult to use as a general screening indicator. In contrast, the activity-based indicators maintained the same direction across both municipal and leachate systems.

The AOB/NOB activity ratio has a stoichiometric reference point that does not depend on facility-specific calibration. Ammonium oxidation to nitrite consumes approximately 3.43 g  $O_2$ /g N, whereas nitrite oxidation consumes approximately 1.14 g  $O_2$ /g N [30,31]. Under excess-substrate conditions, equal nitrogen-processing capacities of AOB and NOB correspond to an AOB/NOB activity ratio of approximately 3.0. Ratios above 3.0 indicate that nitrite-production capacity exceeds nitrite-oxidation capacity. The other two thresholds have a different basis. The sOUR-AOB threshold of 2.0 mg  $O_2$ /g MLVSS•h and the DIC removal threshold of 35% correspond to the cohort medians of 2.11 mg  $O_2$ /g MLVSS•h and 34.3%, rounded to practical operational cutoffs, and are therefore operational rather than mechanistic. These thresholds were not selected to maximize discrimination. Cutoffs based on the Youden index would have been 1.25 mg  $O_2$ /g MLVSS•h and 30%, increasing the AUC of the three-criterion score only modestly from 0.73 to 0.76 (Table S10). The reported thresholds were therefore not optimized for this dataset, but they still require validation at independent facilities before broader application.

Five of the seven municipal facilities had median AOB/NOB activity ratios of 0.4–1.2, indicating substantial excess NOB capacity relative to AOB. The leachate facility had a median ratio of 4.4, consistent with its persistent nitrite accumulation. WWTP-F and WWTP-C illustrate the limitations of using the ratio alone. WWTP-F had a median ratio of 2.6 and remained a persistent emitter near stoichiometric parity, whereas WWTP-C had a median ratio of 4.3 but emitted little  $N_2O$ . The sOUR assay measures potential activity under excess-substrate conditions, while an actual nitrite-oxidation bottleneck develops only when the applied nitrogen flux approaches the available capacity [31]. WWTP-C had a high AOB/NOB activity ratio but a moderate nitrogen load and accumulated little or no nitrite. The ratio may therefore produce false positives when used alone, which is acceptable for a first-stage screening tool intended to identify facilities for direct  $N_2O$  measurement.

The two thresholds were then used to classify campaigns into screening tiers (Fig. 7a). First, sOUR-AOB  $\geq 2.0$  mg  $O_2$ /g MLVSS•h indicated sufficient potential ammonia-oxidizing activity. Second, an AOB/NOB activity ratio  $\geq 3.0$  indicated a potential nitrite-oxidation bottleneck. Campaigns meeting both criteria were classified as high emission potential, those meeting one criterion as moderate, and those meeting neither criterion as low. Median emission factors increased from 0.011% in the low tier ( $n = 20$ ) to 0.134% in the moderate tier ( $n = 21$ ) and 0.817% in the high tier ( $n = 12$ ) (Fig. 7b). The median emission factor was therefore 72-fold higher in the high tier than in the low tier. The proportion of campaigns above the elevated-emission threshold increased from 30% in the low tier to 62% in the moderate tier and 67% in the high tier. A three-criterion score additionally including DIC removal  $\geq 35\%$  achieved an AUC of 0.73 (95% CI, 0.59–0.86), similar to that of the rank-based activity index. Flagging campaigns that met at least two criteria resulted in a sensitivity of 0.63 and a specificity of 0.77 (Table S7). Median emission factors increased from 0.002% at score 0 to approximately 0.97% at score 2. The lower value at score 3 was mainly attributable to WWTP-C (Fig. S4). All three indicators can be measured using standard laboratory respirometry and carbon analysis [39]. The screening approach therefore does not require a gas-phase analyzer, flux chamber, or facility-specific  $N_2O$  calibration and can be applied before direct  $N_2O$  measurements are available.

### 3.5. Engineering implications, integrated pollution control, and limitations

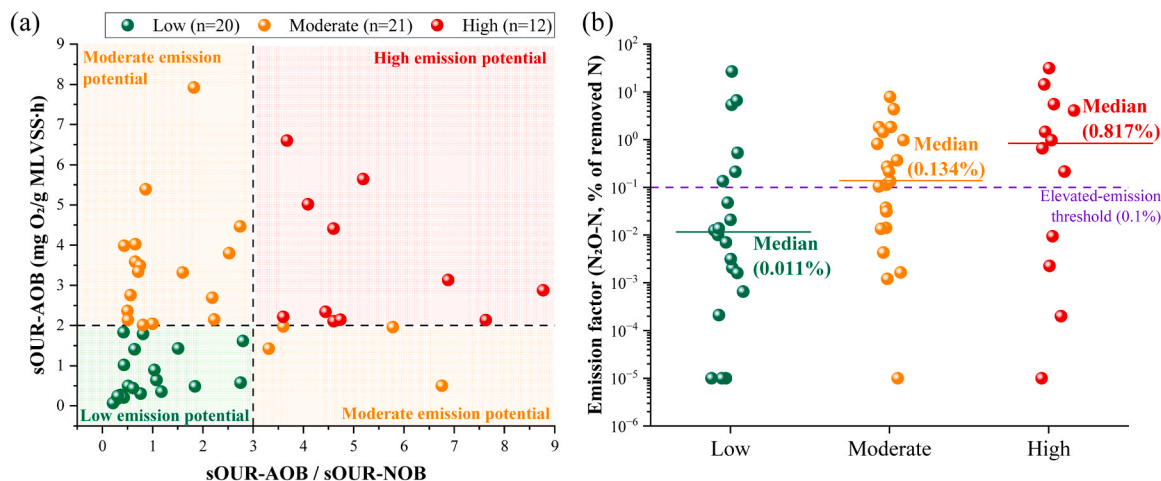
The results support a three-step strategy for  $N_2O$  management when monitoring capacity is limited. First, respirometric screening can be incorporated into routine mixed-liquor sampling. Measuring only the top 25% of campaigns ranked by the activity index captured 42% of the estimated total emissions, a 1.7-fold enrichment over random prioritization, and 35% among municipal facilities alone (Fig. 8a,b). Nitrite captured more in the pooled dataset (73%), but only because of the nitrite-rich leachate facility; among municipal facilities its capture fell below random prioritization (21%), as did DO, while the activity index remained above random (Fig. 8c). The three facilities ranked highest by median activity index accounted for 94% of the estimated total emissions. Second, direct flux-chamber measurements can then be focused on the facilities and operating states identified by screening. Third, mitigation such as aeration control, alkalinity management, or supplemental carbon can be prioritized where measurements confirm elevated emissions [40–42]. The same logic applies to inventories. Because the Tier 1 default overestimated the aggregate emission of the municipal cohort by approximately 25-fold, and the typical campaign by two orders of magnitude, while a single facility contributed 86% of the total, targeted measurement may improve accuracy more effectively than refining a single generic factor, and leachate treatment facilities may warrant separate inventory categories [17].

The strategy is also consistent with integrated pollution control because it prioritizes reduction of  $N_2O$  generation at the source. Screening requires only routine laboratory analysis, and the mitigation measures identified by the screen are primarily operational. Aeration control can change electricity demand, while alkalinity or external-carbon dosing can increase chemical use and sludge production. These trade-offs should therefore be evaluated on a  $CO_2$ -equivalent basis against the reduction in  $N_2O$  emissions. End-of-pipe options such as off-gas biofiltration instead require treatment of the collected gas and may generate spent media.

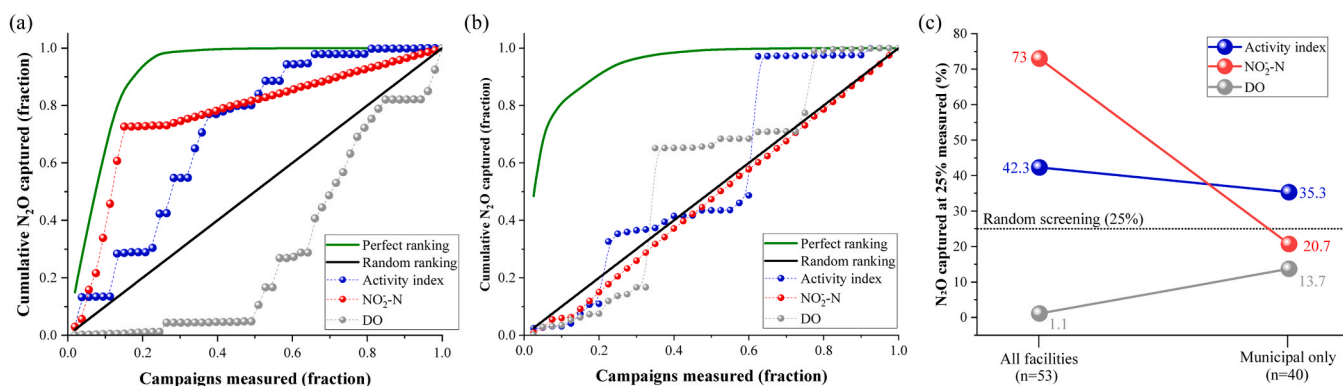
Four limitations should be considered. First, monthly campaigns may miss episodic emissions [43], and the surface flux chamber may under-collect  $N_2O$  released through coarse aeration bubbles. Absolute emission factors are therefore uncertain and may be conservative, although relative comparisons among facilities are less affected because all facilities were monitored using the same protocol. Second, only the AOB/NOB activity-ratio threshold is derived from stoichiometry. The sOUR-AOB and DIC removal thresholds correspond to the cohort medians and were evaluated using the same dataset in which the two-condition framework was identified. Independent prospective validation is therefore required before these thresholds are applied more broadly. The indicators were evaluated for cross-facility screening within the present cohort; their performance at facilities outside this cohort has not been tested and should therefore not be assumed. Third, within-facility rank consistency was weak (Table S8), indicating that the current indicators are better suited to cross-facility screening than to temporal tracking within a single facility. Finally, the study included only eight facilities, and the leachate regime was represented by a single system. Extending respirometric screening to additional high-strength treatment systems and pooling campaign-level data across studies may support the development of more robust regime-specific models.

## 4. Conclusion

Across eight full-scale facilities monitored using the same flux-chamber protocol,  $N_2O$  emission factors varied by more than four orders of magnitude and showed little relationship with process configuration. On an influent-nitrogen basis, the median emission factor for municipal facilities was approximately 160-fold lower than the IPCC



**Fig. 7.** Screening tiers defined by one stoichiometric and one operational threshold. (a) Classification of low, moderate, and high emission potential using an AOB/NOB activity ratio of 3.0 and sOUR-AOB of 2.0 mg O<sub>2</sub>/g MLVSS-h. One high-tier campaign from WWTP-H (sOUR-AOB = 17.9 mg O<sub>2</sub>/g MLVSS-h; AOB/NOB ratio = 26.5) falls outside the displayed axes but is included in the tier counts and panel (b). (b) Removal-based emission factors by tier; the dashed line marks the elevated-emission threshold of 0.1%.



**Fig. 8.** Emission captured by indicator-guided measurement. Cumulative share of estimated N<sub>2</sub>O captured by prioritized campaigns for (a) all facilities and (b) municipal facilities, compared with perfect and random ranking. (c) Emission captured when 25% of campaigns are measured; NO<sub>2</sub>-N performs below random prioritization after excluding the leachate facility.

Tier 1 default, while the landfill leachate facility showed chronically elevated emissions with a median removal-based emission factor of 5.3%. Machine-learning models based on routine water-quality variables failed leave-one-facility-out validation in all 15 combinations of algorithm and physicochemical or activity feature set. Models using facility identity alone outperformed the water-quality models under random cross-validation, indicating that random-split performance mainly reflected facility-specific patterns rather than transferable relationships. This poor transferability was consistent with the different relationships between routine water-quality variables and N<sub>2</sub>O emissions across operating regimes. In contrast, activity-based indicators derived from sOUR-AOB, the AOB/NOB activity ratio, and DIC removal identified elevated-emission states with an activity-index AUC of 0.78 and a threshold-based AUC of 0.73 while the three-indicator block improved model fit after accounting for facility-level effects. Two thresholds separated campaigns into screening tiers, with a 72-fold difference in median emission factor between the high- and low-potential tiers, and measuring only the top 25% of campaigns ranked by the activity index captured 42% of the estimated emissions. These findings support respirometric screening as a candidate approach for prioritizing direct N<sub>2</sub>O measurement and mitigation at facilities and operating states with high emission potential, subject to validation at independent facilities.

#### CRedit authorship contribution statement

**Seongbong Heo:** Writing – review & editing, Writing – original draft, Visualization, Formal analysis, Conceptualization. **Hyeongju Park:** Writing – review & editing, Methodology, Formal analysis. **Chamsol Kwon:** Visualization. **Young Mo Kim:** Writing – review & editing, Supervision, Project administration, Data curation, Conceptualization.

#### Declaration of Competing Interest

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.

#### Acknowledgement

This work was supported by a grant from the National Research Foundation of Korea (NRF), funded by the Korean government (MSIT) (No. RS-2026-25485864).

#### Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at [doi:10.1016/j.dwt.2026.101944](https://doi.org/10.1016/j.dwt.2026.101944).

## Data availability

Data will be made available on request.

## References

- Kampschreur MJ, Temmink H, Kleerebezem R, Jetten MSM, van Loosdrecht MCM. Nitrous oxide emission during wastewater treatment. *Water Res* 2009;43: 4093–103. <https://doi.org/10.1016/j.watres.2009.03.001>.
- Law Y, Ye L, Pan Y, Yuan Z. Nitrous oxide emissions from wastewater treatment processes. *Philos Trans R Soc B Biol Sci* 2012;367:1265–77. <https://doi.org/10.1098/rstb.2011.0317>.
- IPCC, Refinement to the 2006 IPCC Guidelines for National Greenhouse Gas Inventories, Geneva, 2019. <https://www.ipcc-nggip.iges.or.jp/public/2019rf/1/ndex.html> (accessed August 27, 2026).
- Daelman MRJ, van Voorthuizen EM, van Dongen UGJM, Volcke EIP, van Loosdrecht MCM. Seasonal and diurnal variability of N<sub>2</sub>O emissions from a full-scale municipal wastewater treatment plant. *Sci Total Environ* 2015;536:1–11. <https://doi.org/10.1016/j.scitotenv.2015.06.122>.
- Vasilaki V, Massara TM, Stanchev P, Fatone F, Katsou E. A decade of nitrous oxide (N<sub>2</sub>O) monitoring in full-scale wastewater treatment processes: a critical review. *Water Res* 2019;161:392–412. <https://doi.org/10.1016/j.watres.2019.04.022>.
- de Haas D, Andrews J. Nitrous oxide emissions from wastewater treatment - revisiting the IPCC 2019 refinement guidelines. *Environ Chall* 2022;8. <https://doi.org/10.1016/j.envc.2022.100557>.
- Shang Z, Cai C, Guo Y, Huang X, Peng K, Guo R, Wei Z, Wu C, Cheng S, Liao Y, Hung CY, Liu J. Direct and indirect monitoring methods for nitrous oxide emissions in full-scale wastewater treatment plants: a critical review. *J Environ Manag* 2024; 358. <https://doi.org/10.1016/j.jenvman.2024.120842>.
- Schreiber F, Wunderlin P, Udert KM, Wells GF. Nitric oxide and nitrous oxide turnover in natural and engineered microbial communities: biological pathways, chemical reactions, and novel technologies. *Front Microbiol* 2012;3. <https://doi.org/10.3389/fmicb.2012.00372>.
- Wunderlin P, Mohn J, Joss A, Emmenegger L, Siegrist H. Mechanisms of N<sub>2</sub>O production in biological wastewater treatment under nitrifying and denitrifying conditions. *Water Res* 2012;46:1027–37. <https://doi.org/10.1016/j.watres.2011.11.080>.
- Massara TM, Malamis S, Guisasaola A, Baeza JA, Noutsopoulos C, Katsou E. A review on nitrous oxide (N<sub>2</sub>O) emissions during biological nutrient removal from municipal wastewater and sludge reject water. *Sci Total Environ* 2017;596–597: 106–23. <https://doi.org/10.1016/j.scitotenv.2017.03.191>.
- Chandran K, Stein LY, Klotz MG, Van Loosdrecht MCM. Nitrous oxide production by lithotrophic ammonia-oxidizing bacteria and implications for engineered nitrogen-removal systems. *Biochem Soc Trans* 2011;1832–7. <https://doi.org/10.1042/BST20110717>.
- Law Y, Lant P, Yuan Z. The effect of pH on N<sub>2</sub>O production under aerobic conditions in a partial nitrification system. *Water Res* 2011;45:5934–44. <https://doi.org/10.1016/j.watres.2011.08.055>.
- Ahn JH, Kim S, Park H, Rahm B, Pagilla K, Chandran K. N<sub>2</sub>O emissions from activated sludge processes, 2008–2009: results of a national monitoring survey in the united states. *Environ Sci Technol* 2010;44:4505–11. <https://doi.org/10.1021/es903845y>.
- Sabba F, Terada A, Wells G, Smets BF, Nerenberg R. Nitrous oxide emissions from biofilm processes for wastewater treatment. *Appl Microbiol Biotechnol* 2018;102: 9815–29. <https://doi.org/10.1007/s00253-018-9332-7>.
- Foley J, de Haas D, Yuan Z, Lant P. Nitrous oxide generation in full-scale biological nutrient removal wastewater treatment plants. *Water Res* 2010;44:831–44. <https://doi.org/10.1016/j.watres.2009.10.033>.
- Gruber W, von Känel L, Vogt L, Luck M, Biolley L, Feller K, Moosmann A, Krähenbühl N, Kipf M, Loosli R, Vogel M, Morgenroth E, Braun D, Joss A. Estimation of countrywide N<sub>2</sub>O emissions from wastewater treatment in Switzerland using long-term monitoring data. *Water Res X* 2021;13. <https://doi.org/10.1016/j.wroa.2021.100122>.
- Wang X, Jia M, Zhang C, Chen S, Cai Z. Leachate treatment in landfills is a significant N<sub>2</sub>O source. *Sci Total Environ* 2017;596–597:18–25. <https://doi.org/10.1016/j.scitotenv.2017.04.029>.
- Khalil M, AlSayed A, Liu Y, Vanrolleghem PA. Machine learning for modeling N<sub>2</sub>O emissions from wastewater treatment plants: aligning model performance, complexity, and interpretability. *Water Res* 2023;245. <https://doi.org/10.1016/j.watres.2023.120667>.
- Seshan S, Poinapen J, Zandvoort MH, van Lier JB, Kapelan Z. Forecasting nitrous oxide emissions from a full-scale wastewater treatment plant using LSTM-based deep learning models. *Water Res* 2025;268. <https://doi.org/10.1016/j.watres.2024.122754>.
- Ni BJ, Yuan Z. Recent advances in mathematical modeling of nitrous oxides emissions from wastewater treatment processes. *Water Res* 2015;87:336–46. <https://doi.org/10.1016/j.watres.2015.09.049>.
- Ni BJ, Ye L, Law Y, Byers C, Yuan Z. Mathematical modeling of nitrous oxide (N<sub>2</sub>O) emissions from full-scale wastewater treatment plants. *Environ Sci Technol* 2013; 47:7795–803. <https://doi.org/10.1021/es4005398>.
- Guo L, Vanrolleghem PA. Calibration and validation of an activated sludge model for greenhouse gases no. 1 (ASMG1): prediction of temperature-dependent N<sub>2</sub>O emission dynamics. *Bioprocess Biosyst Eng* 2014;37:151–63. <https://doi.org/10.1007/s00449-013-0978-3>.
- Newhart KB, Holloway RW, Hering AS, Cath TY. Data-driven performance analyses of wastewater treatment plants: a review. *Water Res* 2019;157:498–513. <https://doi.org/10.1016/j.watres.2019.03.030>.
- Hwangbo S, Al R, Chen X, Sin G. Integrated model for understanding N<sub>2</sub>O emissions from wastewater treatment plants: a deep learning approach. *Environ Sci Technol* 2021;55:2143–51. <https://doi.org/10.1021/acs.est.0c05231>.
- Chandran K, Smets BF. Estimating biomass yield coefficients for autotrophic ammonia and nitrite oxidation from batch respirograms. *Water Res* 2001;35: 3153–6. [https://doi.org/10.1016/S0043-1354\(01\)00022-7](https://doi.org/10.1016/S0043-1354(01)00022-7).
- Delre A, Münster J, Scheut C. Greenhouse gas emission quantification from wastewater treatment plants, using a tracer gas dispersion method. *Sci Total Environ* 2017;605–606:258–68. <https://doi.org/10.1016/j.scitotenv.2017.06.177>.
- Chandran K. Protocol for the measurement of nitrous oxide fluxes from biological wastewater treatment plants. *Methods Enzymol* 2011:369–85. <https://doi.org/10.1016/B978-0-12-381294-0.00016-X>.
- Rice EW, Bridgewater Laura. Standard methods for the examination of water and wastewater. American Public Health Association; 2012.
- Peng L, Ni BJ, Ye L, Yuan Z. N<sub>2</sub>O production by ammonia oxidizing bacteria in an enriched nitrifying sludge linearly depends on inorganic carbon concentration. *Water Res* 2015;74:58–66. <https://doi.org/10.1016/j.watres.2015.02.003>.
- P. Ginstet, J.-M. Audic, V. Urbain, J.-C. Block, Estimation of Nitrifying Bacterial Activities by Measuring Oxygen Uptake in the Presence of the Metabolic Inhibitors Allylthiourea and Azide, 1998. <https://journals.asm.org/journal/aem>.
- Surmacz-Gorska J, Gernaey K, Demuyneck C, Vanrolleghem P, Verstraete W. Nitrification monitoring in activated sludge by oxygen uptake rate (OUR) measurements. *Water Res* 1996;30:1228–36. [https://doi.org/10.1016/0043-1354\(95\)00280-4](https://doi.org/10.1016/0043-1354(95)00280-4).
- Yan X, Guo D, Qiu D, Zheng S, Jia M, Zhang M, Liu J, Su X, Sun J. Effect of mixed liquor suspended solid concentration on nitrous oxide emission from an anoxic/oxic sequencing bioreactor. *Desalin Water Treat* 2019;163:48–56. <https://doi.org/10.5004/dwt.2019.24457>.
- Wunderlin P, Lehmann MF, Siegrist H, Tuzson B, Joss A, Emmenegger L, Mohn J. Isotope signatures of N<sub>2</sub>O in a mixed microbial population system: constraints on N<sub>2</sub>O producing pathways in wastewater treatment. *Environ Sci Technol* 2013;47: 1339–48. <https://doi.org/10.1021/es303174x>.
- Peng L, Ni BJ, Erler D, Ye L, Yuan Z. The effect of dissolved oxygen on N<sub>2</sub>O production by ammonia-oxidizing bacteria in an enriched nitrifying sludge. *Water Res* 2014;66:12–21. <https://doi.org/10.1016/j.watres.2014.08.009>.
- Rassamee V, Sattayatewa C, Pagilla K, Chandran K. Effect of oxic and anoxic conditions on nitrous oxide emissions from nitrification and denitrification processes. *Biotechnol Bioeng* 2011;108:2036–45. <https://doi.org/10.1002/bit.23147>.
- Kampschreur MJ, van der Star WRL, Wielders HA, Mulder JW, Jetten MSM, van Loosdrecht MCM. Dynamics of nitric oxide and nitrous oxide emission during full-scale reject water treatment. *Water Res* 2008;42:812–26. <https://doi.org/10.1016/j.watres.2007.08.022>.
- Mahanna H, ELRahsidy N, Kaloop MR, El-Sapakh S, Alluqmani A, Hassan R. Prediction of wastewater treatment plant performance through machine learning techniques. *Desalin Water Treat* 2024;319. <https://doi.org/10.1016/j.dwt.2024.100524>.
- Vasilaki V, Conca V, Frison N, Eusebi AL, Fatone F, Katsou E. A knowledge discovery framework to predict the N<sub>2</sub>O emissions in the wastewater sector. *Water Res* 2020;178. <https://doi.org/10.1016/j.watres.2020.115799>.
- Eaton AD, Franson MAH, Clesceri LS, Rice EW, Greenberg AE. Standard methods for the examination of water & wastewater 2005.
- Duan H, van den Akker B, Thwaites BJ, Peng L, Herman C, Pan Y, Ni BJ, Watt S, Yuan Z, Ye L. Mitigating nitrous oxide emissions at a full-scale wastewater treatment plant. *Water Res* 2020;185. <https://doi.org/10.1016/j.watres.2020.116196>.
- Mannina G, Ekama G, Caniani D, Cosenza A, Esposito G, Gori R, Garrido-Baserba M, Rosso D, Olsson G. Greenhouse gases from wastewater treatment - a review of modelling tools. *Sci Total Environ* 2016;551–552:254–70. <https://doi.org/10.1016/j.scitotenv.2016.01.163>.
- Soo A, Gao L, Shon HK. Pathways to mitigating nitrous oxide emissions throughout the nutrient wastewater circular economy. *Desalin Water Treat* 2026;325. <https://doi.org/10.1016/j.dwt.2025.101611>.
- Domingo-Félez C, Jensen MM, Bang A, Smets BF. Variability and uncertainty analysis of N<sub>2</sub>O emissions from WWTP to improve the accuracy of emission factors and the design of monitoring strategies. *ACS ES T Water* 2024;4:2542–52. <https://doi.org/10.1021/acsestwater.4c00048>.