Identification of primary effecters of N$_2$O emissions from full-scale biological nitrogen removal systems using random forest approach

Min Joon Song$^a$, Sangki Choi$^b$, Wo Bin Bae$^b$, Jaejin Lee$^c$, Heejoo Han$^a$, Daehyun D. Kim$^a$, Miye Kwon$^a$, Jaewook Myung$^a$, Young Mo Kim$^d$, Sukhwan Yoon$^a$,*

$^a$ Department of Civil and Environmental Engineering, KAIST, Daejeon, 34141, Republic of Korea
$^b$ School of Earth Sciences and Environmental Engineering, Gwangju Institute of Science and Technology, Gwangju, 61005, Republic of Korea
$^c$ Department of Agricultural and Biosystems Engineering, Iowa State University, Ames, IA, 50011, United States
$^d$ Department of Civil and Environmental Engineering, Hanyang University, Seongdong-gu, Seoul, 04763, Republic of Korea

**A R T I C L E  I N F O**

Article history:
Received 18 March 2020
Received in revised form 30 June 2020
Accepted 2 July 2020
Available online 6 July 2020

Keywords:
Nitrous oxide
Wastewater treatment plants
Biological nitrogen removal
Nitrification
Denitrification
Random forest

**A B S T R A C T**

Wastewater treatment plants (WWTPs) have long been recognized as point sources of N$_2$O, a potent greenhouse gas and ozone-depleting agent. Multiple mechanisms, both biotic and abiotic, have been suggested to be responsible for N$_2$O production from WWTPs, with basis on extrapolation from laboratory results and statistical analyses of metadata collected from operational full-scale plants. In this study, random forest (RF) analysis, a machine-learning approach for feature selection from highly multivariate datasets, was adopted to investigate N$_2$O production mechanism in activated sludge tanks of WWTPs from a novel perspective. Standardized measurements of N$_2$O effluxes coupled with exhaustive metadata collection were performed at activated sludge tanks of three biological nitrogen removal WWTPs at different times of the year. The multivariate datasets were used as inputs for RF analyses. Computation of the permutation variable importance measures returned biomass-normalized dissolved inorganic carbon concentration (DIC$/VSS$/C$^0$/1) and specific ammonia oxidation activity (sOURAOB) as the most influential parameters determining N$_2$O emissions from the aerated zones (or phases) of activated sludge bioreactors. For the anoxic tanks, dissolved-organic-carbon-to-NO$_2$/NO$_3$/C$^0$/N$_+_+$NO$_3$/N$^{-1}$ ratio (DOC$/($NO$_2$/N$_+_+$NO$_3$/N$^{-1}$)) was singled out as the most influential. These data analysis results clearly indicate disparate mechanisms for N$_2$O generation in the oxic and anoxic activated sludge bioreactors, and provide evidences against significant contributions of N$_2$O carryover across different zones or phases or niche-specific microbial reactions, with aerobic NH$_3$/NH$_4^+$ oxidation to NO$_2$ and anoxic denitrification predominantly responsible from aerated and anoxic zones or phases of activated sludge bioreactors, respectively.

© 2020 Elsevier Ltd. All rights reserved.

**1. Introduction**

Nitrous oxide (N$_2$O) has recently attracted immense scientific interest amid the escalating concern for global warming and climate change. The global contribution of N$_2$O to the net greenhouse gas emissions is relatively minor (−6% in terms of CO$_2$eq) as compared to those of CO$_2$ (−76%) and CH$_4$ (−16%). Developing strategies for mitigating N$_2$O emissions is nevertheless crucial for curbing climate change, given its high global warming potential, −300 times greater than that of CO$_2$ in a 100-year scale (IPCC, 2014).

The contribution of the wastewater treatment sector to the net anthropogenic N$_2$O emissions used to be regarded insignificant at −3.5%. With the recent upward adjustment of the emission factor from 0.00035 to 0.016 kg N$_2$O−N/kg TN load, however, WWTPs would now be one of the most potent N$_2$O sources (IPCC et al., 2014; 2019). Further, a steady increase in water usage is projected in the foreseeable future, which would lead to further growth in global N$_2$O emissions from the wastewater treatment sector (Flörke et al., 2013). Recently, WWTPs are increasingly viewed as water resource recovery facilities (WRRFs) (Lam et al., 2020). Coping with N$_2$O emissions would undoubtedly be crucial for the transition of current WWTPs to carbon-neutral WRRFs, as N$_2$O is often considered as the largest contributor to life-cycle greenhouse gas emissions from WWTPs (Vasilaki et al., 2019).
Despite decades of investigation, the \( \text{N}_2\text{O} \)-producing mechanisms in biological nitrogen removal (BNR) wastewater treatment plants (WWTPs) are not yet fully understood. In the conventional BNR systems, where \( \text{NH}_4\text{-N} \) is removed via nitrification (\( \text{NH}_3/\text{NH}_4^- \text{to}-\text{NO}_3^- \) oxidation via \( \text{NO}_2^- \)) and denitrification (\( \text{NO}_3^- \text{to}-\text{N}_2 \) reduction via \( \text{NO}_2^- \), \( \text{NO}^- \), and \( \text{N}_2\text{O} \)), >90% of \( \text{N}_2\text{O} \) production originates from bioreactors where these N-dissimilation reactions take place (Kampschreur et al., 2009). Ammonia-oxidizing bacteria (AOB) and heterotrophic denitrifying bacteria are the microbial guilds deemed majorly responsible for \( \text{N}_2\text{O} \) production from nitrification and denitrification, respectively (Kampschreur et al., 2009). Several hypotheses regarding conditions leading to increased \( \text{N}_2\text{O} \) production from these microbial guilds have been suggested based on in vitro observations. High pHe, low dissolved oxygen (DO) concentration, high \( \text{NH}_4^+ \) and \( \text{NO}_2^- \) concentrations, and dissolved inorganic carbon (DIC) excess or deficiency have been suggested to increase \( \text{N}_2\text{O} \) formation from ammonia oxidation, and low C:N (or COD/N) ratio, low pH, \( \text{O}_2 \) intrusion, and high \( \text{NO}_2^- \) concentration have been hypothesized to increase \( \text{N}_2\text{O} \) release from incomplete denitrification (Itokawa et al., 2001; Pan et al., 2012; Peng et al., 2014; Peng et al., 2015a; b; Wunderlin et al., 2012). Whether \( \text{N}_2\text{O} \) emissions from BNR bioreactors are actually affected by these conditions, however, has not been sufficiently supported with datasets collected from full-scale operational WWTPs (Kosonen et al., 2016; Rodríguez-Caballero et al., 2015).

Over decades, several researchers have attempted to statistically identify the effectors with greatest influence on \( \text{N}_2\text{O} \) emissions from AS bioreactors (Aboobakar et al., 2013; Ahn et al., 2010; Vasilaki et al., 2018). Considering the practical difficulty of \( \text{N}_2\text{O} \) flux measurements in situ and the large numbers of parameters characterizing the AS operations in WWTPs, limited sample size \((n)\) and high dimensionality were the defining characters of the datasets used as inputs for these statistical analyses. The multiple regression approach and feature extraction approach taken in these previous analyses may not have been the optimal methodological approaches given the circumstances. Multiple regression fitting too many effectors to a single outcome may often result in noisy and even arbitrary interpretations due mainly to the curse of dimensionality, especially when the number of features far exceeds the sample size (Hastie et al., 2005; Krzywinski and Altman, 2015). Expansion of the datasets by compiling previously published \( \text{N}_2\text{O} \) identification from these microbial guilds have been suggested based on in vitro observations. High pHe, low dissolved oxygen (DO) concentration, high \( \text{NH}_4^+ \) and \( \text{NO}_2^- \) concentrations, and dissolved inorganic carbon (DIC) excess or deficiency have been suggested to increase \( \text{N}_2\text{O} \) formation from ammonia oxidation, and low C:N (or COD/N) ratio, low pH, \( \text{O}_2 \) intrusion, and high \( \text{NO}_2^- \) concentration have been hypothesized to increase \( \text{N}_2\text{O} \) release from incomplete denitrification (Itokawa et al., 2001; Pan et al., 2012; Peng et al., 2014; Peng et al., 2015a; b; Wunderlin et al., 2012). Whether \( \text{N}_2\text{O} \) emissions from AS bioreactors are actually affected by these conditions, however, has not been sufficiently supported with datasets collected from full-scale operational WWTPs (Kosonen et al., 2016; Rodríguez-Caballero et al., 2015).

Over decades, several researchers have attempted to statistically identify the effectors with greatest influence on \( \text{N}_2\text{O} \) emissions from AS bioreactors (Aboobakar et al., 2013; Ahn et al., 2010; Vasilaki et al., 2018). Considering the practical difficulty of \( \text{N}_2\text{O} \) flux measurements in situ and the large numbers of parameters characterizing the AS operations in WWTPs, limited sample size \((n)\) and high dimensionality were the defining characters of the datasets used as inputs for these statistical analyses. The multiple regression approach and feature extraction approach taken in these previous analyses may not have been the optimal methodological approaches given the circumstances. Multiple regression fitting too many effectors to a single outcome may often result in noisy and even arbitrary interpretations due mainly to the curse of dimensionality, especially when the number of features far exceeds the sample size (Hastie et al., 2005; Krzywinski and Altman, 2015). Expansion of the datasets by compiling previously published \( \text{N}_2\text{O} \) emissions data can easily be considered an option as the solution to this dilemma. Although no shortage of on-site \( \text{N}_2\text{O} \) emission monitoring campaigns have been conducted at pilot- and full-scale WWTPs with different configurations and operating conditions over a 10-month period from April 2018 to January 2019. The collected datasets were then utilized to compute the variable importance measures (VIM) of the AS parameters, i.e., the statistic indicators used for selection of the most influential parameters (Breiman, 2001).

### 2. Materials and methods

#### 2.1. Description of the full-scale wastewater treatment plants

The monitoring campaign was carried out at three WWTPs located near Gwangju, South Korea over a 10-month span from April 2018 to January 2019. The names and the exact locations of the WWTPs will not be disclosed due to confidentiality issues, and the WWTPs will be referred to as plants A, B, and C instead (Fig. 1, Table 1). The plant A is a conventional 8-train A2O (anaerobic-anoxic-oxic) WWTP with the return activated sludge (RAS) rate of 1,300 m³ h⁻¹ train⁻¹. Only one of the eight trains was monitored in the campaign. The plant B is an AO (anoxic-oxic) membrane bioreactor (MBR) that utilizes the membrane modules in the oxic tank to sustain high solids retention time (SRT) without RAS. Overflow from the oxic tank returns the AS to the anoxic tank, where denitrification takes place. The effluent is partially reused for in-facility and gardening purposes. The plant C is a four-step ([fill — aeration — sedimentation — draw]) sequencing batch reactor (SBR) with 2.37 h as the duration of each complete cycle. At each cycle, 250 m³ of 2,250 m³ mixed liquor in the bioreactor tank is withdrawn at the draw step and refilled with influent wastewater at the fill step of the following cycle. Heterotrophic denitrification takes place during the fill step, removing \( \text{NO}_2^- \) generated from nitrification of \( \text{NH}_4^+ \) in the oxic steps (the aeration and sedimentation steps) of the preceding cycle. No organic carbon supplement is added at any of the three WWTPs. Detailed operating parameters of the three WWTPs are presented in Table 1.

#### 2.2. Measurement of \( \text{N}_2\text{O} \) emissions from the WWTPs

The \( \text{N}_2\text{O} \) fluxes from the AS bioreactors were measured using a
floating flux chamber constructed according to the specification in the US EPA guideline (Chandran, 2011). The main body of the flux chamber was constructed with aluminum and glued to a rubber tire for flotation (the schematic diagram presented in Fig. S1). One of the five gas ports fitted on the aluminum dome was connected via PFA Teflon™ tubing to a Teledyne T320E infrared gas analyzer (San Diego, CA) for on-line measurement of N₂O in effluent gas. A gas port was used for streaming the tracer gas (10% v/v He in N₂; all gases were supplied by Daeduk Gas Co. Daejeon, Korea) through the chamber, and another port for supply of carrier gas (>99.999% N₂) to meet the minimum flowrate for the gas analyzer when measuring N₂O flux from quiescent bioreactors. Before each measurement, two-point calibration was performed using >99.999% N₂ gas and N₂ gas carrying 100 ppmv N₂O. The floating flux chamber was positioned at the geometric center of each bioreactor zone, where the measured N₂O flux would best represent the average flux from the zone (Chandran, 2011). Tracer gas was passed through the chamber at a flowrate of 1 L min⁻¹ and the He concentration in the effluent gas was measured with a HP-5890 gas chromatograph equipped with a thermal conductivity detector and Porapak Q column (Agilent Technologies, Palo Alto, CA). For N₂O flux measurements at the aerated zones or phases of AS bioreactors, N₂O and He concentrations in the off-gases exiting through the flux chamber were measured without aid of external carrier gas. For measurements of N₂O flux from quiescent tanks, high-purity N₂ sweep gas was blown into the headspace of the flux chamber at a constant flowrate of 5 L min⁻¹, and the N₂O and He concentrations in the effluent sweep gas were measured. For the plants A and B, the N₂O flux was monitored for 30 min at each zone with different aeration regime. The gas flowrates from the chambers placed on the aerated and quiescent zones or phases were calculated using Equations (1) and (2), respectively (Chandran, 2011).
Table 1 Description of the BNR processes at the examined WWTPs.

<table>
<thead>
<tr>
<th>Plant</th>
<th>A (A2O)</th>
<th>B (MBR)</th>
<th>C (SBR)</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tank volume per train</td>
<td>2.064</td>
<td>387</td>
<td>2,250</td>
<td>m³</td>
</tr>
<tr>
<td>Anaerobic</td>
<td>338</td>
<td>Anoxic</td>
<td>154</td>
<td></td>
</tr>
<tr>
<td>Anoxic</td>
<td>413</td>
<td>Oxic (aerated)</td>
<td>233</td>
<td></td>
</tr>
<tr>
<td>Oxic (aerated)</td>
<td>1,313</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of trains</td>
<td>B</td>
<td>2</td>
<td>2</td>
<td>–</td>
</tr>
<tr>
<td>External return activated sludge</td>
<td>370</td>
<td></td>
<td></td>
<td>m³ h⁻¹</td>
</tr>
<tr>
<td>Internal return activated sludge</td>
<td></td>
<td>250</td>
<td></td>
<td>m³ h⁻¹</td>
</tr>
<tr>
<td>Design HRT</td>
<td>Total: 6.7</td>
<td>Anoxic: 3.2</td>
<td>Fill: 0.12</td>
<td>h</td>
</tr>
<tr>
<td></td>
<td>Anoxic: 1.1</td>
<td>Oxic: 4.8</td>
<td>Aeration: 1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Oxic: 4.2</td>
<td>Sedimentation: 1</td>
<td>Draw: 0.25</td>
<td></td>
</tr>
<tr>
<td>Actual HRT</td>
<td>Total: 7.4</td>
<td>Anoxic: 6.1</td>
<td>Operates at design HRT</td>
<td>h</td>
</tr>
<tr>
<td></td>
<td>Anaoxic: 1.2</td>
<td>Oxic: 9.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Oxic: 4.7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Design SRT</td>
<td>12.1</td>
<td>30</td>
<td>20</td>
<td>d</td>
</tr>
<tr>
<td>Actual SRT</td>
<td>7.6</td>
<td>20.6</td>
<td>22.5</td>
<td>d</td>
</tr>
<tr>
<td>Aeration rate at theoxic zone/phase</td>
<td>27,200</td>
<td>66,400</td>
<td>56,600</td>
<td>m³ d⁻¹</td>
</tr>
<tr>
<td></td>
<td>16,000</td>
<td></td>
<td>4,500</td>
<td>m³ d⁻¹</td>
</tr>
</tbody>
</table>

\[
\begin{align*}
Q_e &= \frac{Q_{\text{tracer}} \times (C_{\text{He-tracer}} - C_{\text{He-fc}})}{C_{\text{He-fc}}} \quad (1) \\
Q_e &= \frac{Q_{\text{tracer}} \times (C_{\text{He-tracer}} - C_{\text{He-fc}})}{C_{\text{He-fc}}} - Q_{\text{sweep}} \quad (2)
\end{align*}
\]

where \(Q_e\) is the flowrate of the gas exiting the chamber (L³·T⁻¹); \(Q_{\text{tracer}}\) is the flowrate of the tracer gas introduced into the chamber (L³·T⁻¹); \(C_{\text{He-tracer}}\) and \(C_{\text{He-fc}}\) are the concentrations of He in the influent stream of the tracer gas and the effluent stream from the chamber, respectively (M·L⁻³); and \(Q_{\text{sweep}}\) is the flowrate of the sweep gas introduced into the chamber (L³·T⁻¹).

The \(N_2O\) concentration in the effluent gas from the chamber was recorded every minute and the average of the concentrations was taken as the representative value. As the plant C is an SBR with the different approach was taken for \(N_2O\) emission profile changing over time within each step, a slightly different approach was taken for the \(N_2O\) flux measurement. At the aeration and sedimentation steps, the \(N_2O\) concentration in the chamber effluent was monitored over the entire step (e.g., for the entire hour of the aeration step) and the concentrations were averaged for calculation of the representative \(N_2O\) flux value. Due to technical difficulty, the \(N_2O\) monitoring was not possible during the fill and draw steps. The on-line \(N_2O\) measurements were intermittently cross-checked with the \(N_2O\) concentrations measured with HP6890 Series gas chromatograph equipped with an electron capture detector (Agilent, Palo Alto, CA, USA). The \(N_2O\) flux was calculated using Equation (3).

\[
J_{N_2O} = \frac{Q_e \times C_{N_2O}}{A_w} \quad (3)
\]

where, \(J_{N_2O}\) is the \(N_2O\) flux (M·L⁻²·T⁻¹); \(C_{N_2O}\) is the representative effluent \(N_2O\) concentration (M·L⁻³); and \(A_w\) is the water surface area covered by the flux chamber (L²).

2.3. Determination of the specific ammonia oxidation activity of activated sludge samples

The AOB-specific oxygen uptake rate (sOUR\textsubscript{AOB}) was measured using a batch respirometric assay as an estimate of the NH\textsubscript{3} oxidation activity of anoxic AS sample (Chandran and Smets, 2005). The culture medium contained per liter, 500 mg KHCO₃, 272 mg KH₂PO₄, 300 mg MgSO₄·7H₂O, 180 mg CaCl₂·2H₂O, and 1 mL each of the trace element solutions I and II (van de Graaf et al., 1996). Two sets of triplicate reaction vessels were prepared per sample, with 30-mL culture medium distributed to 40-mL Plexiglass respirometry chambers (Unisense, Aarhus, Denmark). The respirometric assays performed with the triplicate vessels amended with 5 mg NH₄Cl·L⁻¹ estimated the sOUR\textsubscript{AOB}−NOB−HET value, i.e., the sum of sOURs owing to respiration of AOB, NOB (nitrite-oxidizing bacteria), and heterotrophs. The other set of triplicate vessels amended with 5 mg NO₂⁻·L⁻¹ was used to estimate the sOUR\textsubscript{NOB−HET} value. The vessels were vigorously bubbled with compressed air prior to addition of the AS biomass. The assays were performed on site with the AS samples collected during the \(N_2O\) flux measurements at the aerated zones or phases. The pellets collected from centrifugation of 30-mL samples were washed with 2 mM phosphate buffer solution (pH 7.0) and resuspended in 2 mL buffer solution. A 0.5-mL aliquot of the suspension was added to each of the triplicate respirometry chambers carrying aerated medium. The decrease in the DO (dissolved oxygen) concentration was monitored with a MicroRespiration system equipped with an OX-MR microelectrode (Unisense). For computation of the sOUR value, the maximum observed linear oxygen consumption rate within the first 30 min of incubation was normalized with the biomass calculated from the volatile suspended solids (VSS) concentration. The sOUR\textsubscript{AOB} value was obtained by subtracting sOUR\textsubscript{NOB−HET} from sOUR\textsubscript{AOB−NOB−HET}. No oxygen consumption was observed in the abiotic controls prepared identically but incubated without addition of AS biomass. All measurements were performed in triplicates and the average values were taken for further analysis. The standard deviations of the sOUR\textsubscript{AOB} values were calculated with the propagation of error method.

2.4. Physico-chemical characterization of collected wastewater samples

The water temperature, pH, and DO concentration of AS were measured with an Orion Star™ A329 portable multiparameter meter (Thermo Scientific, Waltham, MA) immediately before each \(N_2O\) flux measurement. Simultaneously, AS samples were collected and 100-mL subsamples were passed through 0.2-µm filters (Advantec, Inc., Tokyo, Japan). The filtrates were stored at −20 °C.
2.5. Feature selection with random forest analysis

The random forest (RF) method was used for feature selection from the collected datasets. The RF algorithm is an ensemble-learning algorithm, in which the final prediction is reached from the majority voting of the decisions made with multiple decision trees constructed with randomly permuted features and observations via recursive partitioning (Breiman, 2001). The data collected from the WWTP monitoring campaigns (Tables S2–S4) were subdivided into in-bag samples utilized as training sets (15 and 10 samples for aerated and anoxic AS, respectively) for growing the ‘random forest’ and out-of-bag (OOB) samples as test sets for evaluating the prediction performance (5 and 3 samples for aerated and anoxic AS, respectively). The random permutations of the in-bag observations and features were used to grow the forest, creating multiple decision trees with weakened intrinsic association between the observations and the features in the original dataset. The nodes of each individual tree were hierarchically structured to minimize the Gini impurity, i.e., maximize the information gain, at each bifurcation (Equation (4)) (Breiman, 2001).

\[
I_G = 1 - f_1^2 - f_2^2
\]

where \(I_G\) is the gini index for a binary split and \(f_i\) is the proportion of the samples classified as the \(i\) th group at the node.

Individual features were ranked according to their variable importance measures (VIM) (Strobl et al., 2007). The VIM value of each feature was computed following its permutation in each of the trees in the random forest (Equation (5)) (Janitza et al., 2013).

\[
\text{VIM}(X_j) = \frac{1}{n_{\text{tree}}} \sum_{t=1}^{n_{\text{tree}}} [E_t(X_j) - E_t]
\]

where, \(\text{VIM}(X_j)\) is the variable importance measure of the feature \(X_j\), \(n_{\text{tree}}\) is the number of trees in the forest, \(E_t\) and \(E_t(X_j)\) are the misclassification rates of the tree \(t\) before and after permuting \(X_j\), respectively, as tested with the OOB bag samples.

The RF classification was performed using the randomForest function of the randomForest package in RStudio v. 1.2.5 (Liaw and Wiener, 2012). For binary classification, the response variable (N\(_2\)O flux) was discretized to categorical attributes with the threshold set to the median value (303 and 163 mg N\(_2\)O–N m\(^{-2}\)d\(^{-1}\) for aerated and anoxic zones or phases, respectively), such that the outcome of the RF prediction can be classified as either ‘high’ or ‘low’ N\(_2\)O emission. The variables for construction of individual trees were randomly selected by bootstrapping with replacement. The hyperparameters were set as follows: \(n_{\text{tree}} = 500\), \(mtry = \text{default}\) (the square root of the number of variables), \(\text{metric} = \text{accuracy}\), and \(\text{importance} = \text{true}\). The classification performance was evaluated with the accuracy (Equation (6)), sensitivity (Equation (7)), and specificity (Equation (8)) presented in a confusion matrix (Powers, 2011). The VIM values were computed using the function \(\text{varimp}\) with the scale and type parameters set to false and 1, respectively.

\[
\text{Accuracy} = \frac{\# \text{ of correct predictions (TH + TL)}}{\text{Total} \# \text{ of predictions (TH + TL + FH + FL)}}
\]

\[
\text{Sensitivity} = \frac{\# \text{ correctly predicted as 'high' (TH)}}{\text{Total} \# \text{ predicted as 'high' (TH + FL)}}
\]

\[
\text{Specificity} = \frac{\# \text{ correctly predicted as 'low' (TL)}}{\text{Total} \# \text{ predicted as 'low' (TL + FH)}}
\]

where, TH, TL, FH, and FL stand for true high, true low, false high, and false low, respectively.

2.6. Statistical tests and correlation analyses

All statistical analyses were performed using RStudio v. 1.2.5 (Rstudio Team, 2015). Given that only a few of the AS parameters fully satisfied the assumptions for parametric statistical tests (e.g., normality and homoscedasticity), all statistical tests were performed using non-parametric statistical tests. Statistical significances of pairwise comparisons were determined with the Mann-Whitney test (Fay and Proschan, 2010). The Kruskal-Wallis test was used when three measurements were compared (Kruskal and Wallis, 1952). Bivariate associations between the AS parameters were evaluated using tie-corrected Spearman tests (Dodge, 2008).

3. Results and discussion

3.1. Long-term monitoring of BNR WWTP performances and N\(_2\)O emissions

Twenty sets of data including the operating parameters, AS characteristics, and N\(_2\)O flux records were collected from the three WWTPs during the monitoring period (Tables S1–S4). The dissolved organic carbon (DOC) and total nitrogen (TN) removal performances were inconsistent, with the removal efficiencies ranging between 7.8–57.9% and 17.0–95.5%, respectively. As with the removal efficiencies, many of the operating parameters and AS characteristics critical for proper functioning of the BNR systems exhibited large temporal fluctuations, as well as variations across the WWTPs. The DO, DOC, DIC, TN, NH\(_4\)-N, and NO\(_3\)-N concentrations in all examined AS bioreactors exhibited temporal fluctuations amounting to percent relative ranges larger than 150% over the ten-month span of the monitoring campaign. The only parameter with statistically significant difference across the WWTPs was the VSS concentration. Activated sludge sample from the aerated zone of plant B had significantly higher (\(p = 0.00\)) VSS concentration (7050 ± 1320 mg VSS L\(^{-1}\)) than aerated AS from the plants A (1660 ± 250 mg VSS L\(^{-1}\)) and C (2110 ± 216 mg VSS L\(^{-1}\)). The anoxic zone of the plant B (3020 ± 1010 mg VSS L\(^{-1}\)) also had significantly higher VSS concentrations (\(p = 0.03\)) than the anoxic zone of the plant A (1260 ± 500 mg VSS L\(^{-1}\)). The prolonged monitoring period enabled incorporation of temperature as an independent variable in the statistical analyses, as the water temperature varied between −15 and −30 °C at the WWTPs during the period.

The N\(_2\)O flux data were collected from the AS bioreactors of the BNR systems (Fig. 2). The three BNR systems exhibited substantially different N\(_2\)O emission profiles. At the full-scale A2O plant (plant A)
significantly larger N₂O fluxes were measured at the aerated zone than at the anoxic zone on all eight occasions, indicating that aerobic nitrification was majorly responsible for N₂O emissions. Contrastingly, the N₂O flux was significantly higher at the anoxic zone than the aerated zone on three out of five occasions at the plant B, suggesting that contribution of denitrification to N₂O emissions was substantial at this MBR plant. The effect of temperature was ambivalent from the N₂O emissions data. More intensified N₂O emission was recorded in the warmer months (August to October) than in the colder months (April and November to January) at the plant A. Contrastingly, the largest measure of N₂O emissions were recorded in winter (December) at the plant C. No clear seasonal distinction was observed at the plant B, where the water temperature of the underground bioreactors was less affected by the outdoor air temperature. Although statistical comparisons across the WWTPs do not tell much due to the large standard deviations, both the maximum and the median values for the N₂O fluxes from bothoxic and anoxic AS bioreactors were substantially higher at the plant A than at the plants B and C.

Previously, the larger N₂O emission figures often observed from the aerated AS were attributed to carryover of dissolved N₂O produced from the anoxic counterpart, as well as N₂O production from nitrification (Sun et al., 2013; Toyoda et al., 2010). The N₂O flux measurements at the plant B in the months of September and October suggested such carryover of N₂O was not likely. The N₂O fluxes measured at the anoxic zone were substantial (98.5 ± 15.1 mg N₂O-N m⁻²d⁻¹ and 701.5 ± 140.2 mg N₂O-N m⁻²d⁻¹ in September and October, respectively) during this period indicating N₂O production in the anoxic AS. The near absence of N₂O flux from the aerated tank at the same time points suggested that N₂O carryover from the anoxic zone was not a substantial contributor to N₂O emissions from the aerobic zone. The inconsistent ratio of N₂O flux from the anoxic tank to that from the aerated tank at plant A (0.086) also suggested insignificance of N₂O carryover from the anoxic zones to the aerated zones.

3.2. Random forest analyses and identification of key variables affecting N₂O emission fromoxic and anoxic tanks

Random forest analyses were implemented to rank the nineteen variables (listed in Table S5) in accordance to their impacts on N₂O emissions from the anoxic and anoxic AS of the BNR WWTPs (Fig. 3). The first round of RF analysis with the dataset obtained from the aerated AS screened out ten redundant variables with negative VIM values (temperature, pH, DO, DOC, NH₄-N, NO₂-N, NO₃-N, DO·VSS⁻¹, TN·VSS⁻¹, NO₂-N·VSS⁻¹, and NO₃-N·VSS⁻¹). The second round of RF analysis using the remaining nine variables identified DIC·VSS⁻¹ and sOURₐ₀ as the most influential variables. The RF analysis with the dataset obtained from the anoxic AS identified DOC·(NO₂-N + NO₃-N)⁻¹ as the most influential parameters affecting N₂O emissions, out of the four variables with positive VIM values in the first round of RF analysis. The RF classification performance was evaluated by constructing a 2 × 2 confusion matrix and computing the evaluation metrics (Table S6). The accuracy, sensitivity, and specificity values were 0.70, 0.75, and 0.67, respectively, for dichotomous categorization of the aerate AS datasets to high or low N₂O emissions. For anoxic AS datasets, the same calculations returned 0.77, 0.80, and 0.75, respectively. These metrics supported that the RF classifications were reliable, given the limited data size and the temporally and spatially heterogeneous nature of the WWTPs, which allowed for only limited accuracy in the measurements. The variables identified with high VIM all significantly correlated with the N₂O fluxes from the respective zones or phases of AS bioreactors (Table S7). The N₂O fluxes from the aerated AS were significantly associated with the DIC·VSS⁻¹ (r = 0.49; p =
0.03) and sOURAOB ($r = 0.47; p = 0.03$). Similarly, the DOC (NO$_2^{-}$N + NO$_3^{-}$N)$^{-1}$ values were significantly associated with N$_2$O fluxes ($r = -0.62; p = 0.02$) from the anoxic AS.

### 3.3. Analyses of the key variables

The two parameters identified as the most influential determinants of N$_2$O emissions from the aerated AS bioreactors, DIC-VSS$^{-1}$ and sOURAOB, were significantly correlated ($r = 0.50; p = 0.03$), suggesting that DIC availability has significant positive impacts on NH$_3$ oxidation activity and N$_2$O production thereof (Table S7). Only a few studies have examined the relationship between DIC availability and N$_2$O production from nitrifying isolates or consortia and the results of these studies were largely contradictory (Jiang et al., 2015; Mellbye et al., 2016; Peng et al., 2015b). Jiang et al. (2015) and Mellbye et al. (2016) reported increased N$_2$O emissions at DIC-limiting conditions in axenic *Nitrosomonas europaea* cultures. The response of a complex nitrifying culture enriched from AS to varied DIC supply, as reported by Peng et al. (2015b), was contradictory to these pure culture results, in that a positive linear correlation was observed between DIC concentration and N$_2$O emissions within the range of DIC concentration between 4.8 and 56 mg C L$^{-1}$. The observations in the current study generally agrees with the latter, as the positive correlation between DIC availability and N$_2$O flux could be best explained with elevated N$_2$O-producing AOB activity at high DIC availability. The range of DIC concentrations in the aerated AS, 12.2–49.6 mg C L$^{-1}$, was sufficiently high not to cause carbon limitation to AOB, as the estimate of inorganic carbon needed for AOB growth per 1 mg NH$_3$-N L$^{-1}$ is 77 µg C L$^{-1}$ (van Loosdrecht et al., 2016). Nevertheless, mechanistic or physiological explanation of the observed correlations is not possible at current stage, due to the lack of information regarding the effect of excess DIC on AOB physiology.

Previously, nitrifier denitrification and heterotrophic denitrification occurring upon oxygen deficiency had been regarded as the major mechanisms of N$_2$O emissions from the aerated AS (Kampschreur et al., 2009; Tallec et al., 2006). That sOURAOB was selected as one of the variables most influential to N$_2$O flux from the aerated AS and DO concentration was an insignificant parameter suggested that the majority of N$_2$O produced from the aerated AS bioreactors originated from biological NH$_3$ oxidation at oxygen-replete condition, on the contrary to the previous perception. The aerated AS bioreactors were generally maintained at high DO concentrations (the average of 4.3 ± 2.2 mg L$^{-1}$ over all 20 measurements taken at the three facilities), and DO concentrations below 2.0 mg L$^{-1}$ were only recorded twice (April and August) at the aerated zone of the plant B (MBR). Neither heterotrophic denitrification nor nitrifier denitrification activity was likely to be substantial at such well-oxygenated conditions (Tallec et al., 2006). Thus, the major N$_2$O production mechanism from the aerated AS reactors was presumably aerobic N$_2$O production by AOB. Although moderate in quantity as compared to nitrifier denitrification at oxygen-limited condition, AOB has been known to produce N$_2$O as a byproduct of NH$_3$ oxidation in oxygen-replete conditions (Jiang and Bakken, 1999). Several recent studies have suggested involvement of hydroxylamine (NH$_2$OH), the reactive intermediate of NH$_3$ oxidation to NO$_2$-N in the oxidative N$_2$O production by AOB (Liu et al., 2017; Song et al., 2020). Abiotic oxidation and n-nitrosation, as well as incomplete oxidation of NH$_2$OH by hydroxylamine oxidoreductase with nitric oxide (NO) as the intermediate, have been suggested as the mechanisms of aerobic production of N$_2$O as a byproduct of NH$_3$ oxidation (Liu et al., 2017; Terada et al., 2017). N$_2$O emitted from the aerated AS appears to have originated from these aerobic reactions previously regarded as minor N$_2$O sources from nitrification.

The parameter with the largest impact to N$_2$O flux from the anoxic AS bioreactors of the plant A and B, according to the RF analysis, was DOC (NO$_2^{-}$N + NO$_3^{-}$N)$^{-1}$. The high VIM value of the parameter and its significant negative Spearman’s rank correlation (Table S7; $r = -0.62; p = 0.02$) with the N$_2$O flux both supported that heterotrophic denitrification was likely the major N$_2$O-producing mechanism in the anoxic AS. Denitrification is carried out in multiple steps, and due to its modular nature, each step may be carried out by different guilds of organisms in a complex microbial consortia such as AS (Graf et al., 2014). In the environments with low electron donor availability, as signified by the low values of DOC (NO$_2^{-}$N + NO$_3^{-}$N)$^{-1}$ or more conventional C/N or COD/N ratio, limited electrons may be unevenly distributed to the different steps of denitrification (Pan et al., 2013). The intensified competition for limited electrons may result in increased N$_2$O production if electron donors are preferably consumed by the organisinal groups responsible for the steps of denitrification upstream of N$_2$O. Experimental results both supporting and contradicting this hypothesis have been reported over past few decades (Itokawa et al., 2001; Kishida et al., 2004; Lu and Chandran, 2010; Pan et al., 2013). The more recent reports generally dispute against the significance of electron donor limitation to N$_2$O emissions, with basis on the batch kinetics of denitrifying AS consortia provided with a single

![Unscaled variable importance measure (VIM) values as computed from the second round of random forest analyses performed with the datasets acquired from the (a) aerated and (b) anoxic zones or phases of the examined WWTPs. Only the variables that returned positive VIM values in the first rounds of the analyses were examined. The variables were rearranged in the order of decreasing VIM values.](image)
organic electron donor. For interpretation of the current result, however, dated results from less controlled reactor experiments, in which COD/N ratios turned out to be critical determinants of N₂O emissions, appear to be more applicable due to the following reasons. (1) Neither plant A nor B use organic supplements for nitrogen removal enhancement; therefore, the source of labile organic electron donor, presumably a small fraction of DOC, is of complex makeup. Different groups of denitrifiers favor different organic compounds as the sources of electrons, and thus, the kinetics of the denitification steps in AS with complex organics may substantially differ from those observed with single organic species, e.g., methanol (Lu and Chandran, 2010; Pan et al., 2013). (2) The dissolved N₂O concentrations used in the batch kinetic assays for determining the kinetic parameters may be much lower than the linear N₂O reduction rates observed in the kinetic assays (Yoon et al., 2016).

4. Conclusion

Simultaneous monitoring of AS properties, WWTP performance, and N₂O flux at a full-scale WWTP is an onerous and time-consuming endeavor. Unless an automated monitoring system is already installed at the study site, the quantity of data obtainable from in situ monitoring would be always limited as relative to the number of manifold potential effectors. Not many researchers are aided with such luxury, as also was the case in this study, in which only 13–20 datasets were in hand for statistical analyses of the association between AS parameters and N₂O fluxes. Although several successful cases of fitting bench-scale reactor data or short-term (several days at the longest) to modified activated sludge models (ASM) have previously been reported, the irregularity and sheer magnitudes of fluctuations in the N₂O flux data collected in this study obviously indicated impossibility of developing a universal mathematical model estimating the N₂O emissions from full-scale AS bioreactors at different oxygen regimes (Ni et al., 2013; Spérandio et al., 2016). The approach taken here instead, the machine-learning based RF analysis, allowed for reliable binary categorization of aerated and anoxic AS conditions to high or low N₂O-emitting conditions. Besides, the WWTP parameters identified as the most influential effectors supported that the overarching microbial metabolism responsible for N₂O emissions from exoc and anoxic AS are NH₃ oxidation and denitrification, respectively.

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 project is supported by the “R&D Center for reduction of Non–CO₂ Greenhouse gases (2017002420002)” funded by Korea Ministry of Environment (MOE) as “Global Top Environment R&D Program”.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.watres.2020.116144.

References

Liu, S., Han, P., Hink, L., Prosser, J.I., Wagner, M., Brüggemann, N., 2017. Abiotic conversion of extracellular NH₄OH contributes to N₂O emission during