Influence of sampling strategies on the estimated nitrous oxide emission from wastewater treatment plants

Matthijs R.J. Daelman\textsuperscript{a,b, *}, Bernard De Baets\textsuperscript{c}, Mark C.M. van Loosdrecht\textsuperscript{b,d}, Eveline I.P. Volcke\textsuperscript{a}

\textsuperscript{a} Department of Biosystems Engineering, Ghent University, Coupure Links 653, 9000 Gent, Belgium
\textsuperscript{b} Department of Biotechnology, Delft University of Technology, Julianalaan 67, 2628 BC Delft, Netherlands
\textsuperscript{c} Department of Mathematical Modelling, Statistics and Bioinformatics, Ghent University, Coupure Links 653, 9000 Gent, Belgium
\textsuperscript{d} KWR Watercycle Research Institute, P.O. Box 1072, 3433 PE Nieuwegein, Netherlands

\textbf{Article info}

Article history:
Received 27 October 2012
Received in revised form 6 March 2013
Accepted 10 March 2013
Available online 21 March 2013

Keywords:
Greenhouse gas
Monitoring
Nitrous oxide
Sampling strategy
Wastewater treatment

\textbf{Abstract}

In the last few years, the emission of nitrous oxide from wastewater treatment plants has become a topic of increased interest, given its considerable impact on the overall climate footprint of wastewater treatment plants. Various sampling strategies to estimate nitrous oxide emission from wastewater treatment plants have been applied in different studies. The present study addresses the influence of sampling strategies on the estimated emission by analysing the variability of an extensive dataset of nitrous oxide emissions resulting from a long-term online monitoring campaign at a full-scale municipal wastewater treatment plant. It is shown that short-term sampling is inadequate to accurately estimate the average nitrous oxide emissions from a particular wastewater treatment plant, while online monitoring is indispensable to capture the short-term variability (diurnal dynamics).

© 2013 Elsevier Ltd. All rights reserved.

\section{Introduction}

Wastewater treatment plants (WWTPs) are known as emission sources of the greenhouse gas nitrous oxide (N$_2$O) (Hanaki et al., 1992; Kampschreur et al., 2009; Desloover et al., 2012; Law et al., 2012). Nitrous oxide is expected to be emitted during biological nitrogen removal from wastewater, through nitrification and subsequent denitrification (Kampschreur et al., 2009). In one study, nitrous oxide was found to make up 88% of the emitted carbon dioxide equivalents of a particular WWTP (STOWA, 2010), while Daelman et al. (in press) established the share of nitrous oxide at the same WWTP as 78% based on the long-term, online monitoring campaign that yielded the dataset that is used in the present study. Since nitrous oxide has a global warming potential of 298 CO$_2$-equivalents over a 100 year time horizon, even a low emission contributes significantly to a WWTP’s greenhouse gas footprint (IPCC, 2007).

Table 1 provides an overview of the reported monitoring studies of full-scale municipal wastewater treatment plants and the applied sampling strategies. The various strategies differ in aspects such as the duration of the sampling campaign (ranging from a single day to 1.5 years) and the sampling frequency (ranging from a single grab sample to online sampling). The nitrous oxide emissions obtained from these studies display a huge variability, both over time and...
Overview of sampling strategies for the determination of nitrous oxide emissions from full-scale WWTPs, applied in previous studies.

<table>
<thead>
<tr>
<th>Study</th>
<th>Sampling strategy (frequency and duration)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Czepiel et al. (1995)</td>
<td>Weekly grab samples from one WWTP during 15 weeks</td>
</tr>
<tr>
<td>Wicht and Beier (1995)</td>
<td>Single grab samples from 25 WWTPs</td>
</tr>
<tr>
<td>Sümer et al. (1995)</td>
<td>Biweekly grab samples from one WWTP over one year</td>
</tr>
<tr>
<td>Kimochi et al. (1998)</td>
<td>Online over four 2 h aeration cycles of a single WWTP</td>
</tr>
<tr>
<td>Sommer et al. (1998)</td>
<td>Biweekly grab samples from one WWTP over 1.5 years</td>
</tr>
<tr>
<td>Peu et al. (2006)</td>
<td>Online over one/three weeks from two WWTPs</td>
</tr>
<tr>
<td>STOWA (2010)</td>
<td>2 WWTPs: online over one week</td>
</tr>
<tr>
<td>Ahn et al. (2010a,b)</td>
<td>Online over 24 h at 12 different WWTPs</td>
</tr>
<tr>
<td>Foley et al. (2010)</td>
<td>Grab samples taken in the morning and afternoon of two consecutive days for 7 WWTP</td>
</tr>
</tbody>
</table>

Table 1 – Overview of sampling strategies for the determination of nitrous oxide emissions from full-scale WWTPs, applied in previous studies.

between different WWTPs. In their review, Kampschreur et al. (2009) reported nitrous oxide emissions ranging from 0 to 14.6% of the nitrogen load. During a study in the Netherlands (STOWA, 2010), the emission of nitrous oxide was observed at three WWTPs. Two plants were monitored during one week each, while the third plant was monitored during one week in October 2008 and one week in February 2009. The emission showed a variability of 0.040–6.1% of the incoming nitrogen between the WWTPs, and of 0.42–6.1% between October and February respectively, for the same WWTP. Ahn et al. (2010b) also demonstrated the short-term variability of nitrous oxide emission over a single day.

The influence of different sampling strategies on the reported nitrous oxide emission of different studies has not been assessed up till now. It can, however, be reasonably expected that the sampling strategy will influence the estimated total nitrous oxide emission from such a dynamic process. In the field of wastewater treatment, the challenge of monitoring dynamic phenomena has also been addressed by Ort and Gujer (2006) and Ort et al. (2010a,b) in the case of pharmaceuticals and personal care products in sewer systems, while Gevaert et al. (2009) discussed the monitoring of the substances on the priority list of the European Water Framework Directive. In these studies, monitoring scenarios are tested by applying different sampling modes to a simulated dataset that results from dynamic models describing the presence of pollutants in sewers and water bodies. In the present study, a similar approach is taken, but instead of using simulated data, this contribution compares and evaluates the reported monitoring strategies for nitrous oxide emissions by applying strategies from previous studies to the dataset of a long-term, online measuring campaign.

The objective of this study is to expose the caveats to be expected during monitoring of nitrous oxide emission from wastewater treatment by illustrating the effect of sampling strategies on the estimation of the emission. These days, numerous research groups are engaging in sampling campaigns in order to quantify the nitrous oxide emission and to identify the mechanisms behind it. Often, the complexity and the cost of sampling limit these campaigns to a small number of samples or to a short measurement period. The present study demonstrates that results obtained from such limited measurements lead to unreliable estimates of the amount of nitrous oxide emitted.

2. Materials and methods

2.1. The long-term, online dataset

To assess the influence of monitoring strategies on the estimated nitrous oxide emissions from WWTPs, several monitoring scenarios have been applied to an extensive dataset of a long-term online monitoring campaign at a completely covered WWTP. The nitrous oxide mass flow rate was calculated from measurement data of the concentration and volumetric gas flow rate of the off-gas coming from the covered activated sludge tanks. The sampling protocol is added as supplementary material. The dataset contained 23,280 data points of the total nitrous oxide emission from the plant and covered the entire measurement period from October 14, 2010 to January 26, 2012, with a month-long interruption in October 2011 due to a technical failure. In total, the dataset covered 416 days, one data point being available for every 25 min during this period. By integrating all the data points of the entire dataset, an estimate of the plant’s total emission was obtained: 91 kg N2O–N d−1. This value is further referred to as the true average emission.

The emission pattern shows a distinct variability, both on the long term (seasonal) and on the short term (diurnal). Fig. 1 shows the seasonal variability of the daily nitrous oxide emission. In October and November 2010 the emission is lower than 10 kg N2O–N d−1. November 2010 has even two weeks during which no nitrous oxide is emitted at all. In December 2010, the emission starts to increase to reach a

![Fig. 1 – Daily nitrous oxide emission from 14 October 2010 to 26 January 2012 (the gap in October 2010 is due to equipment downtime).](image-url)
maximum daily emission of 455 kg N\textsubscript{2}O–N d\textsuperscript{−1} in March 2011. From April onwards, the emission starts to decrease again. In the autumn and early winter of 2011–2012, it is hard to distinguish any trend.

Fig. 2 shows the diurnal variability during one week in December 2010. During this week, the emission peaked at midnight and was lowest in the morning. The results of this week were chosen as a mere illustration of the diurnal variability that could possibly be encountered. It should be noted that the diurnal emission pattern during this week in December 2010 is not at all representative for the entire period.

2.2. Implementation of different sampling strategies

The long-term online dataset that came forth from this monitoring campaign was subjected to different sampling strategies based on the ones that have been used in the previous studies mentioned in Table 1. The evaluated sampling strategies are presented in Table 2.

2.2.1. 24 h online sampling

A 24 h online sampling strategy was applied by Ahn et al. (2010b). The exact sampling frequency is not mentioned. Presumably, such a monitoring campaign would last e.g. from 11 a.m. on a working day until 11 a.m. the next working day. The average over this 24 h period would yield an estimated daily emission flow rate (kg N\textsubscript{2}O–N d\textsuperscript{−1}). This strategy was applied to every 24 h period of the present online long-term dataset, starting at 11 a.m. on a Monday, a Tuesday, a Wednesday or a Thursday (it was presumed that a WWTP cannot be accessed during the weekend, so a measurement campaign has to start and end on a weekday). Of each such period, the average emission was calculated, resulting in 237 calculated values of the daily nitrous oxide emission of the plant. Each of these 237 emission estimates is a possible outcome of a 24 h online sampling campaign. All the possible outcomes were collected in a histogram with a bin width of 10 kg N\textsubscript{2}O–N d\textsuperscript{−1} (Fig. 6).

2.2.2. 7 Day online sampling

Online sampling campaigns lasting for one week were carried out by Peu et al. (2006) and by STOWA (2010). By averaging the emission over seven days of online sampling, an estimate of the daily emission (kg N\textsubscript{2}O–N d\textsuperscript{−1}) is obtained. For the present study, this strategy was applied to every period of 7 consecutive days of the complete dataset, excluding the 7 day periods that started on Saturday or Sunday (again it was presumed that a WWTP cannot be accessed during the weekend, so also a 7 day online measurement campaign has to start and end on a weekday). 7-day moving averages over the long-term online dataset were calculated, starting and ending at 11 a.m., excluding the 7 day periods that started on Saturday or Sunday, resulting in 289 calculated values of the daily nitrous oxide emission of the plant. Similarly to the previous case, each of these 289 emission estimates is a possible outcome of a 7 day online sampling campaign. All the possible outcomes were collected in a histogram with a bin width of 10 kg N\textsubscript{2}O–N d\textsuperscript{−1} (Fig. 6).

2.2.3. Long-term weekly grab sample

Czepiel et al. (1995), Sümer et al. (1995) and Sommer et al. (1998) took (bi)-weekly grab samples over a longer period (fifteen weeks, 1 year and 1.5 year, respectively). According to this sampling scheme, one takes a single grab sample on a specific day of the week, e.g. each Monday, at some time between 9 a.m. and 5 p.m., during several weeks or longer. The average of all the grab samples is the final estimate of the average emission. This long-term weekly grab sampling strategy was implemented by randomly selecting a daytime emission value from the long-term online dataset for each working day of the week, mimicking five long-term weekly grab sampling campaigns of 58, 59 or 60 weeks (not all days of the week were equally represented in the long-term online dataset due to interruptions in the online sampling campaign), each of which was performed on a specific day of the week. For all days of the week of the dataset, these 58, 59 or 60 randomly selected values were averaged per day, resulting in seven calculated values of the daily emission, one for each day of the week. With about 20 data points available between 8 a.m. and 5 p.m. each day, and 58–60 days per sampling campaign, the possible ways in which to conduct a long-term weekly sampling campaign are all but infinite (2\textsuperscript{206}). In contrast to the three previous strategies, it is therefore nearly impossible to obtain the full number of possible outcomes for each of the seven weekly sampling campaigns. Instead, the procedure was repeated 1000 times for each day, so each repetition yielded 1000 calculated values per day of the week. This procedure was also performed for a random working day per sampling week, mimicking a monitoring campaign during which a weekly sample is taken on a random working day.
instead of a fixed day of the week. This procedure was repeated 1000 times as well. For each of the seven sampling campaigns using a fixed day and for the sampling campaign using a random weekday, the 1000 values are possible outcomes of the respective sampling strategies. All the possible outcomes were collected in a histogram with a bin width of 10 kg N₂O–N d⁻¹ (Fig. 7).

2.2.4. Single grab sample
Wicht and Beier (1995) determined the nitrous oxide emission of 25 WWTPs by taking a single grab sample at each plant. To evaluate grab sampling, every value of the online dataset that was logged on working days between 9 a.m. and 5 p.m. is considered a potential value of a grab sample, since grab samples are usually taken during working hours. This resulted in 5504 possible estimates. Again, each of these 5504 emission estimates is a possible outcome of a sampling campaign consisting of a single grab sample. All the possible outcomes were collected in a histogram with a bin width of 10 kg N₂O–N d⁻¹ (Fig. 6).

2.3. Precision as function of sampling campaign duration

For each of the investigated sampling strategies, the relative error between the true average emission and the simulated estimate is determined in function of the number of sampling instances, similarly as Ort and Gujer (2006). For the 24 h and the 7 day online sampling strategy it concerns the number of random, not necessarily consecutive 24 h periods; or random, not necessarily consecutive weeks of online monitoring, respectively. For the long-term weekly grab sampling strategy it is determined how many weeks such a campaign should last to yield a satisfying estimate. Finally, the number of completely random grab samples that is required, is determined as well.

2.3.1. 24 h online sampling

From the collected dataset of 24 h averages that start and end on 11 a.m. on a working day, containing 237 values (cf. Section 2.2), a number of 24 h averages could be selected randomly. This number is called \( n \), and it does not necessarily concern consecutive periods. These \( n \) 24 h averages represent a subset of the long-term online dataset. Fig. 3 illustrates the way in which a sampling scheme of \( n = 20 \) randomly picked 24 h periods is mimicked. The average of the entire dataset represents the true average nitrous oxide emission from the WWTP, while the average of the subset of 20 randomly picked sampling days represents a possible estimate of the true average nitrous oxide emission.

This procedure was repeated for every number \( n \) of sampling periods (from 1 to 237, cf. Section 2.2), resulting in a different simulated estimate for any value of \( n \). For each value of \( n \), the relative error \( \varepsilon \) was calculated as

\[
\varepsilon = \frac{\text{Simulated average emission} - \text{True average emission}}{\text{True average emission}}
\]  

(1)

When \( \varepsilon \) is plotted against \( n \), a graph like in Fig. 4 is obtained. Per value of \( n \), this was repeated 1000 times. This yielded a distribution of the relative errors as a function of the number of sampling periods, \( \varepsilon(n) \). Fig. 5 shows this distribution of \( \varepsilon \) for \( n = 20 \).

The average \( \mu \) and the standard deviation \( \sigma \) of this distribution were determined for each value of \( n \), \( \mu - 2\sigma \) and \( \mu + 2\sigma \) are uncertainty bounds for \( \varepsilon \). If the 1000 instances of \( \varepsilon \) for a certain \( n \) are normally distributed (such as in Fig. 5), this interval is a 95% confidence interval around the mean. For each \( n \), the normality of \( \varepsilon \) was verified using the Kolmogorov–Smirnov test.

2.3.2. 7 day online sampling

The precision of the estimate as a function of the number of online sampling weeks was evaluated similarly as for the number of 24 h online sampling periods. In total, 289 moving averages were calculated, one for each period of seven consecutive days starting and ending at 11 a.m. on a weekday.
(cf. Section 2.2). A monitoring campaign could consist of any number of random non-overlapping weeks during which the nitrous oxide emission was measured online, up to a maximum of 58, the maximum number of non-overlapping periods starting and ending at 11 a.m. on a weekday in the long-term, online dataset. As in the case of 24 h online sampling, all the data points over the 7 day periods are averaged for each n. This average is an estimate of the plant’s total average emission. For each value of n, the relative error ε was calculated according to eq. (1). For each value of n, this was repeated 1000 times, resulting in 1000 simulated estimates and relative errors per value of n. Again, this yielded a distribution of the relative errors as a function of the number of sampling periods, ε(n). The average μ and the standard deviation σ of this distribution were determined for each value of n. The uncertainty bounds for ε are μ − 2σ and μ + 2σ. If the 1000 instances of ε for a certain n are normally distributed, this interval is a 95% confidence interval. For each n, the normality of ε was verified using the Kolmogorov–Smirnov test.

2.3.3. Long-term weekly grab sampling

Long-term weekly grab sampling was mimicked without the restriction that the sample was taken on a specific day of the week. Such a strategy would take a grab sample in every consecutive week, no matter what day of the week, during n weeks. To simulate this, the daytime values of the weekdays were pooled per week. Next, n weekly grab sampling campaigns of 1 to n weeks were simulated by randomly selecting a value from the weekly pooled data for n consecutive weeks, starting with a random week. However, when n consecutive weeks are randomly chosen from 62 weeks, the first and the last n − 1 weeks have a lower probability to be selected than the weeks in between. To avoid this misrepresentation, any series of n consecutive weeks that started less than n weeks before the end of the dataset was completed with the required number of consecutive weeks at the beginning of the dataset. This ensures that the weeks in the beginning and the end of the long-term dataset have an equal chance to be randomly chosen.

For each value of n, the n weekly emission values were averaged, resulting in 62 estimates. For each estimate, the relative error ε was calculated according to eq. (1). Again, this procedure was repeated 1000 times, resulting in 1000 simulated estimates and 1000 values of ε for each of the 62 values of n. For each n, the average μ and the standard deviation σ of the 1000 relative errors were determined, with μ − 2σ and μ + 2σ as uncertainty bounds. Again, the Kolmogorov–Smirnov test was used to determine for which values of n the 1000 relative errors were normally distributed. In the latter case, μ − 2σ and μ + 2σ are the bounds of a 95% confidence interval.

2.3.4. Completely random grab sampling

The long-term online dataset contained 5504 values that were logged on working days between 9 a.m. and 5 p.m. All these values are possible results of grab samples. A grab sampling campaign could consist of one to any number of completely random grab samples, but from a practical point of view, the maximum number of grab samples is limited to n = 200. To mimic n completely random sampling campaigns, n random values were picked from the 5504 daytime values of the long-term online dataset. For each n, the n grab samples were averaged, resulting in 200 simulated estimates of the nitrous oxide emission, one for each value of n. 1000 repetitions of this procedure result in 1000 simulated estimates and 1000 relative errors (calculated according to eq. (1)) for each value of n. For each n, the average μ and the standard deviation σ of the 1000 relative errors were determined, with μ − 2σ and μ + 2σ as uncertainty bounds. Again, the Kolmogorov–Smirnov test was used to determine for which values of n the 1000 relative errors were normally distributed. In the latter case, μ − 2σ and μ + 2σ are the bounds of a 95% confidence interval.

3. Results

3.1. Evaluation of the different sampling strategies when applied to the long-term online dataset

A histogram of the possible estimates of the daily nitrous oxide emission was made for each simulated sampling strategy that was applied to the long-term online dataset. Fig. 6 shows the histograms of the three short-term sampling strategies, while Fig. 7 shows the histograms of the possible emission estimates that result from implementing the long-term weekly grab sampling strategy to the long-term, online dataset.

For the three short-term sampling strategies shown in Fig. 6, about 20–30% of the simulated estimates are in the lowest quantile of the histogram. This implies that these sampling schemes have a relatively high chance to severely underestimate the true average nitrous oxide emission of 91 kg N₂O—N d⁻¹. However, these three distributions also have long tails: the chance to end up with an estimate equal to the double of the true emission is about the same as the chance to obtain an estimate that is equal to the true emission (i.e. 2%). In contrast, the long-term weekly grab sampling campaigns (Fig. 7) appear to have a relatively high chance to obtain an

![Fig. 5](image-url) Histogram of the relative error for 1000 simulations of a sampling scheme consisting of 20 days of 24 h online measurements.
accurate estimate of the true emission, except for the ones during which the samples are taken on a Saturday, a Sunday or a Monday. The latter three distributions have their mean left of the true average emission of 91 kg N₂O–N d⁻¹, resulting in an underestimation of the emission. This agrees with the true average emission per day, calculated from all the online data points per day, shown in Table 3. Besides a higher accuracy, the long-term weekly grab sampling strategy also results in a higher precision than the short-term sampling campaigns, as indicated by the lower dispersion of the simulation results (note different scales of X-axes).

3.2. Precision as function of number of samples or duration of sampling campaign

3.2.1. 24 h online sampling
Using the long-term online dataset, it was possible to simulate 237 sampling periods of 24 h, starting and ending at 11 a.m. on a working day. A sampling campaign could consist of any number \( n \) of random, not necessarily consecutive 24 h periods, with \( n \) ranging from 1 to 237. Each \( n \) results in an estimate that deviates to some extent from the true average emission, resulting in a relative error \( \varepsilon \) for each value of \( n \). The relative error \( \varepsilon \) will generally approach to zero with increasing \( n \), as confirmed in Fig. 4. Nonetheless, the relative error does not coincide with zero when \( n \) becomes 237, because these 237 sampling periods only comprise 24 h periods starting and ending at 11 a.m. on weekdays, while the true average emission is estimated using the entire long-term online dataset of 417 days, including weekends. The precision of the estimate in function of the number of random 24 h online sampling days is inferred from Fig. 8A. By running 1000 simulations for each \( n \), 1000 values of the relative error are obtained for each \( n \). For each \( n \), one can plot \( \mu - 2\sigma \) and \( \mu + 2\sigma \) as uncertainty bounds. For \( n > 7 \) and \( n < 230 \), the 1000 relative errors are normally distributed. In those cases, the uncertainty bounds are the bounds of a 95% confidence interval around the average.

Fig. 6 – Histograms of the possible emission estimates that result from implementing different short-term sampling strategies: a single 24 h period of online measurements, a single 7 d period of online measurements and a single grab-sample, respectively. The vertical line indicates the true average nitrous oxide emission as determined by the long-term, online sampling campaign.

From Fig. 8 it is clear that ten random 24 h periods of online sampling would yield an estimate whose relative error has 95% chance to lie between −62 and 78%. A subset of 50 sampling periods of 24 h yields a relative error between −22 and 35% in 95% of the cases. An increase in the number of sampling days results in an ever decreasing marginal improvement in precision. It should be noted that the graph shows a slight bias towards a positive error. That is because the true emission is based on the entire long-term, online dataset (417 days) while the simulation of the sampling strategy is limited to 24 h periods starting and ending on working days only (237 days). Therefore, the maximum number of 24 h sampling days results in a simulated dataset that still contains less data than the long-term, online dataset that includes the weekends as well. When \( n \) equals 237, the estimate and the concomitant relative error is the same for each of the 1000 iterations, since there is only one way in which to sample 237 days. At that point, the lower and upper uncertainty bounds coincide.

3.2.2. 7 day online sampling
A similar approach was adopted for determining the relative error in function of the number of not necessarily consecutive 7 day periods during which the emission is measured online. In total, there were 289 periods that qualified for simulation. Taking into account that those periods were not allowed to overlap, a sampling campaign could consist of any number \( n \) of random, not necessarily consecutive 7 d periods, with \( n \) ranging from 1 to 57. The precision of the estimate in function of the number of random 7 d online sampling periods can be inferred from Fig. 8B, in a similar way as for the number of 24 h online sampling periods. Only for \( n > 7 \) and \( n < 57 \) are the uncertainty bounds also the bounds of the 95% confidence interval around the average, since the relative error \( \varepsilon \) follows
a normal distribution for those values of \( n \). Sampling online for 10 not necessarily consecutive weeks would result in an estimate whose relative error has a 95% confidence interval between 55% and 59%. To obtain an estimate with a relative error that has a 95% confidence interval of about 10%, one would need to sample during more than 53 (not necessarily consecutive) weeks. In the case of online sampling periods, the upper and lower uncertainty bounds do not coincide when the maximum number of sampling weeks is obtained, since a campaign of 57 online sampling periods of 7 days can start on each working day of the week. So, even with 57 sampling periods of 7 days, the estimate will differ depending on the day of the week on which the mimicked campaign starts.

Fig. 7 – Histograms of the possible emission estimates that result from implementing the long-term (58, 59 or 60 weeks, depending on the day of the week) weekly grab sampling strategy to the long-term, online dataset. The first seven plots show the outcome per day of the week separately, while the last plot shows the result when the weekly sample is taken on a random workday. The vertical line indicates the true average nitrous oxide emission as determined by the long-term, online sampling campaign.

Table 3 – Average true emission of N2O from the Kralingseveer WWTP per day of the week. The left column contains the values from midnight to midnight, while the right column contains the average of all daytime values (9 a.m.–5 p.m.).

<table>
<thead>
<tr>
<th>Day of the week</th>
<th>True average daily emission over 24 h [kg N(_2)O–N d(^{-1})]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monday</td>
<td>75.8</td>
</tr>
<tr>
<td>Tuesday</td>
<td>106.4</td>
</tr>
<tr>
<td>Wednesday</td>
<td>103.2</td>
</tr>
<tr>
<td>Thursday</td>
<td>108.8</td>
</tr>
<tr>
<td>Friday</td>
<td>97.0</td>
</tr>
<tr>
<td>Saturday</td>
<td>81.7</td>
</tr>
<tr>
<td>Sunday</td>
<td>66.6</td>
</tr>
</tbody>
</table>
3.2.3. Long-term weekly grab samples
In Section 3.1 it was established that long-term weekly grab sampling on a random working day during the entire length of the dataset yields the most accurate estimate of the average nitrous oxide emission. When such a sampling strategy is simulated for different campaign durations, i.e. a weekly grab sample over a different number of consecutive weeks, Fig. 8C is obtained. Only from 59 weekly grab samples onward, the relative errors are normally distributed. So, only for those values of \( n \) are the uncertainty bounds also the bounds of the 95% confidence interval around the mean. Fig. 8C demonstrates that the relative error decreases with an increasing length of the sampling campaign. Just as in the case of 7 day online sampling, the upper and lower uncertainty bounds do not coincide, because even at the maximum of 62 sampling weeks, the possible ways to obtain weekly samples are sheer infinite.

3.2.4. Completely random grab sampling
Finally, the accuracy of a campaign consisting of completely random grab samples is verified. In this case, the simulated relative errors are normally distributed for each \( n \) starting from 3.

Fig. 8D shows that a sampling scheme of 20 completely random samples results in a 95% confidence interval of [−64%, 42%], narrowing down to [−29%, 13%] for 100 samples, but the marginal improvement becomes smaller with an increasing number of samples.

4. Discussion

4.1. What is the best sampling strategy?

4.1.1. Duration and timing of the sampling campaign
The long-term, online dataset of the present measurement campaign exhibits an enormous variability, both on the long term (Fig. 1) and on the short term (Fig. 2). The exact causes of this variability are beyond the scope of the present study, but this variability has severe implications for the sampling strategy that is used to estimate the average amount of nitrous oxide that is emitted from a WWTP.

From Fig. 6 it is evident that the simulated short-term sampling campaigns, be it a single grab sample, 24 h or 7 day online monitoring, are unable to result in a good
estimation of the yearly nitrous oxide emission from the municipal WWTP under study. If such a short-term monitoring campaign was simulated for October 2010, the resulting estimate of the nitrous oxide emission would result in a severe underestimation of the yearly nitrous oxide emission from the WWTP, while a simulated short-term campaign during Spring 2011 would lead to a serious overestimation (Fig. 1). This emission pattern is consistent with the results from a previous study at the same WWTP, where the emission was measured during one week in October 2008 and one week in February 2009 (STOWA, 2010). In that study, the emission was low in October and high in February.

The histograms of the simulated estimates resulting from short term sampling campaigns are heavily skewed to the right, indicating that such campaigns have a higher chance to result in an underestimation of the emission. Therefore short-term sampling strategies are ill suited to obtain an accurate estimate of the average nitrous oxide emission of the Kralingseveer plant under study. The long-term weekly sampling strategy performs better, as shown in Fig. 7. Nonetheless, the resulting simulated estimates from a long-term weekly grab sampling strategy depend on the day on which the sample is taken. If the weekly sample is taken on a Saturday, a Sunday or a Monday, the resulting estimate is an underestimate of the true average emission. Indeed, for reasons that are beyond the scope of this paper, the true emission on those days of the week is lower than on the other days (Table 3). In fact, a long-term weekly grab sampling strategy during which samples are taken on a fixed day of the week, results in an estimate of the true daytime emission on that particular day of the week. Since it is not possible to know beforehand which days of the week have a deviant emission, sampling on a random day is the better option to get an estimate of the yearly average nitrous oxide emission.

4.1.2. Sampling frequency
Although long-term weekly grab samples give the most accurate and precise estimate of the yearly nitrous oxide emission from the WWTP, such a sampling strategy cannot discern the diurnal emission pattern that is illustrated in Fig. 2. If during this period samples are taken during daytime only, as is usually the case with grab sampling, the total emission during this period is underestimated since the emissions at this particular WWTP and during this particular period are at their lowest during the daytime. As already mentioned above, this diurnal pattern is not representative for the entire measurement period. During other periods the daytime emissions were actually higher than the night-time emissions. That is why the long-term grab sampling strategy, although using only daytime samples, still results in acceptable simulated estimates of the yearly emission, but this cannot be known before actually measuring online during 24 h.

For research purposes, diurnal patterns may be very helpful to identify the mechanisms behind the nitrous oxide emission, since many operational parameters of a WWTP also show diurnal variability (e.g. dissolved oxygen, nitrogen load, influent flow etc.). In order to correlate these parameters with emissions, high frequency data of both the emission and the parameters are needed, and that can only be achieved with high frequency, online sampling. Kampschreur et al. (2008) already mentioned that the registration of their dynamics is a prerequisite to determine the relationship between the emissions and the operational conditions. Their online measurements on a full-scale nitritation reactor showed that nitrous oxide emissions depend mainly on dissolved oxygen level, nitrite concentration and aeration rate. In their turn, Ahn et al. (2010a) were able to establish a correlation between diurnal variability in the emission of nitrous oxide and the diurnal total Kjeldahl nitrogen loadings on a full-scale municipal WWTP. In a pilot-scale study, Lotito et al. (2012) exposed a diurnal pattern of nitrous oxide emissions during the day. This pattern was correlated to the ammonia and nitrite peaks in the tank. Finally, Aboobakar et al. (2013) used online data of emissions and several process conditions of a full-scale WWTP in a statistical analysis and found out that the emissions were negatively correlated with the dissolved oxygen level.

From the studies mentioned above, it is clear that a consensus about the causes of the emission is still lacking. Yet, it is obvious that insight in the short-term dynamics of both the nitrous oxide emission and the process conditions is essential in the discussion about which process conditions induce nitrous oxide emissions. As the present study demonstrates, the short-term dynamics of the nitrous oxide emission can only be revealed by high-frequency (online) sampling.

4.2. Precision as function of sampling campaign duration

The precision of the estimate resulting from a sampling campaign increases with the length of the campaign. This is also evident from Fig. 8. However, sampling requires a considerable effort and cost, resulting in a trade-off between precision and resources. To decide how long a sampling campaign should last or how many samples should be taken, the precision of the final estimate as a function of the campaign length or the number of samples was assessed in this study.

4.2.1. Grab sampling
From the evaluation of different sampling strategies in Section 4.1 it was concluded that, of all the discussed sampling strategies (cf. Table 1), a weekly sampling campaign over the entire measurement period, during which the samples are taken on a random working day results in the most precise estimate of the true nitrous oxide emission from this WWTP. This was concluded after simulating such a campaign for the entire length of the long-term online monitoring period, i.e. 62 weeks. Nevertheless, depending on the desired precision, a shorter monitoring period might have been sufficient. In Fig. 8C it was shown that the relative error decreases linearly with the number of consecutive sampling weeks, until about 60 weeks, when there is a 95% chance that the absolute value of the relative error is smaller than 25%. When grab samples are taken completely randomly instead of weekly, the same precision can never be obtained (Fig. 8D). If the 60 grab samples are taken randomly instead of weekly, the absolute value of the relative error has a 95% chance to be smaller than 41% (Fig. 8D). However, when only 30 samples are taken completely at random, the relative error still has a 95% chance to be smaller than 50% (Fig. 8D), while 30 grab samples taken
in consecutive weeks result in uncertainty bounds between –112% and 91% (Fig. 8C). Considering the cost of sampling, this may be a clue to determine the length of a weekly grab sampling strategy that should have been applied to estimate the yearly nitrous oxide emission from Kralingseveer WWTP.

Weekly grab sampling campaigns that are shorter than the complete dataset, perform relatively poorly, because such campaigns do not capture the seasonal variability in the complete dataset. When planning a weekly sampling campaign, one should try to cover the variability that can be reasonably expected. On a WWTP, the seasonal variability is caused by the seasonal variations in ambient temperature. To cover the entire temperature range, one has to sample at least one year, which amounts to about fifty weekly grab samples. However, in the case of grab sampling, the cost is primarily associated with taking and analysing the sample, rather than with the duration of a campaign. Considering this, it is better to take only a limited number of samples spread randomly over the year, than to take the same number of samples every week.

4.2.2. Online sampling

Grab sampling strategies cannot bring insight into the diurnal variability of the emission, in contrast to online monitoring strategies such as the 24 h of 7 day online monitoring campaigns. Of course, one could consider to repeat several 24 h or 7 day online monitoring campaigns over a longer period in order to take the long-term variability into account as well. The precision of the simulated estimate in function of the number of such sampling events is shown in Fig. 8A and B. To obtain a relative error that has 95% chance to be smaller than 50%, one would need 25 random periods of 24 h online sampling or 12 random periods of 7 day online measurements. Considering the effort it takes to install, calibrate and dismantle the measuring equipment it is worthwhile to measure online over a longer period, preferably one year to cover the entire temperature range, instead of several short periods. But when budgetary or other constraints limit the use of online analysing equipment to a limited number of days or weeks, it is better in terms of precision to spread the measurements over the entire year than to measure during consecutive periods.

4.3. What about the emission from other WWTPs?

The present study is entirely based on a dataset of nitrous oxide emission of the WWTP of Kralingseveer from 14 October 2010 to 26 January 2012. These data were used to identify the way of sampling that would have resulted in the best estimate of the average emission at this particular WWTP in this particular period. One could argue that the conclusions from this study are not necessarily valid for other plants. Indeed, the values of the confidence interval bounds, for instance, cannot be directly applied to other studies. The variability that was encountered in the nitrous oxide emission from Kralingseveer WWTP is not necessarily representative for other plants. Yet, the lack of other long-term online datasets of full-scale nitrous oxide emission makes it impossible to assess to what extent the tremendous variability at Kralingseveer is generally applicable. In two other water-related fields, this lack of data was countered by testing different monitoring scenarios against simulated data from dynamic models of the presence of pollutants in sewers (Ort and Gujer, 2006) and water bodies (Gevaert et al., 2009). For nitrous oxide emission, a reliable mechanistic model that is able to describe the dynamics of nitrous oxide emission from full-scale WWTPs is not available (yet). Overall, rather than prescribing the exact way in which to sample, this study reveals the caveats involved in sampling nitrous oxide emissions from full-scale WWTPs.

5. Conclusions

A long-term dataset of continuous nitrous oxide emission measurements from a municipal WWTP was used to evaluate different sampling strategies that have been used in previous studies to monitor these emissions. A reliable determination of the actual average nitrous oxide emission of a WWTP requires long-term sampling, be it online or grab sampling, covering the entire temperature range that can possibly be encountered. For long-term grab sampling, night-time and weekend samples contribute significantly to a more accurate estimate.

If the interest of the monitoring campaign is in the short-term variability, rather than in the mere average nitrous oxide emission, (long-term) grab sampling is not sufficient. In that case, e.g. when researchers want to compare the diurnal dynamics of the emission with the diurnal dynamics of the plant’s process conditions, high-frequency (online) sampling is indispensable. Especially in the case of online sampling, long-term sampling is very resource-demanding. As a guideline to help balancing cost and precision, a method was presented to obtain the number of grab samples or online sampling periods that would have been required to obtain a sufficiently precise estimation of the emission.

Acknowledgements

This research was financed by Stichting Toegepast Onderzoek Waterbeheer (STOWA), the Dutch Foundation for Applied Water Research. The authors are much obliged to Hoogheemraadschap van Schieland en Krimpenerwaard, the Water Board of Schieland and Krimpenerwaard. Udo van Dongen is greatly acknowledged for his help in collecting the data. The authors thank Arjan Hensen (Energy research Centre of the Netherlands) for his constructive remarks.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.watres.2013.03.016.

References


Ort, C., Lawrence, M.G., Rieckermann, J., Joss, A., 2010b. Sampling for pharmaceuticals and personal care products (PPCPs) and illicit drugs in wastewater systems: are your conclusions valid? A critical review. Environmental Science & Technology 44 (16), 6024–6035.


STOWA, 2010. Emissies van broeikasgassen van RWZI’s. (Amersfoort, the Netherlands).
