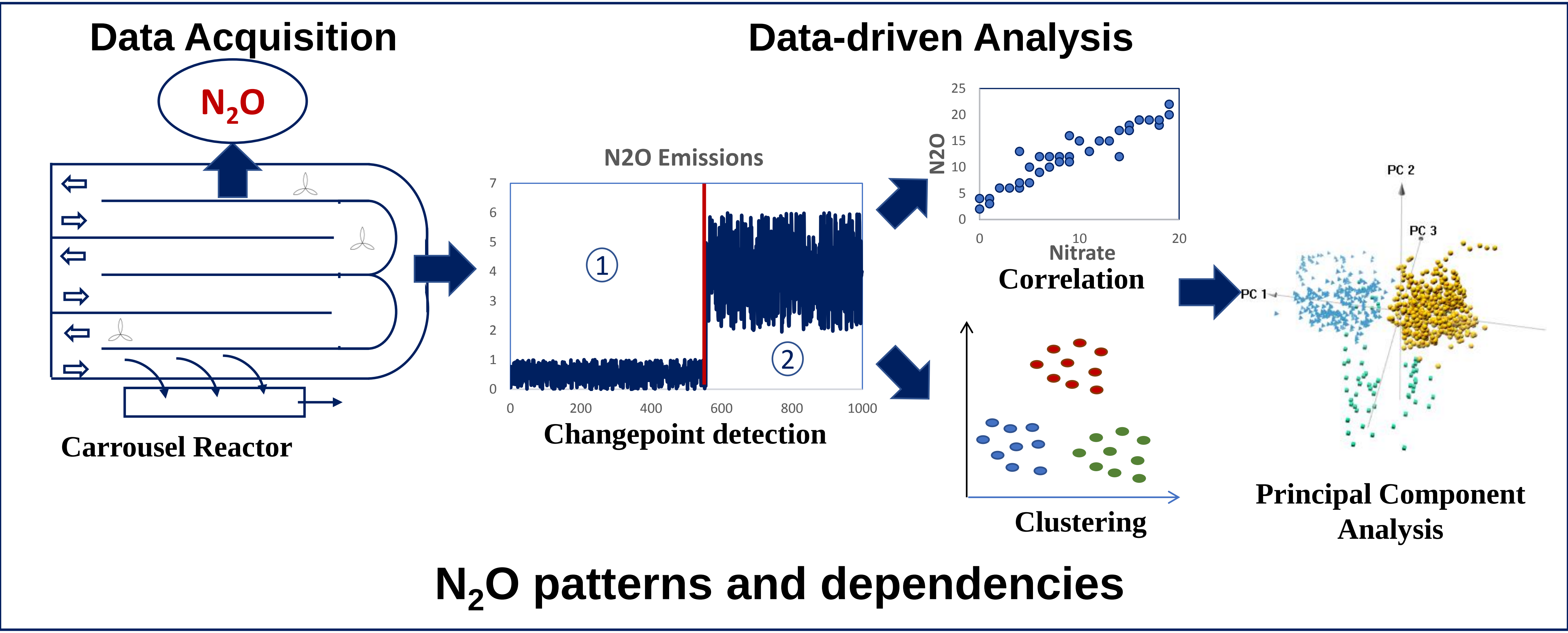
***Highlights (for review)**

Highlights

- Multivariate analysis of N2O emissions and process variables in full-scale reactor
- Correlation between N_2O, DO, NH_4-N and NO_3-N fluctuates in different periods
- Clusters group the different ranges of the process variables and N2O emissions
- PCA shows the combined effect of process variables on the system and N2O emissions

Graphical Abstract



Relating N_2O emissions during biological nitrogen removal with operating conditions using multivariate statistical techniques

Vasilaki V.[a], Volcke, E.I.P.[b], Nandi A.K.[c], van Loosdrecht[d], M.C.M., Katsou E.[a*]

[a] Department of Civil & Environmental Engineering, Brunel University London, Uxbridge UB8 3PH, UK

[b] Department of Green Chemistry and Technology, Ghent University, Coupure Links 653, 9000 Gent, Belgium

[c] Department of Electronic and Computer Engineering, Brunel University London, Uxbridge UB8 3PH, UK

[d] Department of Biotechnology, Delft University of Technology, Van der Maasweg 9, 2629 HZ Delft, The Netherlands

*Corresponding author. Department of Civil & Environmental Engineering, Brunel University London, Uxbridge UB8 3PH, UK. Email: evina.katsou@brunel.ac.uk

Keywords: N_2O emissions, long-term monitoring campaign, principal component analysis, hierarchical k-means clustering

Abstract

Multivariate statistical analysis was applied to investigate the dependencies and underlying patterns between N_2O emissions and online operational variables (dissolved oxygen and nitrogen component concentrations, temperature and influent flow-rate) during biological nitrogen removal from wastewater. The system under study was a full-scale reactor, for which hourly sensor data were available. The 15-month long monitoring campaign was divided into 10 sub-periods based on the profile of N_2O emissions, using Binary Segmentation. The dependencies between operating variables and N_2O emissions fluctuated according to Spearman's rank correlation. The correlation between N_2O emissions and nitrite concentrations ranged between 0.51-0.78. Correlation > 0.7 between N_2O emissions and nitrate concentrations was observed at sub-periods with average temperature lower than 12 °C. Hierarchical k-means clustering and principal component analysis linked N_2O emission peaks with precipitation events and ammonium concentrations higher than 2 mg/L, especially in sub-periods characterized by low N_2O fluxes. Additionally, the highest ranges of measured N_2O fluxes belonged to clusters corresponding with NO_3-N concentration less than 1 mg/L in the upstream plug-flow reactor (middle of oxic zone), indicating slow nitrification rates. The results showed that the range of N_2O emissions partially depend on the prior behavior of the system. The principal component analysis validated the findings from the clustering analysis and showed that ammonium, nitrate, nitrite and temperature explained a considerable percentage of the variance in the system for the majority of the sub-periods. The applied statistical methods, linked the different ranges of emissions with the system variables, provided insights on the effect of operating conditions on N_2O emissions in each sub-period and can be integrated into N_2O emissions data processing at wastewater treatment plants.

Abbreviations

AOR: Ammonia oxidation rate

CH_4: Methane

CO_2: Carbon dioxide

DO: Dissolved oxygen

GHG: Greenhouse gas

N_2O: Nitrous oxide

NH_4-N: Ammonium nitrogen

NO_2-N: Nitrite nitrogen

NO_3-N: Nitrate nitrogen

PC: Principal component

PCA: Principal component analysis

PLS: Partial least squares

TN: Total nitrogen

WWTP: Wastewater treatment plant

1. Introduction

The increasing demand to reduce the carbon footprint of municipal wastewater treatment plants (WWTPs) by reducing greenhouse gas (GHG) emissions and energy consumption, is posing new challenges for the water industry (Flores-Alsina et al., 2014). The climate change pressures prompt the quantification and minimization of GHG emissions generated in WWTPs (Haas et al., 2014). Three main sources of GHG emissions prevail in WWTPs (Monteith et al., 2005; Mannina et al., 2016): (i) the direct emissions mainly linked to biological processes, (ii) the indirect internal emissions generated by the use of imported energy to the plants, and (iii) the indirect external emissions associated with the sources that are controlled outside the WWTPs (e.g. chemicals production, disposal of sewage sludge, transportation). The GHGs emitted into the atmosphere from biological wastewater treatment processes are carbon dioxide (CO_2), methane (CH_4) and nitrous oxide (N_2O) (Kampschreur et al., 2009b).

With the potential contribution of 265 times more than CO_2 for a 100-year time horizon to global warming (IPCC, 2013), N_2O is a potent GHG and the most significant contributor to ozone depletion (Ravishankara et al., 2009). WWTPs are significant generators of N_2O and are responsible for 3.1% of the N_2O emissions in Europe (EEA Report, 2017). N_2O is generated mainly during the autotrophic nitrification and heterotrophic denitrification (Kampschreur et al., 2008) and can contribute up to 78% (Daelman et al., 2013) of the footprint of a WWTP's operation. Recent studies have focused on the understanding, quantification, control and minimization of N_2O emissions (Aboobakar et al., 2013; Mampaey et al., 2016; Pan et al., 2016). However, several studies have resulted in contradicting findings on the influence of operating and environmental variables on N_2O generation (Liu et al., 2016; Massara et al., 2017). For instance, several studies have reported increasing N_2O emissions with decreasing DO concentrations during nitrification (Kampschreur et al., 2009b). However, Rodriguez-Caballero et al. (2014) found that N_2O emission profiles in a full-scale biological reactor did not change even for DO variations higher than 1.5 mg/L. The latter, was attributed to the high nitrification efficiency and the potential biomass adaptation to continuously varying DO concentrations. Results from real-field N_2O monitoring campaigns cannot fully explain long-term causes of N_2O emissions and the combined effect of operating, environmental and external factors that influence the biological systems

(Jönsson et al., 2015). Long-term full-scale monitoring campaigns have shown that N_2O fluxes are highly dynamic with significant diurnal fluctuations and seasonal variations; however, the dynamics cannot be fully explained (Daelman et al., 2015; Kosonen et al., 2016).

Several mechanistic process models describing N_2O emissions from wastewater treatment plants have been developed over the last few years (Massara et al., 2017). While they have been successfully applied to identify N_2O formation mechanisms and pathways from experimental data (Ni et al., 2015; Pocquet et al., 2016), their calibration and validation to long-term process data remains a challenge. Domingo-Félez and F. Smets (2016) reported that substrate affinity constants for NO_2 and NO reduction in existing N_2O models differ by a factor of about 100. Additionally, calibration of models under specific operational conditions (i.e. dry weather) can affect their performance and accuracy when the system varies (Gernaey et al., 2004; Guo and Vanrolleghem, 2014). Moreover, full-scale N_2O emission data show long-term trends that cannot be explained by commonly available operational data (Daelman et al., 2015) but are possibly caused by microbial population changes, which are hard to catch with the current models, typically describing single functional groups with fixed parameter sets. Multivariate statistical techniques are capable of identifying relationships between N_2O emissions and a multitude of influencing factors, at the same time identifying various operating sub-periods for which this behaviour may differ. This will lead to increased understanding of experimental data, on its turn facilitating the application, calibration and validation of mechanistic models. As such, multivariate statistical techniques maximize the information acquired from N_2O monitoring campaign data.

Statistical techniques have been used for the analysis of data from full-scale monitoring campaigns, to identify interconnections between operating and environmental variables on the one hand and N_2O formation on the other hand. Through multiple linear regression analyses, Aboobakar et al. (2013) showed dependencies between N_2O emissions and nitrogen load, temperature and dissolved oxygen (DO) in various compartments of a plug-flow reactor for biological nitrogen removal. Multi-regression analysis of one year of data with bi-monthly sampling frequency, coming from a full-scale SBR (Sun et al., 2013) indicated negative correlation between N_2O emissions and temperature, while COD/N ratio lower than 6 resulted in higher emissions. Brotto et al. (2015) used Spearman's rank correlation to explain the behavior of N_2O emissions in an activated sludge process. The analysis showed negative correlation between N_2O emissions and pH but positive correlation between N_2O fluxes and temperature. However, most of the studies did not consider continuous long-term operational data, while further analysis is required to gain a better understanding on the dynamics and trade-offs between N_2O generation and the online monitored and controlled process variables.

Multivariate analysis has been proven to be a suitable method for the identification of patterns and hidden relationships within WWTP data (Rosén and Lennox, 2001) and can be applied to provide insights on the combined effect of operational variables on N_2O emissions in full-scale systems. Chemometric techniques have been applied to the wastewater treatment sector for 40 years (Rosén and Olsson, 1998), enabling the visualization and interpretation of the multi-dimensional interrelations of the operational variables monitored in biological processes (Platikanov et al., 2014). Their application can (i) improve the efficiency of process monitoring (Mirin and Wahab, 2014) and provide further insights of the biological processes (Moon et al., 2009), (ii) identify and isolate process faults (Haimi et al., 2016; Liu et al., 2014; Maere et al., 2012; Rosen and Yuan, 2001), sensor faults (Lee et al., 2004), and iii) predict significant operating variables in the biological systems that affect performance (Rustum et al., 2008). Furthermore, the gradual implementation of online sensors to monitor important parameters in the biological treatment train of WWTPs results in the production of time series, which require the application of specific statistical tools for their interpretation. The most widely applied approaches include methods aiming to reduce the dimensionality of large data-sets (i.e., principal component analysis (PCA), partial least squares (PLS)) and data clustering techniques (i.e., hierarchical clustering, k-means clustering) (Haimi et al., 2013). However, there are limited studies investigating the behavior of N_2O emissions with the application of multivariate statistical techniques, especially utilizing online operational data in long-term monitoring.

The aim of this work is to investigate whether widely applied multivariate statistical techniques can be applied to the online data collected from real-field N_2O monitoring campaigns in order to gain a better understanding on the dynamic behavior of N_2O emissions and explain the combined effect of the operating variables monitored in wastewater treatment processes on N_2O emissions. Hourly data from the operating variables monitored online and N_2O emissions data in a full-scale carrousel reactor from the long-term monitoring campaign published by Daelman et al. (2015) were used for the analysis. A statistical methodological approach was developed, applying changepoint detection techniques to identify changes in the N_2O fluxes behavior combined with hierarchical k-means clustering and PCA, to provide insights on N_2O emissions patterns and generation pathways.

2. Materials and methods

2.1 Process description and data origin

This work was based on the data obtained by Daelman et al. (2015) for the Kralingseveer WWTP, consisting of a plug-flow reactor followed by two carrousel reactors in parallel (Figure 1). The plant treated 80.000 m^3 d^{-1} of domestic wastewater from a combined sewer system. The carrousel reactors were characterized by alternating anoxic/oxic zones; aeration was performed through surface aerators, which were manipulated to control the ammonium concentration in the effluent. Aerator 1 operates under on/off pattern, being on when the ammonium concentration was higher than 1.2 mg N/L), while surface aerators 2 and 3 were always operational to keep the solids from settling but operated at maximum capacity when the ammonium concentration became higher than 0.6 and 0.9 mg/L, respectively. Over the monitoring period the average total nitrogen (TN) removal efficiency was 81 ±10%; the average COD removal efficiency was equal to 87 ±5%.

Ammonium nitrogen (NH_4-N), nitrate nitrogen (NO_3-N) and DO were monitored in the middle of the second oxic zone in the plug flow reactor (location 1, Figure 1). The carrousel reactors were equipped with, NH_4-N, temperature probes, and 3 DO probes (DO1, DO2, DO3) (locations 2, 3, 4, Figure 1). The Northern carrousel reactor was also equipped with a nitrite probe. All the reactors were covered, and the off-gas was collected in ducts and pumped to a Servomex gas analyzer, where N_2O was measured. Table S1 lists all the variables monitored online (Supplementary material). The data matrix developed consists of the variables monitored in the carrousel reactor (DO, NH_4-N C, NO_3-N C, NO_2-N C, N_2O C), the influent flow-rate, as well as the NH_4-N and NO_3-N concentrations from the plug-flow reactor. 24 h composite samples of influent and effluent, available about every 6 days, were used to support the analysis. Figure 2, summarizes the methodological framework applied to the online database.

[Figure 1]

2.2 Methodological framework for data analysis

The monitoring period was divided into distinct sub-periods based on the profile of N_2O fluxes in the carrousel reactor. Spearman's correlation analysis, k-means clustering, hierarchical clustering, and Principal component analysis were applied to the database. The application of clustering algorithms facilitated the identification of operational modes that have historically resulted in specific ranges of N_2O emissions. The PCA reduced the dimensionality of the data-set transforming the sensor signals into useful knowledge that that can be easily interpreted. The methodological framework is extensively described in the following sub-sections.

[Figure 2]

The data-driven approach enabled the utilization of the information and patterns embedded in the real-time monitored variables (from the system sensors) in the biological processes and GHG measurements. Multivariate statistical analysis is an alternative to univariate analysis that is commonly applied for the analysis of WWTP data. It enables the identification of patterns and interrelations in data-sets by examining multiple variables simultaneously (Olsson et al., 2014). R software was used for the statistical analysis (R Core Team, 2017). The complete list of packages used is provided in the supplementary material (Table S2).

2.2.1 *Preliminary data processing*

The preliminary data analysis included: (i) data synchronization under the same time-stamp, and ii) removal of duplicate and unreliable measurements (multiple readings at the same time stamp for the same sensor). The data were aggregated into hourly averages in order to compensate for the missing data due to variation in sampling frequency between the different variables monitored. Exponential moving average imputation was applied when less than 24 consequential data were missing for each variable. Longer periods of missing data were excluded from the analysis.

2.2.2 *Binary segmentation changepoint detection*

Given a series of data, change point analysis investigates abrupt changes in a data-series when specific properties change (i.e., mean and variance) (Kawahara and Sugiyama, 2012). The Binary Segmentation (Scott and Knott, 1974) is a widely applied and computationally efficient changepoint detection algorithm (Killick et al., 2012). The algorithm employs initially single changepoint detection method to the complete data-set as described in (Killick and Eckley, 2014). If a changepoint is identified the procedure is repeated to the two new segments formed; before and after the changepoint. The process continues splitting the data until there are no more changepoints identified. The computational cost of the algorithm is of the order of O(nlog n) with n being the number of data in the data-set and therefore it is applicable in large data-sets. A distribution-free test statistic was applied based on the work of Chen and Gupta, (1997). The penalty for the changepoints identification was equal to log(n). The algorithm requires independent data points. Therefore, first difference transformation of the N_2O timeseries was performed and changes in variance were identified by the Binary segmentation algorithm. The profile of the N_2O emissions was highly variable during the monitoring campaign. Binary segmentation enabled the identification of the sub-periods characterized by different N_2O emissions' profile.

2.2.3 *Spearman's rank correlation*

Spearman's rank correlation coefficient (Spearman, 1904) was used to detect bivariate temporal monotonic trends among the system variables for the different sub-periods; it served as a measure of the association strength. This method is based on the rank of the values and therefore, is less sensitive to outliers than Pearson's correlation. P values lower than .01 were considered to be significant.

2.2.4 *Hierarchical k-means clustering*

Clustering techniques are widely applied in data mining in order to identify and group the underling patterns that exist in high dimensional data sets (Jain, 2010). K-means clustering (Hartigan and Wong, 1979) is a recognized clustering algorithm (Haimi at al., 2013). K-means clustering was applied to categorize the data in groups of similar observations and to investigate the patterns of N_2O emission fluxes, based on Euclidean distance. K-means algorithm begins with the selection of k random centroids of the same dimension within the original data. All the data-points are compared and assigned to the nearest centroid. During each iteration, the nearest data to each centroid are re-defined and centroids are recalculated in a way that squared distances of all points within a cluster to the cluster's centroid are minimized. However, the randomly selected initial centroids can result into locally optimized clustering results (Abu-Jamous et al., 2015). Therefore, hierarchical k-means clustering that was proposed by Arai and Barakbah, (2007), was applied to the dataset. In this method agglomerative hierarchical clustering (Kaufman and Rousseeuw, 1990) is applied for the selection of the centroids; Ward's method is used in order to divide the dataset in clusters (Ward Jr, 1963). The data were normalized before the analysis. NBclust package in R (Charrad et al., 2014) was used to select the number of clusters in each sub-period. The package applies a number cluster validity indexes (i.e. average silhouette value (Rousseeuw, 1987); Hartigan's rule (Hartigan, 1975)).

Hierarchical k-means clustering was applied to the carrousel reactor data matrix from the different sub-periods identified through binary segmentation, to investigate whether different temporal patterns of the operating variables were responsible for the different behavior of N_2O emissions. Hierarchical k-means clustering enabled i) the detection of frequency and persistence of extreme ranges of operating variables, and ii) the comparison of the operational modes between the plug-low and carrousel reactor. Ammonium and nitrate probes in the plug-flow reactor were included in the analysis, since they can provide indirect feedback in terms of the carrousel reactor influent and additional information for the operational behavior of the system. However, the analysis was repeated excluding plug-flow variables (NH_4-N and NO_3-N). Graphical comparisons of the clustered data-points versus time and boxplots of the variables in each identified cluster are displayed in the results' section.

2.2.5 *Principal component analysis*

Principal component analysis (PCA) (Jolliffe, 2002) was applied to the dataset in an effort to reduce the dimensionality of the data by eliminating a small proportion of variance in the data. PCA transforms the original correlated measured variables to uncorrelated variables, i.e., Principal components (PCs), explaining the maximum observed variability. The principal components are linear combinations of the original data variables. The loadings of the variables in each principal component can map their relationship with the respective principal component. PC scores are a linear combination of the data, weighted by the PC loadings for each variable. The scores of the principal components map the different samples in the new dimensional space of the principal components facilitating the investigation of the different relationships between the variables. The data matrices (X) consisting of J columns (variables) and I data rows (number of observations) were normalized with mean equal to 0 and standard deviation equal to 1. Each column of **X**, $x_j =$

$(x_{1j}, \dots x_{Ij})T$, j=1,...J, represents a vector in the I-dimensional space. In PCA, eigenvalue decomposition is used to factorize the data matrix X *(I x J)* and to map the data matrix to a reduced dimensional space*:*

$$X = TP^T + E$$

where, *T:* matrix *(I x S)* representing the score of the principal components, *S:* the number of principal components selected, *P:* matrix *(J x S)* representing the loadings and *E*: matrix of residuals.

The biplot of the first 2 PCs was used in order to visualize the combined behavior of significant variables that affect the system. The biplots enabled the simultaneous visualization of i) the variables' loadings in the first two principal components, ii) the scores of the first two principal components, and iii) the different clusters. The temporal variations of the PC scores enabled the identification of occasions in which the behavior of the system changes. PCA was applied to the data matrix of the carrousel reactor excluding N_2O emissions time series, i) to identify the most significant variables that affect the system, (ii) to analyze the structure of the sensor data, iii) to investigate if changes in the relationship of the system coincide with changes in the N_2O emissions profile, and iv) to validate the results from hierarchical clustering. N_2O emissions time series were excluded from the PCA in order to investigate the relationship between the PC scores and N_2O emissions and to examine which PCs are most significantly linked to the behavior of N_2O emissions.

3. Results and discussion

3.1 N_2O emissions profile and main dependencies

The profile of all the variables monitored was fluctuating during the monitoring period, which can justify the different profiles of N_2O emissions that resulted from the Binary Segmentation algorithm. Overall, high ranges of emissions were reported when nitrate concentration in the plug-flow reactor was low, whereas periods with lower ammonium concentrations in the plug-flow reactor were linked with lower N_2O emissions.

Table 1 shows the average values and standard deviations of the variables monitored online and offline in the Northern carrousel and plug-flow reactors. N_2O fluxes peaked in March 2011 followed by a period characterized by very low N_2O emissions. Gradual decrease was observed until November 2011 and negligible emissions again until January 2011 (Figure 3).

[Table 1]

The application of Binary Segmentation algorithm to the N_2O emissions of the Northern carrousel reactor identified 9 changepoints that correspond to 10 sub-periods with distinct variance of the N_2O timeseries first difference (Figure 3). The analysis identified abrupt temporal changes in the emission dynamics that indicate changes in the underlying mechanisms or environmental conditions responsible for the N_2O formation.

[Figure 3]

Offline data were analyzed in the different sub-periods in order to investigate significant changes that can contribute to the high N_2O emissions in sub-periods 4 and 5. The average COD concentration in the influent of the plug-flow reactor (effluent of primary sedimentation) was 239 ± 80 mg COD/L over the 15-month monitoring period. The average plug-flow reactor influent and carrousel reactor effluent concentrations of COD, TKN, BOD, TP and the effluent pH for all sub-periods are given in the supplementary material (Table S3). In sub-period 5, 27% increase in the influent COD concentration to the plug flow reactor (compared to average value) was observed, which could be attributed to less precipitation events and to the consequently lower average influent flow-rate during this sub-period. Laboratory analyses did not show significant seasonal changes in the plug-flow COD loading (19,934 ± 13310 kg COD/day). The COD loading in sub-period 4 (16,160 ± 2546 kg COD/day) was 17% less than in sub-period 1. TKN and TP loadings were reduced in sub-period 4 compared to sub-period, by 11% and 12% respectively. The COD:TKN:TP ratio remained quite stable, ranging between 1:0.17:0.02 (sub-period 2) and 1:0.20:0.03 (sub-period 4).

Figure 4 shows the different COD to TKN ratios measured for all the sub-periods. There were cases with lower than average COD/TKN in the influent of the plug-flow reactor that coincided with increased N_2O emissions, particularly in sub-periods 4 and 5. However, low ranges of COD/TKN (<5) in sub-periods 1, 2, 7 and 6 corresponded with low N_2O emissions. These observations indicate that limitation of COD cannot be considered the sole contributor of N_2O emissions via heterotrophic denitrification in sub-periods 4 and 5.

[Figure 4]

The COD removal efficiency remained relatively steady during the monitoring campaign ranging from 79% (sub-period 8) to 91% (sub-period 5). The range of TN and TP removal efficiencies ranged from 73 % (sub-periods 1 and 9) to 92% (sub-period 5) and from 67% (sub-period 7) to 87% (sub-period 4). The effluent pH was steady (~ 8) and did not show seasonal variability that could influence the generation of N_2O emissions.

On the other hand, a significant variation is observed for all variables monitored online by analyzing at the complete database. Table 2 summarizes the average values and standard deviations of the online monitored variables considered in the analysis for the target periods. In the carrousel reactor, the nitrite concentration is relatively high in sub-period 4 (average = 2.6 mg/L) and in the first part of sub-period 10 (average = 2.1 mg/L). The average temperature in both cases is ~13 °C. In biological reactors operating in continuous mode, appreciable (> 2 mg N/L) nitrite concentrations are usually not observed, since nitrite is directly oxidized by nitrite oxidizing bacteria into nitrate. However, in certain cases, high nitrite concentrations in biological processes have been observed, which have been linked with low temperatures that affect N_2O reductase during denitrification enhancing N_2O production (Holtan-Hartwig et al., 2002; Adouani et al., 2015).

Analyzing the whole profile, the emissions tended to be low at higher temperatures (sub-periods 6, 7, and 8). Higher emissions were also observed, though, at temperature higher than 18 °C and low nitrite concentrations (i.e., sub-period 5). Ahn et al. (2010) demonstrated that N_2O emissions can be significant at higher temperatures due to the higher enzymatic activities of the bioprocesses producing N_2O. In the carrousel reactor during sub-periods 4 and 5, the temperature increases from 11.8 to 20 °C. Low N_2O emissions were also observed when ammonium concentration was lower than 13 mg/L and nitrate was higher than 2.5 mg/L in the plug-flow reactor. The probe was located in the middle of the second oxic zone; thus, lower ammonium concentrations in the plug-flow reactor can indicate less ammonium loads in the carrousel reactor.

[Table 2]

The analysis of the variables' ranges for the N_2O emission profiles provides limited insight on the dependencies between the system variables monitored online, which is further analyzed in the following sections.

3.2 Spearman's rank correlation analysis for carrousel reactor

The application of Spearman's rank correlation coefficient to the data of the carrousel reactor could not identify significant correlations between the N_2O emissions and the operating variables. The lack of monotonic univariate dependencies could be attributed to i) the temporal fluctuations of the influent characteristics, ii) the continuous variability in the operating conditions of the reactors, and iii) the seasonal variations of the environmental conditions in wastewater treatment processes. Fluctuating correlation coefficients between N_2O emissions and carrousel reactor variables were identified (Supplementary, Figures S1:S2). The findings are in line with the study of Kosonen et al., (2016). The authors compared the results from two monitoring periods at the same biological system and identified different relationships between N_2O emissions and $BOD_{7(ATU)}$ loads.

The correlation coefficient between nitrite and N_2O emissions ranged from 0.78 (sub-period 7) to 0.51 (sub-period 9). As a general remark, nitrite was correlated with N_2O emissions in sub-periods 4, 6 and 7, while lower correlation was observed during sub-periods 5 (Figure 5), 8 and 9. N_2O emissions and NO_3-N concentration in the carrousel reactor exhibited a positive correlation with coefficient higher than 0.7 for sub-periods 2 (Figure 5), 4 and 10 (the temperature was lower than 13 °C in all cases). N_2O emissions and NO_3-N concentrations followed similar diurnal patterns, wherein peaks in nitrate concentration coincided with peaks in N_2O emissions (Daelman et al., 2015). The accumulation of nitrate is potentially linked with higher nitrification than denitrification rates. This is in line with Daelman et al. (2015), considering that the nitrate utilization rate in these sub-periods is affected by the low temperatures (Elefsiniotis and Li, 2006). Additionally, during times when N_2O was positively correlated with DO1 (> 0.5), medium to significant correlations between the N_2O emissions and the ammonium concentration in the carrousel reactor were also observed (sub-periods 1, 6 and 7). Stripping of the already formed N_2O can be a potential explanation. Given that the surface aerator in the location of DO1 probe is manipulated to control the ammonium concentration in the effluent, ammonium peaks trigger the surface aerators to start.

The correlation coefficient between any two of the system variables did not remain stable between the different sub-periods. Figure 5 shows the correlograms for sub-periods 2 and 5. These sub-periods were characterized by low and high ranges of N_2O emissions and temperature respectively (Table 2). In sub-period 2, the average NO_3-N concentration in the plug-flow reactor was equal to 2.5 mg/L (Table 2) and correlated negatively with the influent flow-rate (~ - 0.63) (Figure 5). In sub-period 5 the behavior of nitrate concentration (average equal to 2.1 mg/L) was mainly correlated negatively with ammonium concentration in the same reactor. The ammonium concentration in the carrousel reactor was positively correlated with DO1 only in sub-period 2. NH_4-N concentration in the plug-flow reactor was correlated with the influent-flow rate only in sub-periods 4 and 5. However, the profiles of these two variables showed that in the majority of the sub-periods, abrupt and rapid increase of influent flow-rate (i.e., precipitation events) coincided with increase of the NH_4-N. However, the NH_4-N concentration reduced more rapidly in the system than the influent flow-rate. For example, in sub-period 3 the correlation coefficient between NH_4-N in the plug-flow reactor and influent flow-rate was 0.26. However, when days with significant precipitation events (and thus high influent flow-rate) were omitted, the correlation coefficient was equal to 0.58. The latter shows that, in this example, the lack of correlation between these two variables is most likely to be an indication that the interrelationships are not monotonic and that the method is not appropriate to identify complex relationships within the data. In order to verify that increased influent flow-rate was linked with precipitation events, daily precipitation data were extracted from the Royal Netherlands meteorological institute. Spearman's correlation coefficient between two days moving average of influent flow-rate and daily precipitation in the Netherlands was equal to 0.69. Therefore, there is a direct link between higher than average flow-rates and precipitation events (the timeseries are shown in Figure S3, supplementary material). The correlograms for all sub-periods are provided in the Supplementary material (Figures S1:S2).

Spearman's rank correlation indicated structural changes in the dependencies between the system variables. Therefore, the fluctuating structural dependencies had a different impact on the generation of N_2O emissions. Previous studies have shown that various monitored variables in the biological system (NH_4-N, NO_3-N, NO_2-N, Temperature) can affect N_2O emissions generation. However, further analysis is required to investigate their combined effect in N_2O formation in full-scale complex systems.

[Figure 5]

3.3 Hierarchical k-means clustering

The application of hierarchical k-means clustering enabled the categorization of the different ranges of the operating variables and N_2O emissions within each sub-period.

Hierarchical k-means clustering analysis was repeated excluding NH_4-N and NO_3-N concentrations in the plug-flow reactor. The results showed that the majority of the data points were allocated to the same clusters for each sub-period even when the NH_4-N and NO_3-N concentrations in the plug-flow reactor were excluded. In the majority of the sub-periods (i.e. sub-periods 1-6) more than 85% of the data points were assigned to the same cluster. It can be concluded that specific patterns and ranges of NH_4-N and NO_3-N monitored in plug-flow reactor, systematically resulted in specific responses to the carrousel reactor. The latter is supported by the Spearman's rank correlation analysis, where high correlations were observed between the variables in the two reactors for several sub-periods. For example, the correlation coefficient between NH_4-N in the plug-flow and carrousel reactors is higher than 0.7 for sub-periods 1 to 7. The similarity of the clusters for all the sub-periods is shown in Table S4 in the Supporting Material.

The range of N_2O emissions was differentiated in the majority of the clusters. In all the sub-periods, two major clusters were identified characterized by significant differences in the NH_4-N and NO_3-N concentrations in the plug-flow reactor. In the majority of the sub-periods they represented the diurnal variability of the system nutrient concentrations and influent-flow rate. Additionally, clustering distinguished occasions with high influent flow-rate and ammonium concentration in the carrousel reactor, which can be an indication of precipitation events. In sub-periods characterized by low average N_2O emissions (i.e., 1, 2, 7, 8 and 9), clusters with increased N_2O emissions (yet relatively low) were mainly linked to higher loading rates due to the expected diurnal variability or to precipitation events. However, N_2O emissions higher than 3.8 kg/h were observed when the average NO_3-N concentration was constantly lower than 1 mg/L in the plug-flow reactor and the NO_3-N concentration was lower than 4 mg/L in the carrousel reactor. Table 3 compares the clustered average values for all the variables in sub-period 2 (average N_2O emissions equal to 0.6 kg/h – Table 2) and 4 (average N_2O emissions equal to 5.6 kg/h – Table 2). The average value of N_2O emissions for a set of clusters in a specific sub-period (from Table 3) can be found taking into account the number of data-points in the individual clusters. Sub-period 4 was characterized by very low NO_3-N concentration in the middle of the oxic zone in the plug-flow reactor. The latter indicates slower oxidation of ammonia to nitrate or insufficient DO in the plug-flow nitrification lane. This can lead to higher NH_4-N loading in the carrousel reactor. On the other hand, higher nitrification rates in the plug-flow reactor (i.e. sub-period 2) resulted in lower N_2O emissions in the carrousel reactor. The average values of all the variables in each cluster during all the sub-periods are given as supplementary material (Table S5).

In clusters 2 and 16 the averages of operating variables had similar ranges (Table 3). However, in these two occasions the N_2O emissions were different (0.01 and 0.51 kg/h). Similarly, in clusters 1, 4 and 7, the averages of operating variables were similar yet the N_2O emissions were different (0.09,0.87 and 3.22 kg/h respectively). A corollary to this also existed. In clusters 1 and 2 the averages of operating variables were different but the N_2O emissions were similar (0.09 and 0.01). Similarly, in clusters 5 and 6 the averages of operating variables were different but the N_2O emissions were similar (0.21 and 0.24). Such observations indicate the underlying complexities of the interdependencies. Additionally, it can be concluded that the range of N_2O emissions can partially depend on the preceding operational mode of the system. Figure 6 shows an example of the variables monitored online for two separate occasions in sub-periods 2 and 3 (from 00:00 am until 8:00 am) and the respective N_2O emissions. All the variables showed a similar behavior (in terms of range and trends). N_2O emission profiles had also the same trend; however, their range depended on the initial N_2O fluxes at 00:00 am. The influent flow-rates, NH_4-N and NO_3-N concentrations in the plug-flow reactor also were similar in these two occasions. The average N_2O fluxes were equal to 0.44 and 2.01 kg/h for occasion 1 and 2 respectively. More extensive data are required for quantitative investigation.

[Table 3]

[Figure 6]

3.4 Principal component analysis in the carrousel reactor

PCA was applied to transform the original correlated measured variables to uncorrelated variables (Principal components) and explain the maximum observed variability. In sub-periods with low emissions (1, 2, 7, 8, and 9) the PCA analysis showed that N_2O emissions' peaks are related with NH_4-N and influent flow-rate peaks in the carrousel reactor and with the effect of the diurnal variability of these variables' loading rates.

The current section discusses the PCA results for sub-period 2, as an example. The results for all the sub-periods are given in the supplementary material (Tables S6-S13, Figures S4-S29). The application of PCA reduced the dimensionality of the data with 4 principal components (PCs) explaining ~86% of the total variance (PC1 = 39%, PC2 = 26%, PC3 = 12%, and PC4 = 9%). Loadings for the system variables in the 4 PCs are given in Table 4. The loadings of each component are an indication of the variation in the variables explained by a specific component. Influent flow-rate, ammonium concentration in the carrousel reactor (NH_4-N C) and the three DO (DO1, DO2 and DO3) concentrations had the highest negative loadings in PC1. This means that the first principal component increased with the increase of these variables. Nitrate concentration (NO_3-N PF) in the plug-flow reactor has a relatively high positive loading in PC1 (0.36). Therefore, PC1 describes how the carrousel reactor responds to the behavior of the upstream plug-flow reactor processes and conditions, the variation of the influent flow-rate and variations in ammonium and DO concentrations in the carrousel reactor. The latter can be indirectly connected with the control strategy of the carrousel reactor, since the surface aerators were manipulated based on the effluent ammonium concentration. PC2 linked ammonium concentration in the plug-flow reactor, nitrate concentration in the carrousel reactor and temperature (loadings higher than 0.47). In PC3 ammonium concentration in the carrousel reactor had high negative loading, while DO2 and DO3 concentrations had positive loadings that was not expected considering the control strategy of the system. Investigation of the variables' profiles, though, showed an increasing trend of DO2 and DO3, whereas the ammonium profile did not present a similar trend.

[Table 4]

The biplot of the first 2 PCs is used to visualize the combined behavior of significant variables that affect the system. Data points assigned to cluster 6 (Figure 7), had negative scores in PC2 and PC1. Therefore, ammonium concentration in the carrousel reactor and influent flow rate were higher than average, while the nitrate concentration in the system was lower than average. Figure 8 shows the profile of N_2O emissions and NH_4-N in the carrousel reactor for sub-period 2. The colored points in the diagram represent the identified clusters. Peaks in emissions coincided with peaks in the NH_4-N C profile, whereas peaks in NH_4-N C coincided with precipitation events (cluster 6).

[Figure 7]

The scores of the data-points in cluster 5 were mainly positive in PC1 and negative in PC2 (Figure 7). PC2 increased with the increase of NH_4-N concentration in the plug-flow reactor (Table 4). Given that PC2 had an average equal to 0 (data are standardized), data-points with negative scores in PC2 represent occasions with lower than average NH_4-N concentration in the plug-flow reactor. This is supported by the correlation plot (Figure 7), where the arrow of NH_4-N concentration in the plug-flow reactor points to the direction of increasing concentrations of NH_4-N. Therefore, data-points belonging to cluster 5 were characterized by higher than average ammonium concentration in the plug-flow reactor. Similarly, NO_3-N concentration in the plug-flow reactor had relatively significant positive loading in PC1 (0.36 – Table 4). The latter indicates that NH_4-N and DO concentrations (measured by three probes) in the carrousel reactor (that had negative loadings in PC1 – Table 5) tended to decrease when NO_3-N concentration in the plug-flow reactor increased. Given that all data-points in cluster 5 had positive scores in PC1, it can be concluded that they are characterized by lower than average NH_4-N concentration in the carrousel reactor and higher than average NO_3-N concentration in the plug-flow reactor. According to the clustering results the latter can be an indication of the high nitrogen loadings of the normal diurnal variability in the reactor. This finding is supported from the results presented in Figure 8, where the data-points of cluster 5 correspond to the daily low range of ammonium concentrations in both reactors.

[Figure 8]

Figure 9 summarizes scores of the PC2 and the respective clusters (colored points in the diagram) indicating strong diurnal cyclic fluctuations of the water quality during this sub-period. It also shows that after each precipitation event, a sudden temperature drop occurred; the system was disturbed and cannot recover immediately. Spearman's rank correlation coefficient between PC2 and N_2O emissions is equal to 0.72.

[Figure 9]

In sub-period 4, mechanisms triggering high N_2O emissions in the carrousel reactor prevailed (average = 5.6 kg/h). The PCA loadings were similar to sub-period 2, while the clustering results indicated 3 clusters; clusters 10 and 11 were affected by the diurnal variability and cluster 12 was affected by the precipitation events (Table 3). Again, the DO data obtained from the 3 sensors in the carrousel reactor had significant negative loadings in PC1. However, ammonium concentration in the carrousel reactor was not identified as a significant variable affecting the system in the first two PCs. This can be attributed to the fact that less NH_4-N concentration peaks were observed in the effluent of the carrousel reactor (17 data points belong to cluster 12). The correlation coefficient of PC1with NH_4-N concentration in the carrousel reactor was -0.75. Therefore, PCA analysis shows that PC1 is a good indicator of the ammonium concentration in the carrousel reactor. The DO concentrations in this sub-period especially for cluster 10 (with average NH_4-N concentration in the carrousel reactor equal to 1.26 mg/L) was the highest observed in all the clusters with similar NH_4-N concentrations in the carrousel effluent. The alternation of aerobic and anaerobic conditions observed in this reactor, combined with high NH_4-N and DO concentrations has been identified as a significant cause of nitrification sourced emissions (Yu et al., 2010).

[Table 5]

In PC2, the NO_3-N concentration and temperature had significant positive loadings (Table 5). The score plot of PC2 (Figure 10a) presented an increasing trend and therefore, showed that nitrate and temperature increased. The latter was verified by the profiles of NO_3-N concentrations in the carrousel reactor (Figure 10b) and NO_3-N concentration and temperature in the plug-flow reactor (Supplementary material S30). In the beginning of the sub-period 4 very low concentrations of nitrate were observed in the system and they gradually increased especially after the 28th of March. The Spearman's correlation coefficient between N_2O emissions and PC2 scores were relatively high and equal to 0.62. However, contrary to sub-period 2, the clustering analysis showed that there is no nitrate accumulation (Table 3). The average nitrate concentration in the plug-flow reactor was equal to 0.2 mg/L until the 28th of March and increased up to 1.6 mg/L until the end of the sub-period. Therefore, the observations in section 3.3 are supported by the PCA results (low nitrate in the plug flow resulted in increased loadings in the subsequent carrousel reactor and the denitrification activity in the carrousel reactor is affected by the low temperature resulting in nitrite accumulation).

[Figure 10]

In the section, the combination of hierarchical k-means clustering and PCA was used in order to link the different emission ranges with all the online monitored variables (i.e. Figure 7). Even though, the online dynamics of significant variables that can trigger N_2O emissions in biological processes (i.e. COD, pH) were not available, the applied methodology enabled the identification of a set of variables that are connected with N_2O emissions in each sub-period (i.e. Figure 8). Considering that online data were not available for the influent of the carrousel reactor, higher NH_4-N loadings in the carrousel reactor were linked with clusters characterized by higher than average influent flow-rates and ammonium concentration and lower than average NO_3-N concentration in the plug-flow reactor. The latter can be supported by the fact that the behavior of variables in the carrousel reactor was significantly dependent on the nutrient concentrations in the plug-flow reactor (Table S4 – clustering results). Additionally, more intense aeration in the carrousel reactor (that can affect the stripping of dissolved N_2O) was linked with clusters characterized by higher than average NH_4-N concentration in the carrousel reactor (since the surface aerators were manipulated by the effluent ammonium concentration).

3.5 N_2O generation pathways

In line with Daelman et al. (2015) findings, both AOB pathways can be considered responsible for the N_2O emissions observed in the carrousel rector. The combination of nitrite accumulation and low oxygen concentrations can be linked with the nitrifier denitrification pathway, whereas higher AOR (ammonia oxidation rate), correlation of NH_4-N concentration in the carrousel reactor with N_2O emissions and higher DO concentrations can be linked with the hydroxylamine oxidation pathway (Law et al., 2012). N_2O generation via heterotrophic denitrification can be also significant especially in periods with nitrate accumulation, suggesting insufficient anoxic conditions (Daelman et. al., 2015).

In terms of the offline monitored variables, low pH, accompanied with nitrite accumulation, as observed in sub-period 4 has been identified as a significant factor inhibiting N_2O reduction during denitrification (Pan et al., 2012). Zhou et al. (2008) reported that under these conditions the production of free nitrous acid (FNA) in a denitrifying-Enhanced Biological Phosphorus Removal culture was the main contributor to N_2O emissions production even at low concentrations equal to 0.0007–0.001 mg HNO_2-N/L (nitrite concentration 3-4 mg/L at pH 7). Additionally, high pH values (>7) combined low DO concentration (~0.55 mg/L) have been reported to be responsible for nitrification driven N_2O emissions via the nitrifier denitrification pathway (Law et al., 2011). The latter is attributed to increasing ammonium oxidation rate (due to the pH increase), enhancing the nitrifier denitrification pathway through electrons provision. On the other hand, lower pH (<7) has been linked with elevated nitrification driven N_2O emissions at higher DO concentrations (~3 mg/L) (Li et al., 2015). The authors argued, that at higher pH the electrons available from the ammonium oxidation rate are mainly used to form water from molecular oxygen and H^+. In the current study, the pH in the effluent of the reactor was steady during the monitoring campaign (~8±0.2). However, online pH data showing the exact dynamics of the pH in the carrousel reactor were not available.

Low COD/N ratios have been reported to be responsible for denitrification induced N_2O emissions (Schulthess and Gujer, 1996). The offline data showed that COD/TKN ratio in the influent remained relatively steady during the monitoring campaign with a slight decrease in sub-periods 4 and 5 (<5) where emissions were higher (5.6 and 2.6 kg/h respectively). However, low COD/TKN (<5) was also observed in other sub-periods and did not result into high N_2O emissions (Figure 4). The frequency of the offline data (~6 days) did not enable the identification of the exact contribution of COD loading to the system. Figure 4 shows that COD limitation is not the sole contributor to the increased N_2O emissions in sub-period 4. Therefore, the results indicate that heterotrophic denitrification induced by COD/TN limitation was not the main N_2O emissions source in sub-periods 4 and 5.

The results from the application of multivariate statistical techniques can be used for the identification and explanation of potential pathways for N_2O generation. In sub-periods with lower average N_2O emission fluxes (1, 6, and 7), emission peaks coincided with ammonium peaks in the plug-flow reactor and therefore in the influent carrousel reactor. In that case, average emission fluxes ranged from 0.05 kg/h (sub-period 1) to 2.54 kg/h (sub-period 6). Wunderlin et al., (2012) demonstrated that N_2O production through hydroxylamine oxidation is accompanied by excess ammonia, low nitrite concentration and high ammonia oxidation rate. Additionally, in these sub-periods, N_2O emissions were higher at higher temperatures and DO concentrations. The high DO concentrations coincided with peaks in nitrite and nitrate concentrations indicating also insufficient denitrification zones in the reactor. AOB can use nitrite instead of oxygen as electron acceptor (Kampschreur et al., 2009a) especially in oxygen limiting conditions (low DO zones exist even when all surface aerators are under operation); thus, nitrifier denitrification by AOB could potentially contribute in N_2O emissions. Burgess et al. (2002) found strong dependency between nitrite accumulation and N_2O emissions, especially at sudden increase of ammonia loading.

Overall, N_2O emissions increased significantly and peaked at low nitrate concentrations in both reactors (i.e., sub-periods 3 and 4) and high nitrite concentrations in the carrousel reactor (i.e., sub-period 4). Under aerobic conditions, nitrite accumulates in the system when the ammonia oxidation rate to nitrite exceeds the nitrite oxidation rate to nitrate (Guisasola et al., 2005) inducing the nitrifier denitrification pathway. Sub-optimum

DO, COD and pH can also result in nitrite accumulation during denitrification (Schulthess et al., 1994; Yang et al., 2012). Zheng et al., (2015) observed a synergistic N_2O generation between nitrifier denitrification and heterotrophic denitrification in a pilot carrousel reactor where the nitrite built-up during denitrification boosted nitrifier denitrification pathway. The latter is in line with the N_2O profiles observed in this study in sub-periods with high emissions. The combined results of PCA and hierarchical k-means clustering can guide through the most significant N_2O production pathways in different sub-periods (supplementary material).

Conclusions

N_2O emissions depend on a set of interacting biological and chemical conversions and physical processes. This complex interaction obscures the determination of the governing processes in individual treatment plants. With multivariate analysis correlations between influential factors in a complex system might be revealed.

- A data-driven approach consisting of statistical-based methods was applied to analyze long-term N_2O emission dynamics and generation mechanisms based on available high temporal resolution (hourly) data. Applying binary segmentation to the N_2O emission profile allowed to split up the 15-month N_2O monitoring campaign into 10 sub-periods.
- Spearman's rank correlation analysis showed significant univariate correlations between N_2O emissions and ammonium, nitrate and nitrite concentrations. The correlation coefficients fluctuated between the 10 sub-periods. Low values for the correlation coefficients indicated non-monotonic interrelationships that Spearman's rank correlation cannot identify.
- Hierarchical k-means clustering provided information on the existence of reoccurring patterns and their effect on N_2O emissions. N_2O emission peaks were linked with the diurnal behavior of the nutrients' concentrations and with rain events, whereas low nitrate concentrations in the preceding plug flow reactor (<1 mg/L) resulted in increased ammonium loadings and high N_2O emissions in the subsequent carrousel reactor.
- Principal component analysis validated the findings from the clustering analysis and showed that ammonium, nitrate, nitrite, influent flow-rate and temperature, explained more than 65% of the variance in the system for the majority of the sub-periods. The first principal component corresponded to the control strategy of the reactor.
- The proposed methodological approach can detect and visualize disturbances in the system (i.e., precipitation events, high NH_4-N concentrations, etc.) and their effect on N_2O emissions.

Additionally, the ranges of operating variables that have historically resulted in low or high ranges of N_2O emissions can be identified. Overall, multivariate analysis can assist researchers and operators to understand and control the N_2O emissions using long term historical data.

Acknowledgements

This paper is supported by the Horizon 2020 research and innovation programme, SMART-Plant under grant agreement No 690323. The authors acknowledge Alex Sengers and David Philo from Hoogheemraadschap van Schieland en de Krimpenerwaard, the Water Board of Schieland and Krimpenerwaard. for sharing their knowledge regarding the Kralingseveer WWTP operation.

References

Aboobakar, A., Cartmell, E., Stephenson, T., Jones, M., Vale, P., Dotro, G., 2013. Nitrous oxide emissions and dissolved oxygen profiling in a full-scale nitrifying activated sludge treatment plant. Water Res. 47, 524–534. https://doi.org/10.1016/j.watres.2012.10.004

Abu-Jamous, B., Nandi, A.K., Fa, R., 2015. Integrative Cluster Analysis in Bioinformatics. John Wiley & Sons.

Adouani, N., Limousy, L., Lendormi, T., Sire, O., 2015. N_2O and NO emissions during wastewater denitrification step: Influence of temperature on the biological process. Comptes Rendus Chim., International Chemical Engineering Congress (ICEC) 2013: From fundamentals to applied chemistry and biochemistry 18, 15–22. https://doi.org/10.1016/j.crci.2014.11.005

Ahn, J.H., Kim, S., Park, H., Katehis, D., Pagilla, K., Chandran, K., 2010. Spatial and Temporal Variability in Atmospheric Nitrous Oxide Generation and Emission from Full-Scale Biological Nitrogen Removal and Non-BNR Processes. Water Environ. Res. 82, 2362–2372. https://doi.org/10.2175/106143010X12681059116897

Arai, K., Barakbah, A.R., 2007. Hierarchical K-means: an algorithm for centroids initialization for K-means. Rep. Fac. Sci. Eng. 36, 25–31.

Brotto, A.C., Kligerman, D.C., Andrade, S.A., Ribeiro, R.P., Oliveira, J.L.M., Chandran, K., Mello, W.Z. de, 2015. Factors controlling nitrous oxide emissions from a full-scale activated sludge system in the tropics. Environ. Sci. Pollut. Res. 22, 11840–11849. https://doi.org/10.1007/s11356-015-4467-x

Burgess, J.E., Colliver, B.B., Stuetz, R.M., Stephenson, T., 2002. Dinitrogen oxide production by a mixed culture of nitrifying bacteria during ammonia shock loading and aeration failure. J. Ind. Microbiol. Biotechnol. 29, 309–313.

Charrad, M., Ghazzali, N., Boiteau, V., Niknafs, A., 2014. NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. Journal of Statistical Software 61. https://doi.org/10.18637/jss.v061.i06

Chen, J., Gupta, A.K., 1997. Testing and Locating Variance Changepoints with Application to Stock Prices. J. Am. Stat. Assoc. 92, 739–747. https://doi.org/10.2307/2965722

Daelman, M.R.J., van Voorthuizen, E.M., van Dongen, U.G.J.M., Volcke, E.I.P., van Loosdrecht, M.C.M., 2015. Seasonal and diurnal variability of N_2O emissions from a full-scale municipal wastewater treatment plant. Sci. Total Environ. 536, 1–11. https://doi.org/10.1016/j.scitotenv.2015.06.122

Daelman, M.R.J., van Voorthuizen, E.M., van Dongen, L.G.J.M., Volcke, E.I.P., van Loosdrecht, M.C.M., 2013. Methane and nitrous oxide emissions from municipal wastewater treatment – results from a long-term study. Water Sci. Technol. 67, 2350–2355. https://doi.org/10.2166/wst.2013.109

Domingo-Félez, C., F. Smets, B., 2016. A consilience model to describe N 2 O production during biological N removal. Environ. Sci. Water Res. Technol. 2, 923–930. https://doi.org/10.1039/C6EW00179C

EEA Report, 2017. Annual European Union greenhouse gas inventory 1990–2015 and inventory report 2017. Technical report No 6/2017, European Environment Agency, Copenhagen, Denmark.

Elefsiniotis, P., Li, D., 2006. The effect of temperature and carbon source on denitrification using volatile fatty acids. Biochem. Eng. J. 28, 148–155.

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

Gernaey, K.V., van Loosdrecht, M.C., Henze, M., Lind, M., Jørgensen, S.B., 2004. Activated sludge wastewater treatment plant modelling and simulation: state of the art. Environ. Model. Softw. 19, 763–783.

Guisasola, A., Jubany, I., Baeza, J.A., Carrera, J., Lafuente, J., 2005. Respirometric estimation of the oxygen affinity constants for biological ammonium and nitrite oxidation. J. Chem. Technol. Biotechnol. 80, 388–396. https://doi.org/10.1002/jctb.1202

Guo, L., Vanrolleghem, P.A., 2014. Calibration and validation of an activated sludge model for greenhouse gases no. 1 (ASMG1): prediction of temperature-dependent N<Subscript>2</Subscript>O emission dynamics. Bioprocess Biosyst. Eng. 37, 151–163. https://doi.org/10.1007/s00449-013-0978-3

Haas, D.W. de, Pepperell, C., Foley, J., 2014. Perspectives on greenhouse gas emission estimates based on Australian wastewater treatment plant operating data. Water Sci. Technol. 69, 451–463. https://doi.org/10.2166/wst.2013.572

Haimi, H., Mulas, M., Corona, F., Marsili-Libelli, S., Lindell, P., Heinonen, M., Vahala, R., 2016. Adaptive data-derived anomaly detection in the activated sludge process of a large-scale wastewater treatment plant. Eng. Appl. Artif. Intell. 52, 65–80. https://doi.org/10.1016/j.engappai.2016.02.003

Haimi, H., Mulas, M., Corona, F., Vahala, R., 2013. Data-derived soft-sensors for biological wastewater treatment plants: An overview. Environ. Model. Softw. 47, 88–107. https://doi.org/10.1016/j.envsoft.2013.05.009

Hartigan, J.A., 1975. Clustering algorithms. CERN Doc. Serv. URL http://cds.cern.ch/record/105051.

Hartigan, J.A., Wong, M.A., 1979. Algorithm AS 136: A k-means clustering algorithm. Journal of the Royal Statistical Society. Series C (Applied Statistics) 28, 100–108.

Holtan-Hartwig, L., Dörsch, P., Bakken, L.R., 2002. Low temperature control of soil denitrifying communities: kinetics of N 2 O production and reduction. Soil Biol. Biochem. 34, 1797–1806.

IPCC, 2013. The physical science basis. Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change. USA: Cambridge University Press.

Jain, A.K., 2010. Data clustering: 50 years beyond K-means. Pattern Recognit. Lett., Award winning papers from the 19th International Conference on Pattern Recognition (ICPR) 31, 651–666. https://doi.org/10.1016/j.patrec.2009.09.011

Jolliffe, I.T., 2002. Principal component analysis and factor analysis, chap 7. Principal component analysis. Springer series in statistics, pp 150–166. Springer, New York

Jönsson, H., Junestedt, C., Willén, A., Yang, J., Tjus, K., Baresel, C., Rodhe, L., Trela, J., Pell, M., Andersson, S., 2015. Minska utsläpp av växthusgaser från rening av avlopp och hantering av avloppsslam. Sven. Vatten Utveckl. Rapp. 2015–02.

Kampschreur, M.J., Poldermans, R., Kleerebezem, R., Star, W.R.L. van der, Haarhuis, R., Abma, W.R., Jetten, M.S.M., Loosdrecht, M.C.M. van, 2009a. Emission of nitrous oxide and nitric oxide from a full-scale single-stage nitritation-anammox reactor. Water Sci. Technol. 60, 3211–3217. https://doi.org/10.2166/wst.2009.608

Kampschreur, M.J., Tan, N.C.G., Kleerebezem, R., Picioreanu, C., Jetten, M.S.M., Loosdrecht, M.C.M. van, 2008. Effect of Dynamic Process Conditions on Nitrogen Oxides Emission from a Nitrifying Culture. Environ. Sci. Technol. 42, 429–435. https://doi.org/10.1021/es071667p

Kampschreur, M.J., Temmink, H., Kleerebezem, R., Jetten, M.S.M., van Loosdrecht, M.C.M., 2009b. Nitrous oxide emission during wastewater treatment. Water Res. 43, 4093–4103. https://doi.org/10.1016/j.watres.2009.03.001

Kaufman, L., Rousseeuw, P.J., 1990. Finding groups in data. John Wiley and Sons, Inc., New York.

Kawahara, Y., Sugiyama, M., 2012. Sequential change-point detection based on direct density-ratio estimation. Stat. Anal. Data Min. ASA Data Sci. J. 5, 114–127.

Killick, R., Eckley, I., 2014. changepoint: An R package for changepoint analysis. J. Stat. Softw. 58, 1–19.

Killick, R., Fearnhead, P., Eckley, I.A., 2012. Optimal detection of changepoints with a linear computational cost. J. Am. Stat. Assoc. 107, 1590–1598. https://doi.org/10.1080/01621459.2012.737745

Kosonen, H., Heinonen, M., Mikola, A., Haimi, H., Mulas, M., Corona, F., Vahala, R., 2016. Nitrous Oxide Production at a Fully Covered Wastewater Treatment Plant: Results of a Long-Term Online Monitoring Campaign. Environ. Sci. Technol. 50, 5547–5554. https://doi.org/10.1021/acs.est.5b04466

Law, Y., Lant, P., Yuan, Z., 2011. The effect of pH on N 2 O production under aerobic conditions in a partial nitritation system. Water Res. 45, 5934–5944.

Law, Y., Ye, L., Pan, Y., Yuan, Z., 2012. Nitrous oxide emissions from wastewater treatment processes. Phil Trans R Soc B 367, 1265–1277. https://doi.org/10.1098/rstb.2011.0317

Lee, C., Choi, S.W., Lee, I.-B., 2004. Sensor fault identification based on time-lagged PCA in dynamic processes. Chemom. Intell. Lab. Syst. 70, 165–178. https://doi.org/10.1016/j.chemolab.2003.10.011

Li, P., Wang, S., Peng, Y., Liu, Y., He, J., 2015. The synergistic effects of dissolved oxygen and pH on N_2O production in biological domestic wastewater treatment under nitrifying conditions. Environ. Technol. 36, 1623–1631. https://doi.org/10.1080/09593330.2014.1002862

Liu, R.-T., Wang, X.-H., Zhang, Y., Wang, M.-Y., Gao, M.-M., Wang, S.-G., 2016. Optimization of operation conditions for the mitigation of nitrous oxide (N_2O) emissions from aerobic nitrifying granular sludge system. Environ. Sci. Pollut. Res. 23, 9518–9528. https://doi.org/10.1007/s11356-016-6178-3

Liu, Y., Pan, Y., Sun, Z., Huang, D., 2014. Statistical Monitoring of Wastewater Treatment Plants Using Variational Bayesian PCA. Ind. Eng. Chem. Res. 53, 3272–3282. https://doi.org/10.1021/ie403788v

Maere, T., Villez, K., Marsili-Libelli, S., Naessens, W., Nopens, I., 2012. Membrane bioreactor fouling behaviour assessment through principal component analysis and fuzzy clustering. Water Res. 46, 6132–6142. https://doi.org/10.1016/j.watres.2012.08.027

Mampaey, K.E., De Kreuk, M.K., van Dongen, U.G.J.M., van Loosdrecht, M.C.M., Volcke, E.I.P., 2016. Identifying N_2O formation and emissions from a full-scale partial nitritation reactor. Water Res. 88, 575–585. https://doi.org/10.1016/j.watres.2015.10.047

Mannina, G., Ekama, G., Caniani, D., Cosenza, A., Esposito, G., Gori, R., Garrido-Baserba, M., Rosso, D., Olsson, G., 2016. Greenhouse gases from wastewater treatment — A review of modelling tools. Sci. Total Environ. 551–552, 254–270. https://doi.org/10.1016/j.scitotenv.2016.01.163

Massara, T.M., Malamis, S., Guisasola, A., Baeza, J.A., Noutsopoulos, C., Katsou, E., 2017. A review on nitrous oxide (N 2 O) emissions during biological nutrient removal from municipal wastewater and sludge reject water. Sci. Total Environ. 596, 106–123.

Mirin, S.N.S., Wahab, N.A., 2014. Fault Detection and Monitoring Using Multiscale Principal Component Analysis at a Sewage Treatment Plant. J. Teknol. 70.

Monteith, H.D., Sahely, H.R., MacLean, H.L., Bagley, D.M., 2005. A Rational Procedure for Estimation of Greenhouse-Gas Emissions from Municipal Wastewater Treatment Plants. Water Environ. Res. 77, 390–403. https://doi.org/10.2175/106143005X51978

Moon, T.S., Kim, Y.J., Kim, J.R., Cha, J.H., Kim, D.H., Kim, C.W., 2009. Identification of process operating state with operational map in municipal wastewater treatment plant. J. Environ. Manage. 90, 772–778. https://doi.org/10.1016/j.jenvman.2008.01.008

Ni, B.-J., Pan, Y., van den Akker, B., Ye, L., Yuan, Z., 2015. Full-Scale Modeling Explaining Large Spatial Variations of Nitrous Oxide Fluxes in a Step-Feed Plug-Flow Wastewater Treatment Reactor. Environ. Sci. Technol. 49, 9176–9184. https://doi.org/10.1021/acs.est.5b02038

Olsson, G., Carlsson, B., Comas, J., Copp, J., Gernaey, K.V., Ingildsen, P., Jeppsson, U., Kim, C., Rieger, L., Rodriguez-Roda, I., others, 2014. Instrumentation, control and automation in wastewater–from London 1973 to Narbonne 2013. Water Sci. Technol. 69, 1373–1385.

Pan, Y., van den Akker, B., Ye, L., Ni, B.-J., Watts, S., Reid, K., Yuan, Z., 2016. Unravelling the spatial variation of nitrous oxide emissions from a step-feed plug-flow full scale wastewater treatment plant. Sci. Rep. 6. https://doi.org/10.1038/srep20792

Pan, Y., Ye, L., Ni, B.-J., Yuan, Z., 2012. Effect of pH on N_2O reduction and accumulation during denitrification by methanol utilizing denitrifiers. Water Res. 46, 4832–4840. https://doi.org/10.1016/j.watres.2012.06.003

Platikanov, S., Rodriguez-Mozaz, S., Huerta, B., Barceló, D., Cros, J., Batle, M., Poch, G., Tauler, R., 2014. Chemometrics quality assessment of wastewater treatment plant effluents using physicochemical parameters and UV absorption measurements. J. Environ. Manage. 140, 33–44.

Pocquet, M., Wu, Z., Queinnec, I., Spérandio, M., 2016. A two pathway model for N_2O emissions by ammonium oxidizing bacteria supported by the NO/N_2O variation. Water Res. 88, 948–959. https://doi.org/10.1016/j.watres.2015.11.029

R Core Team, 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria [WWW Document]. URL https://www.R-project.org/

Ravishankara, A.R., Daniel, J.S., Portmann, R.W., 2009. Nitrous oxide (N_2O): the dominant ozone-depleting substance emitted in the 21st century. science 326, 123–125.

Rodriguez-Caballero, A., Aymerich, I., Poch, M., Pijuan, M., 2014. Evaluation of process conditions triggering emissions of green-house gases from a biological wastewater treatment system. Sci. Total Environ. 493, 384–391. https://doi.org/10.1016/j.scitotenv.2014.06.015

Rosén, C., Lennox, J.A., 2001. Multivariate and multiscale monitoring of wastewater treatment operation. Water Res. 35, 3402–3410.

Rosén, C., Olsson, G., 1998. Disturbance detection in wastewater treatment plants. Water Sci. Technol. 37, 197–205.

Rosen, C., Yuan, Z., 2001. Supervisory control of wastewater treatment plants by combining principal component analysis and fuzzy c-means clustering. Water Sci. Technol. 43, 147–156.

Rousseeuw, P.J., 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. Journal of computational and applied mathematics 20, 53–65.

Rustum, R., Adeloye, A.J., Scholz, M., 2008. Applying Kohonen Self-Organizing Map as a Software Sensor to Predict Biochemical Oxygen Demand. Water Environ. Res. 80, 32–40. https://doi.org/10.2175/106143007X184500

Schulthess, R.V., Gujer, W., 1996. Release of nitrous oxide (N_2O) from denitrifying activated sludge: Verification and application of a mathematical model. Water Res. 30, 521–530. https://doi.org/10.1016/0043-1354(95)00204-9

Schulthess, R. von, Wild, D., Gujer, W., 1994. Nitric and nitrous oxides from denitrifying activated sludge at low oxygen concentration. Water Sci. Technol. 30, 123–132.

Scott, A.J., Knott, M., 1974. A cluster analysis method for grouping means in the analysis of variance. Biometrics 507–512.

Spearman, C., 1904. “General Intelligence,” Objectively Determined and Measured. Am. J. Psychol. 15, 201–292. https://doi.org/10.2307/1412107

Sun, S., Cheng, X., Sun, D., 2013. Emission of N_2O from a full-scale sequencing batch reactor wastewater treatment plant: Characteristics and influencing factors. Int. Biodeterior. Biodegrad. 85, 545–549. https://doi.org/10.1016/j.ibiod.2013.03.034

Ward Jr, J.H., 1963. Hierarchical grouping to optimize an objective function. J. Am. Stat. Assoc. 58, 236–244.

Wunderlin, P., Mohn, J., Joss, A., Emmenegger, L., Siegrist, H., 2012. Mechanisms of N_2O production in biological wastewater treatment under nitrifying and denitrifying conditions. Water Res. 46, 1027–1037. https://doi.org/10.1016/j.watres.2011.11.080

Yang, X., Wang, S., Zhou, L., 2012. Effect of carbon source, C/N ratio, nitrate and dissolved oxygen concentration on nitrite and ammonium production from denitrification process by Pseudomonas stutzeri D6. Bioresour. Technol. 104, 65–72. https://doi.org/10.1016/j.biortech.2011.10.026

Yu, R., Kampschreur, M.J., Loosdrecht, M.C.M. van, Chandran, K., 2010. Mechanisms and Specific Directionality of Autotrophic Nitrous Oxide and Nitric Oxide Generation during Transient Anoxia. Environ. Sci. Technol. 44, 1313–1319. https://doi.org/10.1021/es902794a

Zheng, M., Tian, Y., Liu, T., Ma, T., Li, L., Li, C., Ahmad, M., Chen, Q., Ni, J., 2015. Minimization of nitrous oxide emission in a pilot-scale oxidation ditch: Generation, spatial variation and microbial interpretation. Bioresour. Technol. 179, 510–517.

Zhou, Y.A.N., Pijuan, M., Zeng, R.J., Yuan, Z., 2008. Free nitrous acid inhibition on nitrous oxide reduction by a denitrifying-enhanced biological phosphorus removal sludge. Environmental Science & Technology 42, 8260–8265.

Table 1:Average value and standard deviation (std) of variables monitored in the Northern carrousel reactor (C: carrousel reactor, N: Northern, PF: plug-flow reactor)

Online variables	Average	Std	Offline variables	Average	Std
N_2O (kg/h)	1.4	2.1	COD influent (mg COD/ L)	238.8	79.5
NH_4-N C (mg/L)	1.63	2.2	TKN influent (mg/L)	42.1	10.0
NO_3-N C (mg/L)	5.8	4	TP influent (mg/ L)	7.0	2.1
NO_2-N C (mg/L)	1.2	1.1	Flow-rate (m^3/ d)	85,898	41,786
DO1 (mg/L)	0.6	0.9	COD effluent (mg/ L)	36.9	6.9
DO2 (mg/L)	0.8	0.9	TKN efffluent (mg/ L)	2.8	1.2
DO3 (mg/L)	1.9	0.6	TP effluent (mg/ L)	1.1	0.6
Temperature (°C)	16	3.5	pH effluent	8.0	0.2
N_2O PF (kg/h)	0.71	1.21			
NH_4-N PF (mg/L)	12.41	5.35			
NO_3-N PF (mg/L)	2.38	2.2			
Influent Flow-rate (m^3/h)	3973	2375			
DO PF (mg/L)	2.61	0.65			

Table 2: Average values and standard deviations of the main variables for the 10 sub-periods (C: carrousel reactor, N: Northern, PF: plug-flow reactor).

Variables	N_2O (kg/h)		NO_3-C N (mg/l)		NO_3-N PF (mg/l)		NH_4-N C (mg/l)		NH_4-N PF (mg/l)		NO_2-N C* (mg/l)		Temperature (°C)		DO1 (mg/l)		DO2 (mg/l)		DO3 (mg/l)	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
1	0	0.1	6.1	3.1	1.8	1.6	1.8	2.67	11.4	4.1			15.7	1.4	0.62	0.7	0.62	0.5	1.5	0.4
2	0.6	0.6	7.2	3.1	2.5	2	1.5	1.7	13	4			11.2	1.0	0.77	1	1.31	0.8	2	0.4
3	2.7	1.4	6.1	3.2	1.6	2.1	1.6	2.1	15.2	4.5			11.5	0.7	0.67	0.8	1.49	1	2.07	0.4
4	5.6	2. 6	3	0.1	0.5	0.7	1.3	1.6	15	4.8	2.6	1.9	12.9	1.1	0.64	0.9	1.95	0.9	1.9	0.4
5	2.6	2.2	4.3	4.2	3.1	1.9	1.3	2	11.5	5.2	0.8	1	18.2	1.7	0.34	0.7	0.39	0.8	1.94	0.5
6	0.8	1.4	3.3	3.2	2.3	1.9	2	3.1	14.7	6.1	0.5	0.5	20	1.0	0.42	0.7	0.26	0.5	2.27	0.5
7	0.2	0.3	7.2	5	2.8	2.4	2	3.1	9.8	5.2	0.6	0.4	20	0.7	0.42	0.6	0.29	0.4	2.64	0.5
8	0.1	0.2	10.1	5.7	5.2	2.6	1.4	1	9.6	5.5	0.8	0.5	19.6	0.5	0.27	0.5	0.2	0.5	2.71	0.6
9	0.1	0.2	7.9	3.6	2.8	2.8	2	2	13.2	5.4	1.9	0.8	12.9	2.1	1.12	1.2	1.07	1	1.58	0.4
10	1.3	1.1	6.3	3.5	1.4	0.9	1.6	3.7	16.4	4.3	2.1	0.9	13	0.7	0.58	1.0	1.04	1	1.52	0.3

*NO_2-N concentration was monitored between 11/03/2011 and 19/01/2012

Table 3: Operating variables (average) for all clusters defined by hierarchical clustering in the carrousel reactor (P: Sub-period, Cl: Clusters)

P	Cl	N_2O C	NH_4-N PF	NO_3-N PF	Influent	NH_4-N C	NO_3-N C	DO1	DO2	DO3	NO_2-N
		kg/h	mg/l	mg/l	m^3/h	mg/l	mg/l	mg/l	mg/l	mg/l	mg/l
1	1	0.09	14.13	1.48	3883	1.47	8.66	1.04	0.78	1.72	
	2	0.01	8.55	2.41	3824	0.87	4.26	0.13	0.47	1.25	
	3	0.05	14.74	0.30	8892	7.91	4.63	1.37	0.77	1.58	
2	4	0.87	15.30	2.05	3827	1.51	8.61	0.94	1.53	2.22	
	5	0.21	9.13	3.69	3419	0.74	5.28	0.03	0.62	1.41	
	6	0.24	12.51	0.81	11132	4.52	5.42	2.27	2.31	2.22	
3	7	3.22	16.85	1.52	3383	1.36	7.36	0.87	1.88	2.35	
	8	1.72	10.96	1.91	3672	0.82	4.29	0.05	0.85	1.56	
	9	2.40	21.40	0.12	7935	7.52	4.15	2.10	1.28	2.10	
4	10	6.60	17.30	0.32	3207	1.26	3.79	2.14	0.95	2.41	4.10
	11	3.83	10.82	0.77	2747	0.79	1.80	1.51	0.05	1.20	1.40
	12	6.89	25.45	0.48	6375	10.86	3.62	1.98	2.12	2.34	4.28
6	15	2.54	17.66	0.75	5922	5.00	5.07	1.30	0.73	2.34	1.08
	16	0.51	8.20	2.84	3811	0.98	2.64	0.10	0.10	2.21	0.35

*NO_2-N concentration was monitored between 11/03/2011 and 19/01/2012

Table 4: PCA loadings sub-period 2, carrousel reactor

Variable	PC1	PC2	PC3	PC4
NH_4-N PF	-0.28	0.47	-0.24	0.29
NO_3-N PF	0.36	0.21	0.14	-0.67
Influent	-0.38	-0.31	-0.09	-0.37
NH_4-N C	-0.34	0.03	-0.59	-0.29
NO_3-N C	-0.04	0.58	0.21	-0.31
DO1	-0.43	0.06	-0.15	-0.18
DO2	-0.40	0.08	0.48	-0.17
DO3	-0.37	0.21	0.40	0.28
Temperature	0.22	0.49	-0.33	0.11

Table 5: PCA loadings sub-period 4, carrousel reactor

	PC1	PC2	PC3	PC4
NH_4-N PF	-0.48	0.04	-0.11	0.25
NO_3-N PF	0.26	0.56	-0.04	-0.35
Influent	-0.33	-0.07	-0.52	-0.17
NH_4-N C	-0.28	0.14	-0.50	-0.46
NO_3-N C	-0.17	0.59	0.32	0.04
DO1	-0.37	0.24	-0.13	0.59
DO2	-0.40	0.08	0.41	-0.14
DO3	-0.37	0.01	0.33	-0.40
Temperature	0.23	0.51	-0.27	0.19

Relating N_2O emissions during biological nitrogen removal with operating conditions using multivariate statistical techniques

Vasilaki V.[a], Volcke, E.I.P.[b], Nandi A.K.[c], van Loosdrecht[d], M.C.M., Katsou E.[a*]

[a] Department of Civil & Environmental Engineering, Brunel University London, Uxbridge UB8 3PH, UK

[b] Department of Green Chemistry and Technology, Ghent University, Coupure Links 653, 9000 Gent, Belgium

[c] Department of Electronic and Computer Engineering, Brunel University London, Uxbridge UB8 3PH, UK

[d] Department of Biotechnology, Delft University of Technology, Van der Maasweg 9, 2629 HZ Delft, The Netherlands

*Corresponding author. Department of Civil & Environmental Engineering, Brunel University London, Uxbridge UB8 3PH, UK. Email: evina.katsou@brunel.ac.uk

Keywords: N_2O emissions, long-term monitoring campaign, principal component analysis, hierarchical k-means clustering

Abstract

Multivariate statistical analysis was applied to investigate the dependencies and underlying patterns between N_2O emissions and online operational variables (dissolved oxygen and nitrogen component concentrations, temperature and influent flow-rate) during biological nitrogen removal from wastewater. The system under study was a full-scale reactor, for which hourly sensor data were available. The 15-month long monitoring campaign was divided into 10 sub-periods based on the profile of N_2O emissions, using Binary Segmentation. The dependencies between operating variables and N_2O emissions fluctuated according to Spearman's rank correlation. The correlation between N_2O emissions and nitrite concentrations ranged between 0.51-0.78. Correlation > 0.7 between N_2O emissions and nitrate concentrations was observed at sub-periods with average temperature lower than 12 °C. Hierarchical k-means clustering and principal component analysis linked N_2O emission peaks with precipitation events and ammonium concentrations higher than 2 mg/L, especially in sub-periods characterized by low N_2O fluxes. Additionally, the highest ranges of measured N_2O fluxes belonged to clusters corresponding with NO_3-N concentration less than 1 mg/L in the upstream plug-flow reactor (middle of oxic zone), indicating slow nitrification rates. The results showed that the range of N_2O emissions partially depend on the-prior behavior of the system. The principal component analysis validated the findings from the clustering analysis and showed that ammonium, nitrate, nitrite and temperature explained a considerable percentage of the variance in the system for the majority of the sub-periods. The applied statistical methods, linked the different ranges of emissions with the system variables, provided insights on the effect of operating conditions on N_2O emissions in each sub-period and can be integrated into N_2O emissions data processing at wastewater treatment plants.

Abbreviations

AOR: Ammonia oxidation rate

CH_4: Methane

CO_2: Carbon dioxide

DO: Dissolved oxygen

GHG: Greenhouse gas

N_2O: Nitrous oxide

NH_4-N: Ammonium nitrogen

NO_2-N: Nitrite nitrogen

NO_3-N: Nitrate nitrogen

PC: Principal component

PCA: Principal component analysis

PLS: Partial least squares

TN: Total nitrogen

WWTP: Wastewater treatment plant

1. **Introduction**

The increasing demand to reduce the carbon footprint of municipal wastewater treatment plants (WWTPs) by reducing greenhouse gas (GHG) emissions and energy consumption, is posing new challenges for the water industry (Flores-Alsina et al., 2014). The climate change pressures prompt the quantification and minimization of GHG emissions generated in WWTPs (Haas et al., 2014). Three main sources of GHG emissions prevail in WWTPs (Monteith et al., 2005; Mannina et al., 2016): (i) the direct emissions mainly linked to biological processes, (ii) the indirect internal emissions generated by the use of imported energy to the plants, and (iii) the indirect external emissions associated with the sources that are controlled outside the WWTPs (e.g. chemicals production, disposal of sewage sludge, transportation). The GHGs emitted into the atmosphere from biological wastewater treatment processes are carbon dioxide (CO_2), methane (CH_4) and nitrous oxide (N_2O) (Kampschreur et al., 2009b).

With the potential contribution of 265 times more than CO_2 for a 100-year time horizon to global warming (IPCC, 2013), N_2O is a potent GHG and the most significant contributor to ozone depletion (Ravishankara et al., 2009). WWTPs are significant generators of N_2O and are responsible for 3.1% of the N_2O emissions in Europe (EEA Report, 2017). N_2O is generated mainly during the autotrophic nitrification and heterotrophic denitrification (Kampschreur et al., 2008) and can contribute up to 78% (Daelman et al., 2013) of the footprint of a WWTP's operation. Recent studies have focused on the understanding, quantification, control and minimization of N_2O emissions (Aboobakar et al., 2013; Mampaey et al., 2016; Pan et al., 2016). However, several studies have resulted in contradicting findings on the influence of operating and environmental variables on N_2O generation (Liu et al., 2016; Massara et al., 2017). For instance, several studies have reported increasing N_2O emissions with decreasing DO concentrations during nitrification (Kampschreur et al., 2009b). However, Rodriguez-Caballero et al. (2014) found that N_2O emission profiles in a full-scale biological reactor did not change even for DO variations higher than 1.5 mg/L. The latter, was attributed to the high nitrification efficiency and the potential biomass adaptation to continuously varying DO concentrations. Results from real-field N_2O monitoring campaigns cannot fully explain long-term causes of N_2O emissions and the combined effect of operating, environmental and external factors that influence the biological systems

(Jönsson et al., 2015). Long-term full-scale monitoring campaigns have shown that N_2O fluxes are highly dynamic with significant diurnal fluctuations and seasonal variations; however, the dynamics cannot be fully explained (Daelman et al., 2015; Kosonen et al., 2016).

Several mechanistic process models describing N_2O emissions from wastewater treatment plants have been developed over the last few years (Massara et al., 2017). While they have been successfully applied to identify N_2O formation mechanisms and pathways from experimental data (Ni et al., 2015; Pocquet et al., 2016), their calibration and validation to long-term process data remains a challenge. Domingo-Félez and F. Smets (2016) reported that substrate affinity constants for NO_2 and NO reduction in existing N_2O models differ by a factor of about 100. Additionally, calibration of models under specific operational conditions (i.e. dry weather) can affect their performance and accuracy when the system varies (Gernaey et al., 2004; Guo and Vanrolleghem, 2014). Moreover, full-scale N_2O emission data show long-term trends that cannot be explained by commonly available operational data (Daelman et al., 2015) but are possibly caused by microbial population changes, which are hard to catch with the current models, typically describing single functional groups with fixed parameter sets. Multivariate statistical techniques are capable of identifying relationships between N_2O emissions and a multitude of influencing factors, at the same time identifying various operating sub-periods for which this behaviour may differ. This will lead to increased understanding of experimental data, on its turn facilitating the application, calibration and validation of mechanistic models. As such, multivariate statistical techniques maximize the information acquired from N_2O monitoring campaign data.

Statistical techniques have been used for the analysis of data from full-scale monitoring campaigns, to identify interconnections between operating and environmental variables on the one hand and N_2O formation on the other hand. Through multiple linear regression analyses, Aboobakar et al. (2013) showed dependencies between N_2O emissions and nitrogen load, temperature and dissolved oxygen (DO) in various compartments of a plug-flow reactor for biological nitrogen removal. Multi-regression analysis of one year of data with bi-monthly sampling frequency, coming from a full-scale SBR (Sun et al., 2013) indicated negative correlation between N_2O emissions and temperature, while COD/N ratio lower than 6 resulted in higher emissions. Brotto et al. (2015) used Spearman's rank correlation to explain the behavior of N_2O emissions in an activated sludge process. The analysis showed negative correlation between N_2O emissions and pH but positive correlation between N_2O fluxes and temperature. However, most of the studies did not consider continuous long-term operational data, while further analysis is required to gain a better understanding on the dynamics and trade-offs between N_2O generation and the online monitored and controlled process variables.

Multivariate analysis has been proven to be a suitable method for the identification of patterns and hidden relationships within WWTP data (Rosén and Lennox, 2001) and can be applied to provide insights on the combined effect of operational variables on N_2O emissions in full-scale systems. Chemometric techniques have been applied to the wastewater treatment sector for 40 years (Rosén and Olsson, 1998), enabling the visualization and interpretation of the multi-dimensional interrelations of the operational variables monitored in biological processes (Platikanov et al., 2014). Their application can (i) improve the efficiency of process monitoring (Mirin and Wahab, 2014) and provide further insights of the biological processes (Moon et al., 2009), (ii) identify and isolate process faults (Haimi et al., 2016; Liu et al., 2014; Maere et al., 2012; Rosen and Yuan, 2001), sensor faults (Lee et al., 2004), and iii) predict significant operating variables in the biological systems that affect performance (Rustum et al., 2008). Furthermore, the gradual implementation of online sensors to monitor important parameters in the biological treatment train of WWTPs results in the production of time series, which require the application of specific statistical tools for their interpretation. The most widely applied approaches include methods aiming to reduce the dimensionality of large data-sets (i.e., principal component analysis (PCA), partial least squares (PLS)) and data clustering techniques (i.e., hierarchical clustering, k-means clustering) (Haimi et al., 2013). However, there are limited studies investigating the behavior of N_2O emissions with the application of multivariate statistical techniques, especially utilizing online operational data in long-term monitoring.

The aim of this work is to investigate whether widely applied multivariate statistical techniques can be applied to the online data collected from real-field N_2O monitoring campaigns in order to gain a better understanding on the dynamic behavior of N_2O emissions and explain the combined effect of the operating variables monitored in wastewater treatment processes on N_2O emissions. Hourly data from the operating variables monitored online and N_2O emissions data in a full-scale carrousel reactor from the long-term monitoring campaign published by Daelman et al. (2015) were used for the analysis. A statistical methodological approach was developed, applying changepoint detection techniques to identify changes in the N_2O fluxes behavior combined with hierarchical k-means clustering and PCA, to provide insights on N_2O emissions patterns and generation pathways.

2. **Materials and methods**

2.1 Process description and data origin

This work was based on the data obtained by Daelman et al. (2015) for the Kralingseveer WWTP, consisting of a plug-flow reactor followed by two carrousel reactors in parallel (Figure 1). The plant treated 80.000 m^3 d^{-1} of domestic wastewater from a combined sewer system. The carrousel reactors were characterized by alternating anoxic/oxic zones; aeration was performed through surface aerators, which were manipulated to control the ammonium concentration in the effluent. Aerator 1 operates under on/off pattern, being on when the ammonium concentration was higher than 1.2 mg N/L), while surface aerators 2 and 3 were always operational to keep the solids from settling but operated at maximum capacity when the ammonium concentration became higher than 0.6 and 0.9 mg/L, respectively. Over the monitoring period the average total nitrogen (TN) removal efficiency was 81 ±10%; the average COD removal efficiency was equal to 87 ±5%.

Ammonium nitrogen (NH_4-N), nitrate nitrogen (NO_3-N) and DO were monitored in the middle of the second oxic zone in the plug flow reactor (location 1, Figure 1). The carrousel reactors were equipped with, NH_4-N, temperature probes, and 3 DO probes (DO1, DO2, DO3) (locations 2, 3, 4, Figure 1). The Northern carrousel reactor was also equipped with a nitrite probe. All the reactors were covered, and the off-gas was collected in ducts and pumped to a Servomex gas analyzer, where N_2O was measured. Table S1 lists all the variables monitored online (Supplementary material). The data matrix developed consists of the variables monitored in the carrousel reactor (DO, NH_4-N C, NO_3-N C, NO_2-N C, N_2O C), the influent flow-rate, as well as the NH_4-N and NO_3-N concentrations from the plug-flow reactor. 24 h composite samples of influent and effluent, available about every 6 days, were used to support the analysis. Figure 2, summarizes the methodological framework applied to the online database.

[Figure 1]

2.2 Methodological framework for data analysis

The monitoring period was divided into distinct sub-periods based on the profile of N_2O fluxes in the carrousel reactor. Spearman's correlation analysis, k-means clustering, hierarchical clustering, and Principal component analysis were applied to the database. The application of clustering algorithms facilitated the identification of operational modes that have historically resulted in specific ranges of N_2O emissions. The PCA reduced the dimensionality of the data-set transforming the sensor signals into useful knowledge that that can be easily interpreted. The methodological framework is extensively described in the following sub-sections.

[Figure 2]

The data-driven approach enabled the utilization of the information and patterns embedded in the real-time monitored variables (from the system sensors) in the biological processes and GHG measurements. Multivariate statistical analysis is an alternative to univariate analysis that is commonly applied for the analysis of WWTP data. It enables the identification of patterns and interrelations in data-sets by examining multiple variables simultaneously (Olsson et al., 2014). R software was used for the statistical analysis (R Core Team, 2017). The complete list of packages used is provided in the supplementary material (Table S2).

2.2.1 *Preliminary data processing*

The preliminary data analysis included: (i) data synchronization under the same time-stamp, and ii) removal of duplicate and unreliable measurements (multiple readings at the same time stamp for the same sensor). The data were aggregated into hourly averages in order to compensate for the missing data due to variation in sampling frequency between the different variables monitored. Exponential moving average imputation was applied when less than 24 consequential data were missing for each variable. Longer periods of missing data were excluded from the analysis.

2.2.2 *Binary segmentation changepoint detection*

Given a series of data, change point analysis investigates abrupt changes in a data-series when specific properties change (i.e., mean and variance) (Kawahara and Sugiyama, 2012). The Binary Segmentation (Scott and Knott, 1974) is a widely applied and computationally efficient changepoint detection algorithm (Killick et al., 2012). The algorithm employs initially single changepoint detection method to the complete data-set as described in (Killick and Eckley, 2014). If a changepoint is identified the procedure is repeated to the two new segments formed; before and after the changepoint. The process continues splitting the data until there are no more changepoints identified. The computational cost of the algorithm is of the order of O(nlog n) with n being the number of data in the data-set and therefore it is applicable in large data-sets. A distribution-free test statistic was applied based on the work of Chen and Gupta, (1997). The penalty for the changepoints identification was equal to log(n). The algorithm requires independent data points. Therefore, first difference transformation of the N_2O timeseries was performed and changes in variance were identified by the Binary segmentation algorithm. The profile of the N_2O emissions was highly variable during the monitoring campaign. Binary segmentation enabled the identification of the sub-periods characterized by different N_2O emissions' profile.

2.2.3 *Spearman's rank correlation*

Spearman's rank correlation coefficient (Spearman, 1904) was used to detect bivariate temporal monotonic trends among the system variables for the different sub-periods; it served as a measure of the association strength. This method is based on the rank of the values and therefore, is less sensitive to outliers than Pearson's correlation. P values lower than .01 were considered to be significant.

2.2.4 *Hierarchical k-means clustering*

Clustering techniques are widely applied in data mining in order to identify and group the underling patterns that exist in high dimensional data sets (Jain, 2010). K-means clustering (Hartigan and Wong, 1979) is a recognized clustering algorithm (Haimi at al., 2013). K-means clustering was applied to categorize the data in groups of similar observations and to investigate the patterns of N_2O emission fluxes, based on Euclidean distance. K-means algorithm begins with the selection of k random centroids of the same dimension within the original data. All the data-points are compared and assigned to the nearest centroid. During each iteration, the nearest data to each centroid are re-defined and centroids are recalculated in a way that squared distances of all points within a cluster to the cluster's centroid are minimized. However, the randomly selected initial centroids can result into locally optimized clustering results (Abu-Jamous et al., 2015). Therefore, hierarchical k-means clustering that was proposed by Arai and Barakbah, (2007), was applied to the dataset. In this method agglomerative hierarchical clustering (Kaufman and Rousseeuw, 1990) is applied for the selection of the centroids; Ward's method is used in order to divide the dataset in clusters (Ward Jr, 1963). The data were normalized before the analysis. NBclust package in R (Charrad et al., 2014) was used to select the number of clusters in each sub-period. The package applies a number cluster validity indexes (i.e. average silhouette value (Rousseeuw, 1987); Hartigan's rule (Hartigan, 1975)).

Hierarchical k-means clustering was applied to the carrousel reactor data matrix from the different sub-periods identified through binary segmentation, to investigate whether different temporal patterns of the operating variables were responsible for the different behavior of N_2O emissions. Hierarchical k-means clustering enabled i) the detection of frequency and persistence of extreme ranges of operating variables, and ii) the comparison of the operational modes between the plug-low and carrousel reactor. Ammonium and nitrate probes in the plug-flow reactor were included in the analysis, since they can provide indirect feedback in terms of the carrousel reactor influent and additional information for the operational behavior of the system. However, the analysis was repeated excluding plug-flow variables (NH_4-N and NO_3-N). Graphical comparisons of the clustered data-points versus time and boxplots of the variables in each identified cluster are displayed in the results' section.

2.2.5 *Principal component analysis*

Principal component analysis (PCA) (Jolliffe, 2002) was applied to the dataset in an effort to reduce the dimensionality of the data by eliminating a small proportion of variance in the data. PCA transforms the original correlated measured variables to uncorrelated variables, i.e., Principal components (PCs), explaining the maximum observed variability. The principal components are linear combinations of the original data variables. The loadings of the variables in each principal component can map their relationship with the respective principal component. PC scores are a linear combination of the data, weighted by the PC loadings for each variable. The scores of the principal components map the different samples in the new dimensional space of the principal components facilitating the investigation of the different relationships between the variables. The data matrices (X) consisting of J columns (variables) and I data rows (number of observations) were normalized with mean equal to 0 and standard deviation equal to 1. Each column of **X**, $x_j =$

$(x_{1j}, \dots x_{Ij})T$, j=1,...J, represents a vector in the I-dimensional space. In PCA, eigenvalue decomposition is used to factorize the data matrix X *(I x J)* and to map the data matrix to a reduced dimensional space*:*

$$X = TP^T + E$$

where, *T:* matrix *(I x S)* representing the score of the principal components, *S:* the number of principal components selected, *P:* matrix *(J x S)* representing the loadings and *E*: matrix of residuals.

The biplot of the first 2 PCs was used in order to visualize the combined behavior of significant variables that affect the system. The biplots enabled the simultaneous visualization of i) the variables' loadings in the first two principal components, ii) the scores of the first two principal components, and iii) the different clusters. The temporal variations of the PC scores enabled the identification of occasions in which the behavior of the system changes. PCA was applied to the data matrix of the carrousel reactor excluding N_2O emissions time series, i) to identify the most significant variables that affect the system, (ii) to analyze the structure of the sensor data, iii) to investigate if changes in the relationship of the system coincide with changes in the N_2O emissions profile, and iv) to validate the results from hierarchical clustering. N_2O emissions time series were excluded from the PCA in order to investigate the relationship between the PC scores and N_2O emissions and to examine which PCs are most significantly linked to the behavior of N_2O emissions.

3. Results and discussion

3.1 N_2O emissions profile and main dependencies

The profile of all the variables monitored was fluctuating during the monitoring period, which can justify the different profiles of N_2O emissions that resulted from the Binary Segmentation algorithm. Overall, high ranges of emissions were reported when nitrate concentration in the plug-flow reactor was low, whereas periods with lower ammonium concentrations in the plug-flow reactor were linked with lower N_2O emissions.

Table 1 shows the average values and standard deviations of the variables monitored online and offline in the Northern carrousel and plug-flow reactors. N_2O fluxes peaked in March 2011 followed by a period characterized by very low N_2O emissions. Gradual decrease was observed until November 2011 and negligible emissions again until January 2011 (Figure 3).

[Table 1]

The application of Binary Segmentation algorithm to the N_2O emissions of the Northern carrousel reactor identified 9 changepoints that correspond to 10 sub-periods with distinct variance of the N_2O timeseries first difference (Figure 3). The analysis identified abrupt temporal changes in the emission dynamics that indicate changes in the underlying mechanisms or environmental conditions responsible for the N_2O formation.

[Figure 3]

Offline data were analyzed in the different sub-periods in order to investigate significant changes that can contribute to the high N_2O emissions in sub-periods 4 and 5. The average COD concentration in the influent of the plug-flow reactor (effluent of primary sedimentation) was 239 ± 80 mg COD/L over the 15-month monitoring period. The average plug-flow reactor influent and carrousel reactor effluent concentrations of COD, TKN, BOD, TP and the effluent pH for all sub-periods are given in the supplementary material (Table S3). In sub-period 5, 27% increase in the influent COD concentration to the plug flow reactor (compared to average value) was observed, which could be attributed to less precipitation events and to the consequently lower average influent flow-rate during this sub-period. Laboratory analyses did not show significant seasonal changes in the plug-flow COD loading (19,934 ± 13310 kg COD/day). The COD loading in sub-period 4 (16,160 ± 2546 kg COD/day) was 17% less than in sub-period 1. TKN and TP loadings were reduced in sub-period 4 compared to sub-period, by 11% and 12% respectively. The COD:TKN:TP ratio remained quite stable, ranging between 1:0.17:0.02 (sub-period 2) and 1:0.20:0.03 (sub-period 4).

Figure 4 shows the different COD to TKN ratios measured for all the sub-periods. There were cases with lower than average COD/TKN in the influent of the plug-flow reactor that coincided with increased N_2O emissions, particularly in sub-periods 4 and 5. However, low ranges of COD/TKN (<5) in sub-periods 1, 2, 7 and 6 corresponded with low N_2O emissions. These observations indicate that limitation of COD cannot be considered the sole contributor of N_2O emissions via heterotrophic denitrification in sub-periods 4 and 5.

[Figure 4]

The COD removal efficiency remained relatively steady during the monitoring campaign ranging from 79% (sub-period 8) to 91% (sub-period 5). The range of TN and TP removal efficiencies ranged from 73 % (sub-periods 1 and 9) to 92% (sub-period 5) and from 67% (sub-period 7) to 87% (sub-period 4). The effluent pH was steady (~ 8) and did not show seasonal variability that could influence the generation of N_2O emissions.

On the other hand, a significant variation is observed for all variables monitored online by analyzing at the complete database. Table 2 summarizes the average values and standard deviations of the online monitored variables considered in the analysis for the target periods. In the carrousel reactor, the nitrite concentration is relatively high in sub-period 4 (average = 2.6 mg/L) and in the first part of sub-period 10 (average = 2.1 mg/L). The average temperature in both cases is ~13 °C. In biological reactors operating in continuous mode, appreciable (> 2 mg N/L) nitrite concentrations are usually not observed, since nitrite is directly oxidized by nitrite oxidizing bacteria into nitrate. However, in certain cases, high nitrite concentrations in biological processes have been observed, which have been linked with low temperatures that affect N_2O reductase during denitrification enhancing N_2O production (Holtan-Hartwig et al., 2002; Adouani et al., 2015).

Analyzing the whole profile, the emissions tended to be low at higher temperatures (sub-periods 6, 7, and 8). Higher emissions were also observed, though, at temperature higher than 18 °C and low nitrite concentrations (i.e., sub-period 5). Ahn et al. (2010) demonstrated that N_2O emissions can be significant at higher temperatures due to the higher enzymatic activities of the bioprocesses producing N_2O. In the carrousel reactor during sub-periods 4 and 5, the temperature increases from 11.8 to 20 °C. Low N_2O emissions were also observed when ammonium concentration was lower than 13 mg/L and nitrate was higher than 2.5 mg/L in the plug-flow reactor. The probe was located in the middle of the second oxic zone; thus, lower ammonium concentrations in the plug-flow reactor can indicate less ammonium loads in the carrousel reactor.

[Table 2]

The analysis of the variables' ranges for the N_2O emission profiles provides limited insight on the dependencies between the system variables monitored online, which is further analyzed in the following sections.

3.2 Spearman's rank correlation analysis for carrousel reactor

The application of Spearman's rank correlation coefficient to the data of the carrousel reactor could not identify significant correlations between the N_2O emissions and the operating variables. The lack of monotonic univariate dependencies could be attributed to i) the temporal fluctuations of the influent characteristics, ii) the continuous variability in the operating conditions of the reactors, and iii) the seasonal variations of the environmental conditions in wastewater treatment processes. Fluctuating correlation coefficients between N_2O emissions and carrousel reactor variables were identified (Supplementary, Figures S1:S2). The findings are in line with the study of Kosonen et al., (2016). The authors compared the results from two monitoring periods at the same biological system and identified different relationships between N_2O emissions and $BOD_{7(ATU)}$ loads.

The correlation coefficient between nitrite and N_2O emissions ranged from 0.78 (sub-period 7) to 0.51 (sub-period 9). As a general remark, nitrite was correlated with N_2O emissions in sub-periods 4, 6 and 7, while lower correlation was observed during sub-periods 5 (Figure 5), 8 and 9. N_2O emissions and NO_3-N concentration in the carrousel reactor exhibited a positive correlation with coefficient higher than 0.7 for sub-periods 2 (Figure 5), 4 and 10 (the temperature was lower than 13 °C in all cases). N_2O emissions and NO_3-N concentrations followed similar diurnal patterns, wherein peaks in nitrate concentration coincided with peaks in N_2O emissions (Daelman et al., 2015). The accumulation of nitrate is potentially linked with higher nitrification than denitrification rates. This is in line with Daelman et al. (2015), considering that the nitrate utilization rate in these sub-periods is affected by the low temperatures (Elefsiniotis and Li, 2006). Additionally, during times when N_2O was positively correlated with DO1 (> 0.5), medium to significant correlations between the N_2O emissions and the ammonium concentration in the carrousel reactor were also observed (sub-periods 1, 6 and 7). Stripping of the already formed N_2O can be a potential explanation. Given that the surface aerator in the location of DO1 probe is manipulated to control the ammonium concentration in the effluent, ammonium peaks trigger the surface aerators to start.

The correlation coefficient between any two of the system variables did not remain stable between the different sub-periods. Figure 5 shows the correlograms for sub-periods 2 and 5. These sub-periods were characterized by low and high ranges of N_2O emissions and temperature respectively (Table 2). In sub-period 2, the average NO_3-N concentration in the plug-flow reactor was equal to 2.5 mg/L (Table 2) and correlated negatively with the influent flow-rate (~ - 0.63) (Figure 5). In sub-period 5 the behavior of nitrate concentration (average equal to 2.1 mg/L) was mainly correlated negatively with ammonium concentration in the same reactor. The ammonium concentration in the carrousel reactor was positively correlated with DO1 only in sub-period 2. NH_4-N concentration in the plug-flow reactor was correlated with the influent-flow rate only in sub-periods 4 and 5. However, the profiles of these two variables showed that in the majority of the sub-periods, abrupt and rapid increase of influent flow-rate (i.e., precipitation events) coincided with increase of the NH_4-N. However, the NH_4-N concentration reduced more rapidly in the system than the influent flow-rate. For example, in sub-period 3 the correlation coefficient between NH_4-N in the plug-flow reactor and influent flow-rate was 0.26. However, when days with significant precipitation events (and thus high influent flow-rate) were omitted, the correlation coefficient was equal to 0.58. The latter shows that, in this example, the lack of correlation between these two variables is most likely to be an indication that the interrelationships are not monotonic and that the method is not appropriate to identify complex relationships within the data. In order to verify that increased influent flow-rate was linked with precipitation events, daily precipitation data were extracted from the Royal Netherlands meteorological institute. Spearman's correlation coefficient between two days moving average of influent flow-rate and daily precipitation in the Netherlands was equal to 0.69. Therefore, there is a direct link between higher than average flow-rates and precipitation events (the timeseries are shown in Figure S3, supplementary material). The correlograms for all sub-periods are provided in the Supplementary material (Figures S1:S2).

Spearman's rank correlation indicated structural changes in the dependencies between the system variables. Therefore, the fluctuating structural dependencies had a different impact on the generation of N_2O emissions. Previous studies have shown that various monitored variables in the biological system (NH_4-N, NO_3-N, NO_2-N, Temperature) can affect N_2O emissions generation. However, further analysis is required to investigate their combined effect in N_2O formation in full-scale complex systems.

[Figure 5]

3.3 Hierarchical k-means clustering

The application of hierarchical k-means clustering enabled the categorization of the different ranges of the operating variables and N_2O emissions within each sub-period.

Hierarchical k-means clustering analysis was repeated excluding NH_4-N and NO_3-N concentrations in the plug-flow reactor. The results showed that the majority of the data points were allocated to the same clusters for each sub-period even when the NH_4-N and NO_3-N concentrations in the plug-flow reactor were excluded. In the majority of the sub-periods (i.e. sub-periods 1-6) more than 85% of the data points were assigned to the same cluster. It can be concluded that specific patterns and ranges of NH_4-N and NO_3-N monitored in plug-flow reactor, systematically resulted in specific responses to the carrousel reactor. The latter is supported by the Spearman's rank correlation analysis, where high correlations were observed between the variables in the two reactors for several sub-periods. For example, the correlation coefficient between NH_4-N in the plug-flow and carrousel reactors is higher than 0.7 for sub-periods 1 to 7. The similarity of the clusters for all the sub-periods is shown in Table S4 in the Supporting Material.

The range of N_2O emissions was differentiated in the majority of the clusters. In all the sub-periods, two major clusters were identified characterized by significant differences in the NH_4-N and NO_3-N concentrations in the plug-flow reactor. In the majority of the sub-periods they represented the diurnal variability of the system nutrient concentrations and influent-flow rate. Additionally, clustering distinguished occasions with high influent flow-rate and ammonium concentration in the carrousel reactor, which can be an indication of precipitation events. In sub-periods characterized by low average N_2O emissions (i.e., 1, 2, 7, 8 and 9), clusters with increased N_2O emissions (yet relatively low) were mainly linked to higher loading rates due to the expected diurnal variability or to precipitation events. However, N_2O emissions higher than 3.8 kg/h were observed when the average NO_3-N concentration was constantly lower than 1 mg/L in the plug-flow reactor and the NO_3-N concentration was lower than 4 mg/L in the carrousel reactor. Table 3 compares the clustered average values for all the variables in sub-period 2 (average N_2O emissions equal to 0.6 kg/h – Table 2) and 4 (average N_2O emissions equal to 5.6 kg/h – Table 2). The average value of N_2O emissions for a set of clusters in a specific sub-period (from Table 3) can be found taking into account the number of data-points in the individual clusters. Sub-period 4 was characterized by very low NO_3-N concentration in the middle of the oxic zone in the plug-flow reactor. The latter indicates slower oxidation of ammonia to nitrate or insufficient DO in the plug-flow nitrification lane. This can lead to higher NH_4-N loading in the carrousel reactor. On the other hand, higher nitrification rates in the plug-flow reactor (i.e. sub-period 2) resulted in lower N_2O emissions in the carrousel reactor. The average values of all the variables in each cluster during all the sub-periods are given as supplementary material (Table S5).

In clusters 2 and 16 the averages of operating variables had similar ranges (Table 3). However, in these two occasions the N_2O emissions were different (0.01 and 0.51 kg/h). Similarly, in clusters 1, 4 and 7, the averages of operating variables were similar yet the N_2O emissions were different (0.09,0.87 and 3.22 kg/h respectively). A corollary to this also existed. In clusters 1 and 2 the averages of operating variables were different but the N_2O emissions were similar (0.09 and 0.01). Similarly, in clusters 5 and 6 the averages of operating variables were different but the N_2O emissions were similar (0.21 and 0.24). Such observations indicate the underlying complexities of the interdependencies. Additionally, it can be concluded that the range of N_2O emissions can partially depend on the preceding operational mode of the system. Figure 6 shows an example of the variables monitored online for two separate occasions in sub-periods 2 and 3 (from 00:00 am until 8:00 am) and the respective N_2O emissions. All the variables showed a similar behavior (in terms of range and trends). N_2O emission profiles had also the same trend; however, their range depended on the initial N_2O fluxes at 00:00 am. The influent flow-rates, NH_4-N and NO_3-N concentrations in the plug-flow reactor also were similar in these two occasions. The average N_2O fluxes were equal to 0.44 and 2.01 kg/h for occasion 1 and 2 respectively. More extensive data are required for quantitative investigation.

[Table 3]

[Figure 6]

3.4 Principal component analysis in the carrousel reactor

PCA was applied to transform the original correlated measured variables to uncorrelated variables (Principal components) and explain the maximum observed variability. In sub-periods with low emissions (1, 2, 7, 8, and 9) the PCA analysis showed that N_2O emissions' peaks are related with NH_4-N and influent flow-rate peaks in the carrousel reactor and with the effect of the diurnal variability of these variables' loading rates.

The current section discusses the PCA results for sub-period 2, as an example. The results for all the sub-periods are given in the supplementary material (Tables S6-S13, Figures S4-S29). The application of PCA reduced the dimensionality of the data with 4 principal components (PCs) explaining ~86% of the total variance (PC1 = 39%, PC2 = 26%, PC3 = 12%, and PC4 = 9%). Loadings for the system variables in the 4 PCs are given in Table 4. The loadings of each component are an indication of the variation in the variables explained by a specific component. Influent flow-rate, ammonium concentration in the carrousel reactor (NH_4-N C) and the three DO (DO1, DO2 and DO3) concentrations had the highest negative loadings in PC1. This means that the first principal component increased with the increase of these variables. Nitrate concentration (NO_3-N PF) in the plug-flow reactor has a relatively high positive loading in PC1 (0.36). Therefore, PC1 describes how the carrousel reactor responds to the behavior of the upstream plug-flow reactor processes and conditions, the variation of the influent flow-rate and variations in ammonium and DO concentrations in the carrousel reactor. The latter can be indirectly connected with the control strategy of the carrousel reactor, since the surface aerators were manipulated based on the effluent ammonium concentration. PC2 linked ammonium concentration in the plug-flow reactor, nitrate concentration in the carrousel reactor and temperature (loadings higher than 0.47). In PC3 ammonium concentration in the carrousel reactor had high negative loading, while DO2 and DO3 concentrations had positive loadings that was not expected considering the control strategy of the system. Investigation of the variables' profiles, though, showed an increasing trend of DO2 and DO3, whereas the ammonium profile did not present a similar trend.

[Table 4]

The biplot of the first 2 PCs is used to visualize the combined behavior of significant variables that affect the system. Data points assigned to cluster 6 (Figure 7), had negative scores in PC2 and PC1. Therefore, ammonium concentration in the carrousel reactor and influent flow rate were higher than average, while the nitrate concentration in the system was lower than average. Figure 8 shows the profile of N_2O emissions and NH_4-N in the carrousel reactor for sub-period 2. The colored points in the diagram represent the identified clusters. Peaks in emissions coincided with peaks in the NH_4-N C profile, whereas peaks in NH_4-N C coincided with precipitation events (cluster 6).

[Figure 7]

The scores of the data-points in cluster 5 were mainly positive in PC1 and negative in PC2 (Figure 7). PC2 increased with the increase of NH_4-N concentration in the plug-flow reactor (Table 4). Given that PC2 had an average equal to 0 (data are standardized), data-points with negative scores in PC2 represent occasions with lower than average NH_4-N concentration in the plug-flow reactor. This is supported by the correlation plot (Figure 7), where the arrow of NH_4-N concentration in the plug-flow reactor points to the direction of increasing concentrations of NH_4-N. Therefore, data-points belonging to cluster 5 were characterized by higher than average ammonium concentration in the plug-flow reactor. Similarly, NO_3-N concentration in the plug-flow reactor had relatively significant positive loading in PC1 (0.36 – Table 4). The latter indicates that NH_4-N and DO concentrations (measured by three probes) in the carrousel reactor (that had negative loadings in PC1 – Table 5) tended to decrease when NO_3-N concentration in the plug-flow reactor increased. Given that all data-points in cluster 5 had positive scores in PC1, it can be concluded that they are characterized by lower than average NH_4-N concentration in the carrousel reactor and higher than average NO_3-N concentration in the plug-flow reactor. According to the clustering results the latter can be an indication of the high nitrogen loadings of the normal diurnal variability in the reactor. This finding is supported from the results presented in Figure 8, where the data-points of cluster 5 correspond to the daily low range of ammonium concentrations in both reactors.

[Figure 8]

Figure 9 summarizes scores of the PC2 and the respective clusters (colored points in the diagram) indicating strong diurnal cyclic fluctuations of the water quality during this sub-period. It also shows that after each precipitation event, a sudden temperature drop occurred; the system was disturbed and cannot recover immediately. Spearman's rank correlation coefficient between PC2 and N_2O emissions is equal to 0.72.

[Figure 9]

In sub-period 4, mechanisms triggering high N_2O emissions in the carrousel reactor prevailed (average = 5.6 kg/h). The PCA loadings were similar to sub-period 2, while the clustering results indicated 3 clusters; clusters 10 and 11 were affected by the diurnal variability and cluster 12 was affected by the precipitation events (Table 3). Again, the DO data obtained from the 3 sensors in the carrousel reactor had significant negative loadings in PC1. However, ammonium concentration in the carrousel reactor was not identified as a significant variable affecting the system in the first two PCs. This can be attributed to the fact that less NH_4-N concentration peaks were observed in the effluent of the carrousel reactor (17 data points belong to cluster 12). The correlation coefficient of PC1with NH_4-N concentration in the carrousel reactor was -0.75. Therefore, PCA analysis shows that PC1 is a good indicator of the ammonium concentration in the carrousel reactor. The DO concentrations in this sub-period especially for cluster 10 (with average NH_4-N concentration in the carrousel reactor equal to 1.26 mg/L) was the highest observed in all the clusters with similar NH_4-N concentrations in the carrousel effluent. The alternation of aerobic and anaerobic conditions observed in this reactor, combined with high NH_4-N and DO concentrations has been identified as a significant cause of nitrification sourced emissions (Yu et al., 2010).

[Table 5]

In PC2, the NO_3-N concentration and temperature had significant positive loadings (Table 5). The score plot of PC2 (Figure 10a) presented an increasing trend and therefore, showed that nitrate and temperature increased. The latter was verified by the profiles of NO_3-N concentrations in the carrousel reactor (Figure 10b) and NO_3-N concentration and temperature in the plug-flow reactor (Supplementary material S30). In the beginning of the sub-period 4 very low concentrations of nitrate were observed in the system and they gradually increased especially after the 28th of March. The Spearman's correlation coefficient between N_2O emissions and PC2 scores were relatively high and equal to 0.62. However, contrary to sub-period 2, the clustering analysis showed that there is no nitrate accumulation (Table 3). The average nitrate concentration in the plug-flow reactor was equal to 0.2 mg/L until the 28th of March and increased up to 1.6 mg/L until the end of the sub-period. Therefore, the observations in section 3.3 are supported by the PCA results (low nitrate in the plug flow resulted in increased loadings in the subsequent carrousel reactor and the denitrification activity in the carrousel reactor is affected by the low temperature resulting in nitrite accumulation).

[Figure 10]

In the section, the combination of hierarchical k-means clustering and PCA was used in order to link the different emission ranges with all the online monitored variables (i.e. Figure 7). Even though, the online dynamics of significant variables that can trigger N_2O emissions in biological processes (i.e. COD, pH) were not available, the applied methodology enabled the identification of a set of variables that are connected with N_2O emissions in each sub-period (i.e. Figure 8). Considering that online data were not available for the influent of the carrousel reactor, higher NH_4-N loadings in the carrousel reactor were linked with clusters characterized by higher than average influent flow-rates and ammonium concentration and lower than average NO_3-N concentration in the plug-flow reactor. The latter can be supported by the fact that the behavior of variables in the carrousel reactor was significantly dependent on the nutrient concentrations in the plug-flow reactor (Table S4 – clustering results). Additionally, more intense aeration in the carrousel reactor (that can affect the stripping of dissolved N_2O) was linked with clusters characterized by higher than average NH_4-N concentration in the carrousel reactor (since the surface aerators were manipulated by the effluent ammonium concentration).

3.5 N_2O generation pathways

In line with Daelman et al. (2015) findings, both AOB pathways can be considered responsible for the N_2O emissions observed in the carrousel rector. The combination of nitrite accumulation and low oxygen concentrations can be linked with the nitrifier denitrification pathway, whereas higher AOR (ammonia oxidation rate), correlation of NH_4-N concentration in the carrousel reactor with N_2O emissions and higher DO concentrations can be linked with the hydroxylamine oxidation pathway (Law et al., 2012). N_2O generation via heterotrophic denitrification can be also significant especially in periods with nitrate accumulation, suggesting insufficient anoxic conditions (Daelman et. al., 2015).

In terms of the offline monitored variables, low pH, accompanied with nitrite accumulation, as observed in sub-period 4 has been identified as a significant factor inhibiting N_2O reduction during denitrification (Pan et al., 2012). Zhou et al. (2008) reported that under these conditions the production of free nitrous acid (FNA) in a denitrifying-Enhanced Biological Phosphorus Removal culture was the main contributor to N_2O emissions production even at low concentrations equal to 0.0007–0.001 mg HNO_2-N/L (nitrite concentration 3-4 mg/L at pH 7). Additionally, high pH values (>7) combined low DO concentration (~0.55 mg/L) have been reported to be responsible for nitrification driven N_2O emissions via the nitrifier denitrification pathway (Law et al., 2011). The latter is attributed to increasing ammonium oxidation rate (due to the pH increase), enhancing the nitrifier denitrification pathway through electrons provision. On the other hand, lower pH (<7) has been linked with elevated nitrification driven N_2O emissions at higher DO concentrations (~3 mg/L) (Li et al., 2015). The authors argued, that at higher pH the electrons available from the ammonium oxidation rate are mainly used to form water from molecular oxygen and H^+. In the current study, the pH in the effluent of the reactor was steady during the monitoring campaign (~8±0.2). However, online pH data showing the exact dynamics of the pH in the carrousel reactor were not available.

Low COD/N ratios have been reported to be responsible for denitrification induced N_2O emissions (Schulthess and Gujer, 1996). The offline data showed that COD/TKN ratio in the influent remained relatively steady during the monitoring campaign with a slight decrease in sub-periods 4 and 5 (<5) where emissions were higher (5.6 and 2.6 kg/h respectively). However, low COD/TKN (<5) was also observed in other sub-periods and did not result into high N_2O emissions (Figure 4). The frequency of the offline data (~6 days) did not enable the identification of the exact contribution of COD loading to the system. Figure 4 shows that COD limitation is not the sole contributor to the increased N_2O emissions in sub-period 4. Therefore, the results indicate that heterotrophic denitrification induced by COD/TN limitation was not the main N_2O emissions source in sub-periods 4 and 5.

The results from the application of multivariate statistical techniques can be used for the identification and explanation of potential pathways for N_2O generation. In sub-periods with lower average N_2O emission fluxes (1, 6, and 7), emission peaks coincided with ammonium peaks in the plug-flow reactor and therefore in the influent carrousel reactor. In that case, average emission fluxes ranged from 0.05 kg/h (sub-period 1) to 2.54 kg/h (sub-period 6). Wunderlin et al., (2012) demonstrated that N_2O production through hydroxylamine oxidation is accompanied by excess ammonia, low nitrite concentration and high ammonia oxidation rate. Additionally, in these sub-periods, N_2O emissions were higher at higher temperatures and DO concentrations. The high DO concentrations coincided with peaks in nitrite and nitrate concentrations indicating also insufficient denitrification zones in the reactor. AOB can use nitrite instead of oxygen as electron acceptor (Kampschreur et al., 2009a) especially in oxygen limiting conditions (low DO zones exist even when all surface aerators are under operation); thus, nitrifier denitrification by AOB could potentially contribute in N_2O emissions. Burgess et al. (2002) found strong dependency between nitrite accumulation and N_2O emissions, especially at sudden increase of ammonia loading.

Overall, N_2O emissions increased significantly and peaked at low nitrate concentrations in both reactors (i.e., sub-periods 3 and 4) and high nitrite concentrations in the carrousel reactor (i.e., sub-period 4). Under aerobic conditions, nitrite accumulates in the system when the ammonia oxidation rate to nitrite exceeds the nitrite oxidation rate to nitrate (Guisasola et al., 2005) inducing the nitrifier denitrification pathway. Sub-optimum

DO, COD and pH can also result in nitrite accumulation during denitrification (Schulthess et al., 1994; Yang et al., 2012). Zheng et al., (2015) observed a synergistic N_2O generation between nitrifier denitrification and heterotrophic denitrification in a pilot carrousel reactor where the nitrite built-up during denitrification boosted nitrifier denitrification pathway. The latter is in line with the N_2O profiles observed in this study in sub-periods with high emissions. The combined results of PCA and hierarchical k-means clustering can guide through the most significant N_2O production pathways in different sub-periods (supplementary material).

Conclusions

N_2O emissions depend on a set of interacting biological and chemical conversions and physical processes. This complex interaction obscures the determination of the governing processes in individual treatment plants. With multivariate analysis correlations between influential factors in a complex system might be revealed.

- A data-driven approach consisting of statistical-based methods was applied to analyze long-term N_2O emission dynamics and generation mechanisms based on available high temporal resolution (hourly) data. Applying binary segmentation to the N_2O emission profile allowed to split up the 15-month N_2O monitoring campaign into 10 sub-periods.
- Spearman's rank correlation analysis showed significant univariate correlations between N_2O emissions and ammonium, nitrate and nitrite concentrations. The correlation coefficients fluctuated between the 10 sub-periods. Low values for the correlation coefficients indicated non-monotonic interrelationships that Spearman's rank correlation cannot identify.
- Hierarchical k-means clustering provided information on the existence of reoccurring patterns and their effect on N_2O emissions. N_2O emission peaks were linked with the diurnal behavior of the nutrients' concentrations and with rain events, whereas low nitrate concentrations in the preceding plug flow reactor (<1 mg/L) resulted in increased ammonium loadings and high N_2O emissions in the subsequent carrousel reactor.
- Principal component analysis validated the findings from the clustering analysis and showed that ammonium, nitrate, nitrite, influent flow-rate and temperature, explained more than 65% of the variance in the system for the majority of the sub-periods. The first principal component corresponded to the control strategy of the reactor.
- The proposed methodological approach can detect and visualize disturbances in the system (i.e., precipitation events, high NH_4-N concentrations, etc.) and their effect on N_2O emissions.

Additionally, the ranges of operating variables that have historically resulted in low or high ranges of N_2O emissions can be identified. Overall, multivariate analysis can assist researchers and operators to understand and control the N_2O emissions using long term historical data.

Acknowledgements

This paper is supported by the Horizon 2020 research and innovation programme, SMART-Plant under grant agreement No 690323. The authors acknowledge Alex Sengers and David Philo from Hoogheemraadschap van Schieland en de Krimpenerwaard, the Water Board of Schieland and Krimpenerwaard. for sharing their knowledge regarding the Kralingseveer WWTP operation.

References

Aboobakar, A., Cartmell, E., Stephenson, T., Jones, M., Vale, P., Dotro, G., 2013. Nitrous oxide emissions and dissolved oxygen profiling in a full-scale nitrifying activated sludge treatment plant. Water Res. 47, 524–534. https://doi.org/10.1016/j.watres.2012.10.004

Abu-Jamous, B., Nandi, A.K., Fa, R., 2015. Integrative Cluster Analysis in Bioinformatics. John Wiley & Sons.

Adouani, N., Limousy, L., Lendormi, T., Sire, O., 2015. N_2O and NO emissions during wastewater denitrification step: Influence of temperature on the biological process. Comptes Rendus Chim., International Chemical Engineering Congress (ICEC) 2013: From fundamentals to applied chemistry and biochemistry 18, 15–22. https://doi.org/10.1016/j.crci.2014.11.005

Ahn, J.H., Kim, S., Park, H., Katehis, D., Pagilla, K., Chandran, K., 2010. Spatial and Temporal Variability in Atmospheric Nitrous Oxide Generation and Emission from Full-Scale Biological Nitrogen Removal and Non-BNR Processes. Water Environ. Res. 82, 2362–2372. https://doi.org/10.2175/106143010X12681059116897

Arai, K., Barakbah, A.R., 2007. Hierarchical K-means: an algorithm for centroids initialization for K-means. Rep. Fac. Sci. Eng. 36, 25–31.

Brotto, A.C., Kligerman, D.C., Andrade, S.A., Ribeiro, R.P., Oliveira, J.L.M., Chandran, K., Mello, W.Z. de, 2015. Factors controlling nitrous oxide emissions from a full-scale activated sludge system in the tropics. Environ. Sci. Pollut. Res. 22, 11840–11849. https://doi.org/10.1007/s11356-015-4467-x

Burgess, J.E., Colliver, B.B., Stuetz, R.M., Stephenson, T., 2002. Dinitrogen oxide production by a mixed culture of nitrifying bacteria during ammonia shock loading and aeration failure. J. Ind. Microbiol. Biotechnol. 29, 309–313.

Charrad, M., Ghazzali, N., Boiteau, V., Niknafs, A., 2014. NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. Journal of Statistical Software 61. https://doi.org/10.18637/jss.v061.i06

Chen, J., Gupta, A.K., 1997. Testing and Locating Variance Changepoints with Application to Stock Prices. J. Am. Stat. Assoc. 92, 739–747. https://doi.org/10.2307/2965722

Daelman, M.R.J., van Voorthuizen, E.M., van Dongen, U.G.J.M., Volcke, E.I.P., van Loosdrecht, M.C.M., 2015. Seasonal and diurnal variability of N_2O emissions from a full-scale municipal wastewater treatment plant. Sci. Total Environ. 536, 1–11. https://doi.org/10.1016/j.scitotenv.2015.06.122

Daelman, M.R.J., van Voorthuizen, E.M., van Dongen, L.G.J.M., Volcke, E.I.P., van Loosdrecht, M.C.M., 2013. Methane and nitrous oxide emissions from municipal wastewater treatment – results from a long-term study. Water Sci. Technol. 67, 2350–2355. https://doi.org/10.2166/wst.2013.109

Domingo-Félez, C., F. Smets, B., 2016. A consilience model to describe N 2 O production during biological N removal. Environ. Sci. Water Res. Technol. 2, 923–930. https://doi.org/10.1039/C6EW00179C

EEA Report, 2017. Annual European Union greenhouse gas inventory 1990–2015 and inventory report 2017. Technical report No 6/2017, European Environment Agency, Copenhagen, Denmark.

Elefsiniotis, P., Li, D., 2006. The effect of temperature and carbon source on denitrification using volatile fatty acids. Biochem. Eng. J. 28, 148–155.

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

Gernaey, K.V., van Loosdrecht, M.C., Henze, M., Lind, M., Jørgensen, S.B., 2004. Activated sludge wastewater treatment plant modelling and simulation: state of the art. Environ. Model. Softw. 19, 763–783.

Guisasola, A., Jubany, I., Baeza, J.A., Carrera, J., Lafuente, J., 2005. Respirometric estimation of the oxygen affinity constants for biological ammonium and nitrite oxidation. J. Chem. Technol. Biotechnol. 80, 388–396. https://doi.org/10.1002/jctb.1202

Guo, L., Vanrolleghem, P.A., 2014. Calibration and validation of an activated sludge model for greenhouse gases no. 1 (ASMG1): prediction of temperature-dependent N<Subscript>2</Subscript>O emission dynamics. Bioprocess Biosyst. Eng. 37, 151–163. https://doi.org/10.1007/s00449-013-0978-3

Haas, D.W. de, Pepperell, C., Foley, J., 2014. Perspectives on greenhouse gas emission estimates based on Australian wastewater treatment plant operating data. Water Sci. Technol. 69, 451–463. https://doi.org/10.2166/wst.2013.572

Haimi, H., Mulas, M., Corona, F., Marsili-Libelli, S., Lindell, P., Heinonen, M., Vahala, R., 2016. Adaptive data-derived anomaly detection in the activated sludge process of a large-scale wastewater treatment plant. Eng. Appl. Artif. Intell. 52, 65–80. https://doi.org/10.1016/j.engappai.2016.02.003

Haimi, H., Mulas, M., Corona, F., Vahala, R., 2013. Data-derived soft-sensors for biological wastewater treatment plants: An overview. Environ. Model. Softw. 47, 88–107. https://doi.org/10.1016/j.envsoft.2013.05.009

Hartigan, J.A., 1975. Clustering algorithms. CERN Doc. Serv. URL http://cds.cern.ch/record/105051.

Hartigan, J.A., Wong, M.A., 1979. Algorithm AS 136: A k-means clustering algorithm. Journal of the Royal Statistical Society. Series C (Applied Statistics) 28, 100–108.

Holtan-Hartwig, L., Dörsch, P., Bakken, L.R., 2002. Low temperature control of soil denitrifying communities: kinetics of N 2 O production and reduction. Soil Biol. Biochem. 34, 1797–1806.

IPCC, 2013. The physical science basis. Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change. USA: Cambridge University Press.

Jain, A.K., 2010. Data clustering: 50 years beyond K-means. Pattern Recognit. Lett., Award winning papers from the 19th International Conference on Pattern Recognition (ICPR) 31, 651–666. https://doi.org/10.1016/j.patrec.2009.09.011

Jolliffe, I.T., 2002. Principal component analysis and factor analysis, chap 7. Principal component analysis. Springer series in statistics, pp 150–166. Springer, New York

Jönsson, H., Junestedt, C., Willén, A., Yang, J., Tjus, K., Baresel, C., Rodhe, L., Trela, J., Pell, M., Andersson, S., 2015. Minska utsläpp av växthusgaser från rening av avlopp och hantering av avloppsslam. Sven. Vatten Utveckl. Rapp. 2015–02.

Kampschreur, M.J., Poldermans, R., Kleerebezem, R., Star, W.R.L. van der, Haarhuis, R., Abma, W.R., Jetten, M.S.M., Loosdrecht, M.C.M. van, 2009a. Emission of nitrous oxide and nitric oxide from a full-scale single-stage nitritation-anammox reactor. Water Sci. Technol. 60, 3211–3217. https://doi.org/10.2166/wst.2009.608

Kampschreur, M.J., Tan, N.C.G., Kleerebezem, R., Picioreanu, C., Jetten, M.S.M., Loosdrecht, M.C.M. van, 2008. Effect of Dynamic Process Conditions on Nitrogen Oxides Emission from a Nitrifying Culture. Environ. Sci. Technol. 42, 429–435. https://doi.org/10.1021/es071667p

Kampschreur, M.J., Temmink, H., Kleerebezem, R., Jetten, M.S.M., van Loosdrecht, M.C.M., 2009b. Nitrous oxide emission during wastewater treatment. Water Res. 43, 4093–4103. https://doi.org/10.1016/j.watres.2009.03.001

Kaufman, L., Rousseeuw, P.J., 1990. Finding groups in data. John Wiley and Sons, Inc., New York.

Kawahara, Y., Sugiyama, M., 2012. Sequential change-point detection based on direct density-ratio estimation. Stat. Anal. Data Min. ASA Data Sci. J. 5, 114–127.

Killick, R., Eckley, I., 2014. changepoint: An R package for changepoint analysis. J. Stat. Softw. 58, 1–19.

Killick, R., Fearnhead, P., Eckley, I.A., 2012. Optimal detection of changepoints with a linear computational cost. J. Am. Stat. Assoc. 107, 1590–1598. https://doi.org/10.1080/01621459.2012.737745

Kosonen, H., Heinonen, M., Mikola, A., Haimi, H., Mulas, M., Corona, F., Vahala, R., 2016. Nitrous Oxide Production at a Fully Covered Wastewater Treatment Plant: Results of a Long-Term Online Monitoring Campaign. Environ. Sci. Technol. 50, 5547–5554. https://doi.org/10.1021/acs.est.5b04466

Law, Y., Lant, P., Yuan, Z., 2011. The effect of pH on N 2 O production under aerobic conditions in a partial nitritation system. Water Res. 45, 5934–5944.

Law, Y., Ye, L., Pan, Y., Yuan, Z., 2012. Nitrous oxide emissions from wastewater treatment processes. Phil Trans R Soc B 367, 1265–1277. https://doi.org/10.1098/rstb.2011.0317

Lee, C., Choi, S.W., Lee, I.-B., 2004. Sensor fault identification based on time-lagged PCA in dynamic processes. Chemom. Intell. Lab. Syst. 70, 165–178. https://doi.org/10.1016/j.chemolab.2003.10.011

Li, P., Wang, S., Peng, Y., Liu, Y., He, J., 2015. The synergistic effects of dissolved oxygen and pH on N_2O production in biological domestic wastewater treatment under nitrifying conditions. Environ. Technol. 36, 1623–1631. https://doi.org/10.1080/09593330.2014.1002862

Liu, R.-T., Wang, X.-H., Zhang, Y., Wang, M.-Y., Gao, M.-M., Wang, S.-G., 2016. Optimization of operation conditions for the mitigation of nitrous oxide (N_2O) emissions from aerobic nitrifying granular sludge system. Environ. Sci. Pollut. Res. 23, 9518–9528. https://doi.org/10.1007/s11356-016-6178-3

Liu, Y., Pan, Y., Sun, Z., Huang, D., 2014. Statistical Monitoring of Wastewater Treatment Plants Using Variational Bayesian PCA. Ind. Eng. Chem. Res. 53, 3272–3282. https://doi.org/10.1021/ie403788v

Maere, T., Villez, K., Marsili-Libelli, S., Naessens, W., Nopens, I., 2012. Membrane bioreactor fouling behaviour assessment through principal component analysis and fuzzy clustering. Water Res. 46, 6132–6142. https://doi.org/10.1016/j.watres.2012.08.027

Mampaey, K.E., De Kreuk, M.K., van Dongen, U.G.J.M., van Loosdrecht, M.C.M., Volcke, E.I.P., 2016. Identifying N_2O formation and emissions from a full-scale partial nitritation reactor. Water Res. 88, 575–585. https://doi.org/10.1016/j.watres.2015.10.047

Mannina, G., Ekama, G., Caniani, D., Cosenza, A., Esposito, G., Gori, R., Garrido-Baserba, M., Rosso, D., Olsson, G., 2016. Greenhouse gases from wastewater treatment — A review of modelling tools. Sci. Total Environ. 551–552, 254–270. https://doi.org/10.1016/j.scitotenv.2016.01.163

Massara, T.M., Malamis, S., Guisasola, A., Baeza, J.A., Noutsopoulos, C., Katsou, E., 2017. A review on nitrous oxide (N 2 O) emissions during biological nutrient removal from municipal wastewater and sludge reject water. Sci. Total Environ. 596, 106–123.

Mirin, S.N.S., Wahab, N.A., 2014. Fault Detection and Monitoring Using Multiscale Principal Component Analysis at a Sewage Treatment Plant. J. Teknol. 70.

Monteith, H.D., Sahely, H.R., MacLean, H.L., Bagley, D.M., 2005. A Rational Procedure for Estimation of Greenhouse-Gas Emissions from Municipal Wastewater Treatment Plants. Water Environ. Res. 77, 390–403. https://doi.org/10.2175/106143005X51978

Moon, T.S., Kim, Y.J., Kim, J.R., Cha, J.H., Kim, D.H., Kim, C.W., 2009. Identification of process operating state with operational map in municipal wastewater treatment plant. J. Environ. Manage. 90, 772–778. https://doi.org/10.1016/j.jenvman.2008.01.008

Ni, B.-J., Pan, Y., van den Akker, B., Ye, L., Yuan, Z., 2015. Full-Scale Modeling Explaining Large Spatial Variations of Nitrous Oxide Fluxes in a Step-Feed Plug-Flow Wastewater Treatment Reactor. Environ. Sci. Technol. 49, 9176–9184. https://doi.org/10.1021/acs.est.5b02038

Olsson, G., Carlsson, B., Comas, J., Copp, J., Gernaey, K.V., Ingildsen, P., Jeppsson, U., Kim, C., Rieger, L., Rodriguez-Roda, I., others, 2014. Instrumentation, control and automation in wastewater–from London 1973 to Narbonne 2013. Water Sci. Technol. 69, 1373–1385.

Pan, Y., van den Akker, B., Ye, L., Ni, B.-J., Watts, S., Reid, K., Yuan, Z., 2016. Unravelling the spatial variation of nitrous oxide emissions from a step-feed plug-flow full scale wastewater treatment plant. Sci. Rep. 6. https://doi.org/10.1038/srep20792

Pan, Y., Ye, L., Ni, B.-J., Yuan, Z., 2012. Effect of pH on N_2O reduction and accumulation during denitrification by methanol utilizing denitrifiers. Water Res. 46, 4832–4840. https://doi.org/10.1016/j.watres.2012.06.003

Platikanov, S., Rodriguez-Mozaz, S., Huerta, B., Barceló, D., Cros, J., Batle, M., Poch, G., Tauler, R., 2014. Chemometrics quality assessment of wastewater treatment plant effluents using physicochemical parameters and UV absorption measurements. J. Environ. Manage. 140, 33–44.

Pocquet, M., Wu, Z., Queinnec, I., Spérandio, M., 2016. A two pathway model for N_2O emissions by ammonium oxidizing bacteria supported by the NO/N_2O variation. Water Res. 88, 948–959. https://doi.org/10.1016/j.watres.2015.11.029

R Core Team, 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria [WWW Document]. URL https://www.R-project.org/

Ravishankara, A.R., Daniel, J.S., Portmann, R.W., 2009. Nitrous oxide (N_2O): the dominant ozone-depleting substance emitted in the 21st century. science 326, 123–125.

Rodriguez-Caballero, A., Aymerich, I., Poch, M., Pijuan, M., 2014. Evaluation of process conditions triggering emissions of green-house gases from a biological wastewater treatment system. Sci. Total Environ. 493, 384–391. https://doi.org/10.1016/j.scitotenv.2014.06.015

Rosén, C., Lennox, J.A., 2001. Multivariate and multiscale monitoring of wastewater treatment operation. Water Res. 35, 3402–3410.

Rosén, C., Olsson, G., 1998. Disturbance detection in wastewater treatment plants. Water Sci. Technol. 37, 197–205.

Rosen, C., Yuan, Z., 2001. Supervisory control of wastewater treatment plants by combining principal component analysis and fuzzy c-means clustering. Water Sci. Technol. 43, 147–156.

Rousseeuw, P.J., 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. Journal of computational and applied mathematics 20, 53–65.

Rustum, R., Adeloye, A.J., Scholz, M., 2008. Applying Kohonen Self-Organizing Map as a Software Sensor to Predict Biochemical Oxygen Demand. Water Environ. Res. 80, 32–40. https://doi.org/10.2175/106143007X184500

Schulthess, R.V., Gujer, W., 1996. Release of nitrous oxide (N_2O) from denitrifying activated sludge: Verification and application of a mathematical model. Water Res. 30, 521–530. https://doi.org/10.1016/0043-1354(95)00204-9

Schulthess, R. von, Wild, D., Gujer, W., 1994. Nitric and nitrous oxides from denitrifying activated sludge at low oxygen concentration. Water Sci. Technol. 30, 123–132.

Scott, A.J., Knott, M., 1974. A cluster analysis method for grouping means in the analysis of variance. Biometrics 507–512.

Spearman, C., 1904. “General Intelligence,” Objectively Determined and Measured. Am. J. Psychol. 15, 201–292. https://doi.org/10.2307/1412107

Sun, S., Cheng, X., Sun, D., 2013. Emission of N_2O from a full-scale sequencing batch reactor wastewater treatment plant: Characteristics and influencing factors. Int. Biodeterior. Biodegrad. 85, 545–549. https://doi.org/10.1016/j.ibiod.2013.03.034

Ward Jr, J.H., 1963. Hierarchical grouping to optimize an objective function. J. Am. Stat. Assoc. 58, 236–244.

Wunderlin, P., Mohn, J., Joss, A., Emmenegger, L., Siegrist, H., 2012. Mechanisms of N_2O production in biological wastewater treatment under nitrifying and denitrifying conditions. Water Res. 46, 1027–1037. https://doi.org/10.1016/j.watres.2011.11.080

Yang, X., Wang, S., Zhou, L., 2012. Effect of carbon source, C/N ratio, nitrate and dissolved oxygen concentration on nitrite and ammonium production from denitrification process by Pseudomonas stutzeri D6. Bioresour. Technol. 104, 65–72. https://doi.org/10.1016/j.biortech.2011.10.026

Yu, R., Kampschreur, M.J., Loosdrecht, M.C.M. van, Chandran, K., 2010. Mechanisms and Specific Directionality of Autotrophic Nitrous Oxide and Nitric Oxide Generation during Transient Anoxia. Environ. Sci. Technol. 44, 1313–1319. https://doi.org/10.1021/es902794a

Zheng, M., Tian, Y., Liu, T., Ma, T., Li, L., Li, C., Ahmad, M., Chen, Q., Ni, J., 2015. Minimization of nitrous oxide emission in a pilot-scale oxidation ditch: Generation, spatial variation and microbial interpretation. Bioresour. Technol. 179, 510–517.

Zhou, Y.A.N., Pijuan, M., Zeng, R.J., Yuan, Z., 2008. Free nitrous acid inhibition on nitrous oxide reduction by a denitrifying-enhanced biological phosphorus removal sludge. Environmental Science & Technology 42, 8260–8265.

Table 1:Average value and standard deviation (std) of variables monitored in the Northern carrousel reactor (C: carrousel reactor, N: Northern, PF: plug-flow reactor)

Online variables	Average	Std	Offline variables	Average	Std
N_2O (kg/h)	1.4	2.1	COD influent (mg COD/ L)	238.8	79.5
NH_4-N C (mg/L)	1.63	2.2	TKN influent (mg/L)	42.1	10.0
NO_3-N C (mg/L)	5.8	4	TP influent (mg/ L)	7.0	2.1
NO_2-N C (mg/L)	1.2	1.1	Flow-rate (m^3/ d)	85,898	41,786
DO1 (mg/L)	0.6	0.9	COD effluent (mg/ L)	36.9	6.9
DO2 (mg/L)	0.8	0.9	TKN efffluent (mg/ L)	2.8	1.2
DO3 (mg/L)	1.9	0.6	TP effluent (mg/ L)	1.1	0.6
Temperature (°C)	16	3.5	pH effluent	8.0	0.2
N_2O PF (kg/h)	0.71	1.21			
NH_4-N PF (mg/L)	12.41	5.35			
NO_3-N PF (mg/L)	2.38	2.2			
Influent Flow-rate (m^3/h)	3973	2375			
DO PF (mg/L)	2.61	0.65			

Table 2: Average values and standard deviations of the main variables for the 10 sub-periods (C: carrousel reactor, N: Northern, PF: plug-flow reactor).

Variables	N_2O (kg/h)		NO_3-C N (mg/l)		NO_3-N PF (mg/l)		NH_4-N C (mg/l)		NH_4-N PF (mg/l)		NO_2-N C* (mg/l)		Temperature (°C)		DO1 (mg/l)		DO2 (mg/l)		DO3 (mg/l)	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
1	0	0.1	6.1	3.1	1.8	1.6	1.8	2.67	11.4	4.1			15.7	1.4	0.62	0.7	0.62	0.5	1.5	0.4
2	0.6	0.6	7.2	3.1	2.5	2	1.5	1.7	13	4			11.2	1.0	0.77	1	1.31	0.8	2	0.4
3	2.7	1.4	6.1	3.2	1.6	2.1	1.6	2.1	15.2	4.5			11.5	0.7	0.67	0.8	1.49	1	2.07	0.4
4	5.6	2. 6	3	0.1	0.5	0.7	1.3	1.6	15	4.8	2.6	1.9	12.9	1.1	0.64	0.9	1.95	0.9	1.9	0.4
5	2.6	2.2	4.3	4.2	3.1	1.9	1.3	2	11.5	5.2	0.8	1	18.2	1.7	0.34	0.7	0.39	0.8	1.94	0.5
6	0.8	1.4	3.3	3.2	2.3	1.9	2	3.1	14.7	6.1	0.5	0.5	20	1.0	0.42	0.7	0.26	0.5	2.27	0.5
7	0.2	0.3	7.2	5	2.8	2.4	2	3.1	9.8	5.2	0.6	0.4	20	0.7	0.42	0.6	0.29	0.4	2.64	0.5
8	0.1	0.2	10.1	5.7	5.2	2.6	1.4	1	9.6	5.5	0.8	0.5	19.6	0.5	0.27	0.5	0.2	0.5	2.71	0.6
9	0.1	0.2	7.9	3.6	2.8	2.8	2	2	13.2	5.4	1.9	0.8	12.9	2.1	1.12	1.2	1.07	1	1.58	0.4
10	1.3	1.1	6.3	3.5	1.4	0.9	1.6	3.7	16.4	4.3	2.1	0.9	13	0.7	0.58	1.0	1.04	1	1.52	0.3

*NO_2-N concentration was monitored between 11/03/2011 and 19/01/2012

Table 3: Operating variables (average) for all clusters defined by hierarchical clustering in the carrousel reactor (P: Sub-period, Cl: Clusters)

P	Cl	N_2O C	NH_4-N PF	NO_3-N PF	Influent	NH_4-N C	NO_3-N C	DO1	DO2	DO3	NO_2-N
		kg/h	mg/l	mg/l	m^3/h	mg/l	mg/l	mg/l	mg/l	mg/l	mg/l
1	1	0.09	14.13	1.48	3883	1.47	8.66	1.04	0.78	1.72	
	2	0.01	8.55	2.41	3824	0.87	4.26	0.13	0.47	1.25	
	3	0.05	14.74	0.30	8892	7.91	4.63	1.37	0.77	1.58	
2	4	0.87	15.30	2.05	3827	1.51	8.61	0.94	1.53	2.22	
	5	0.21	9.13	3.69	3419	0.74	5.28	0.03	0.62	1.41	
	6	0.24	12.51	0.81	11132	4.52	5.42	2.27	2.31	2.22	
3	7	3.22	16.85	1.52	3383	1.36	7.36	0.87	1.88	2.35	
	8	1.72	10.96	1.91	3672	0.82	4.29	0.05	0.85	1.56	
	9	2.40	21.40	0.12	7935	7.52	4.15	2.10	1.28	2.10	
4	10	6.60	17.30	0.32	3207	1.26	3.79	2.14	0.95	2.41	4.10
	11	3.83	10.82	0.77	2747	0.79	1.80	1.51	0.05	1.20	1.40
	12	6.89	25.45	0.48	6375	10.86	3.62	1.98	2.12	2.34	4.28
6	15	2.54	17.66	0.75	5922	5.00	5.07	1.30	0.73	2.34	1.08
	16	0.51	8.20	2.84	3811	0.98	2.64	0.10	0.10	2.21	0.35

*NO_2-N concentration was monitored between 11/03/2011 and 19/01/2012

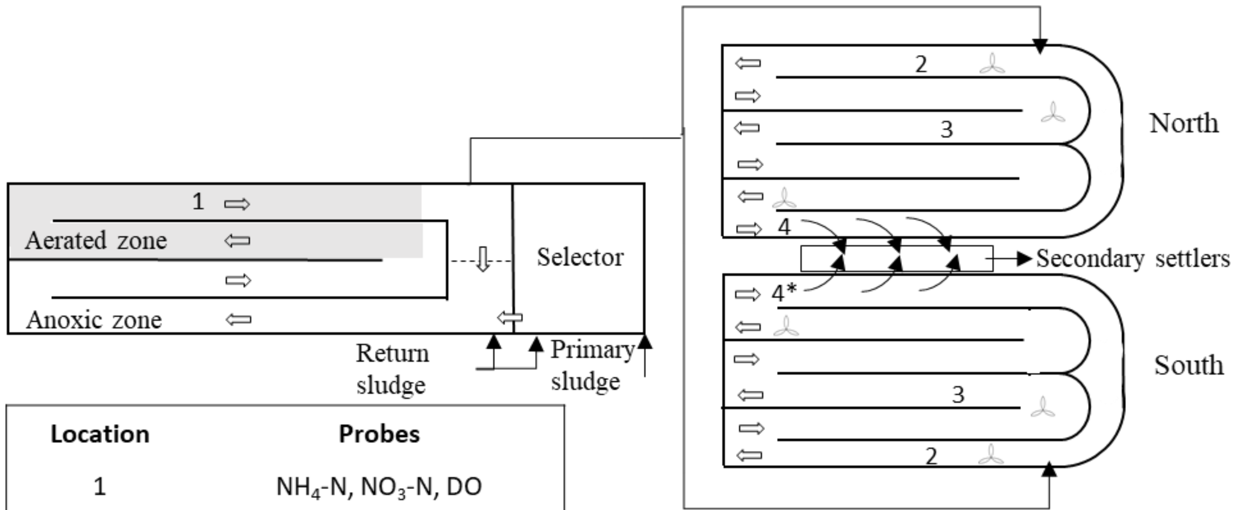
Table 4: PCA loadings sub-period 2, carrousel reactor

Variable	PC1	PC2	PC3	PC4
NH_4-N PF	-0.28	0.47	-0.24	0.29
NO_3-N PF	0.36	0.21	0.14	-0.67
Influent	-0.38	-0.31	-0.09	-0.37
NH_4-N C	-0.34	0.03	-0.59	-0.29
NO_3-N C	-0.04	0.58	0.21	-0.31
DO1	-0.43	0.06	-0.15	-0.18
DO2	-0.40	0.08	0.48	-0.17
DO3	-0.37	0.21	0.40	0.28
Temperature	0.22	0.49	-0.33	0.11

Table 5: PCA loadings sub-period 4, carrousel reactor

	PC1	PC2	PC3	PC4
NH_4-N PF	-0.48	0.04	-0.11	0.25
NO_3-N PF	0.26	0.56	-0.04	-0.35
Influent	-0.33	-0.07	-0.52	-0.17
NH_4-N C	-0.28	0.14	-0.50	-0.46
NO_3-N C	-0.17	0.59	0.32	0.04
DO1	-0.37	0.24	-0.13	0.59
DO2	-0.40	0.08	0.41	-0.14
DO3	-0.37	0.01	0.33	-0.40
Temperature	0.23	0.51	-0.27	0.19

Figure



Location	Probes
1	NH_4-N, NO_3-N, DO
2	DO1, TSS
3	DO2, Temperature
4	DO3, NH_4-N, NO_3-N, NO_2-N
4*	DO3, NH_4-N, NO_3-N

Figure 1: Layout of Kralingseveer WWTP with Plug-flow and Carrousel reactors, adapted from Daelman et al., (2015).

Figure

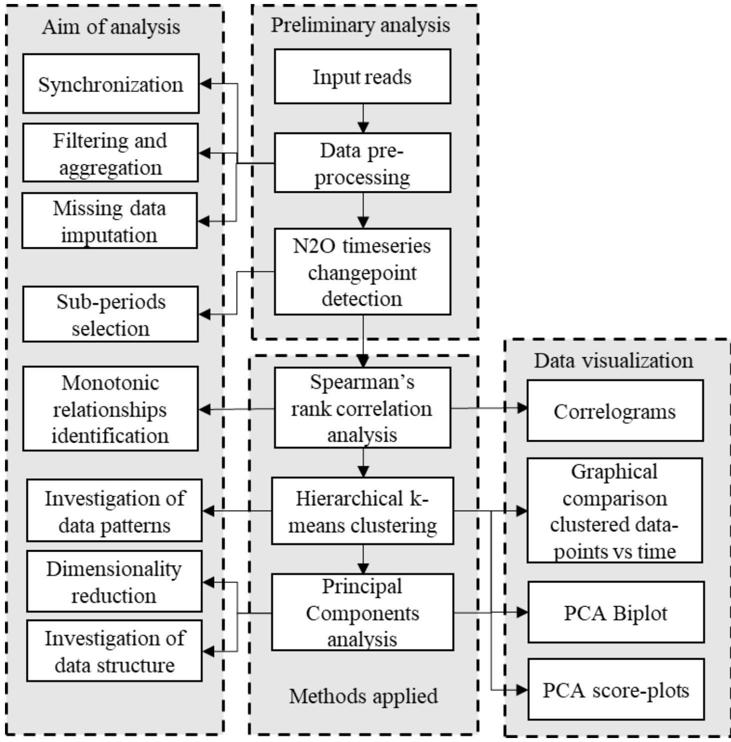


Figure 1: Methodology followed in the current study for data processing and visualization

Figure

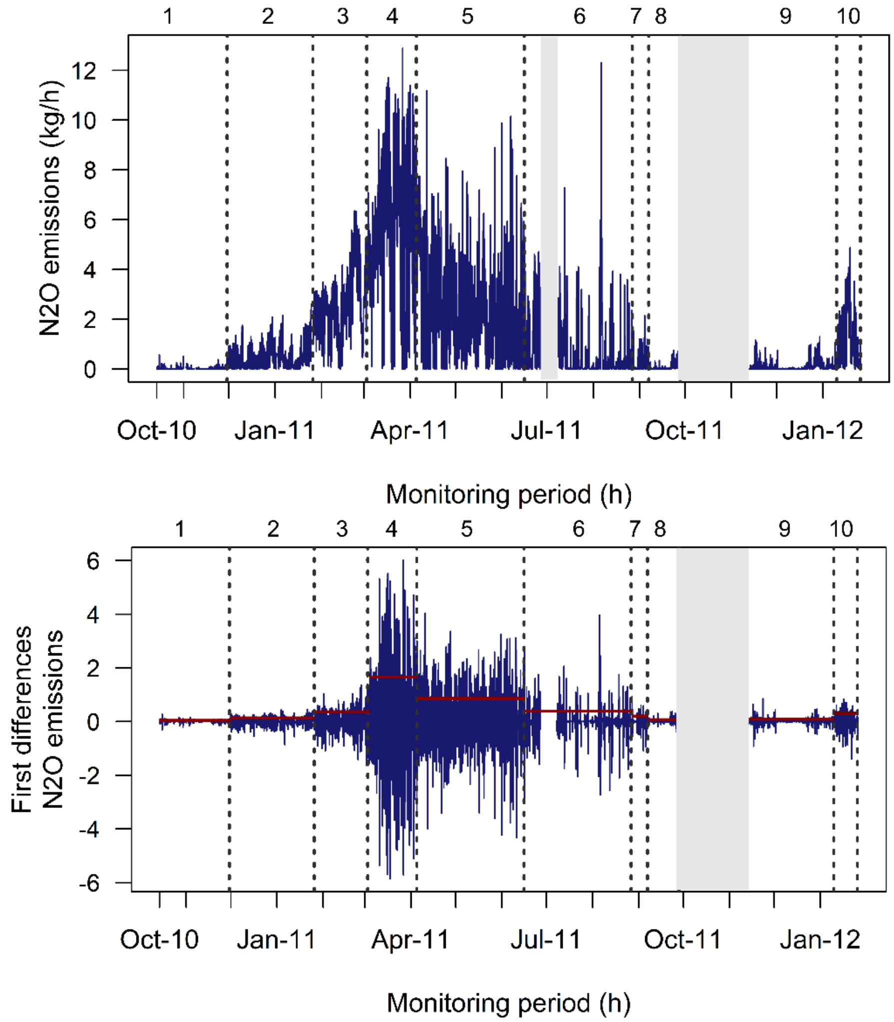


Figure 3 (top): N2O emissions profile in the Northern Carrousel reactor (grey area: periods with missing N2O data) (bottom): First difference of the N2O emissions timeseries (blue line) showing the sub-periods identified by the application of binary segmentation (grey area: periods with missing N2O data, blue dotted lines: changepoints identified by the algorithm, red horizontal lines: standard deviation in each sub-period)

Figure

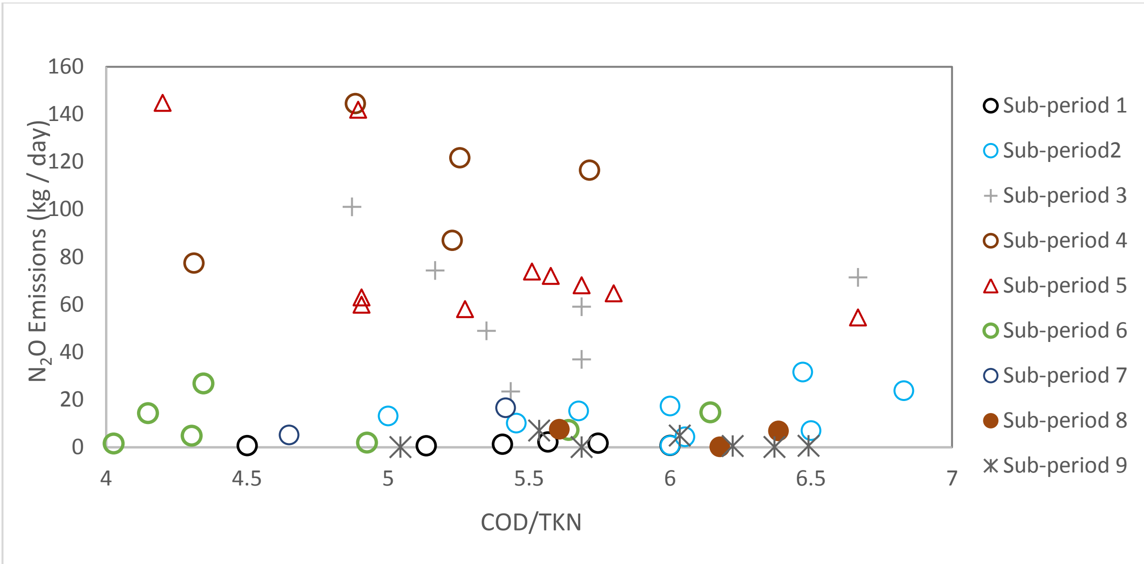


Figure 4: COD/TKN (offline data) for each sub-period

Figure

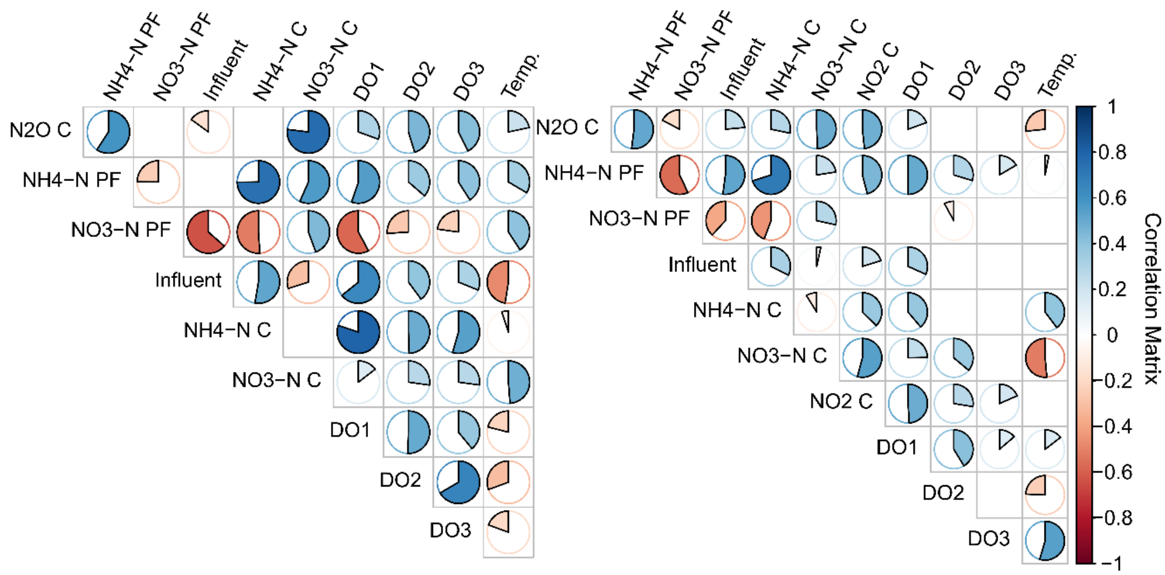


Figure 5: Spearman's rank correlation coefficient for sensor signals in Northern Carrousel reactor. (Left): Sub-period 2. (Right): Sub-period 5. (Red: negative correlation, blue: positive correlation, the coloured part of the circles is proportional to the correlation coefficient, only results with p-value < 0.01 are shown)

Figure

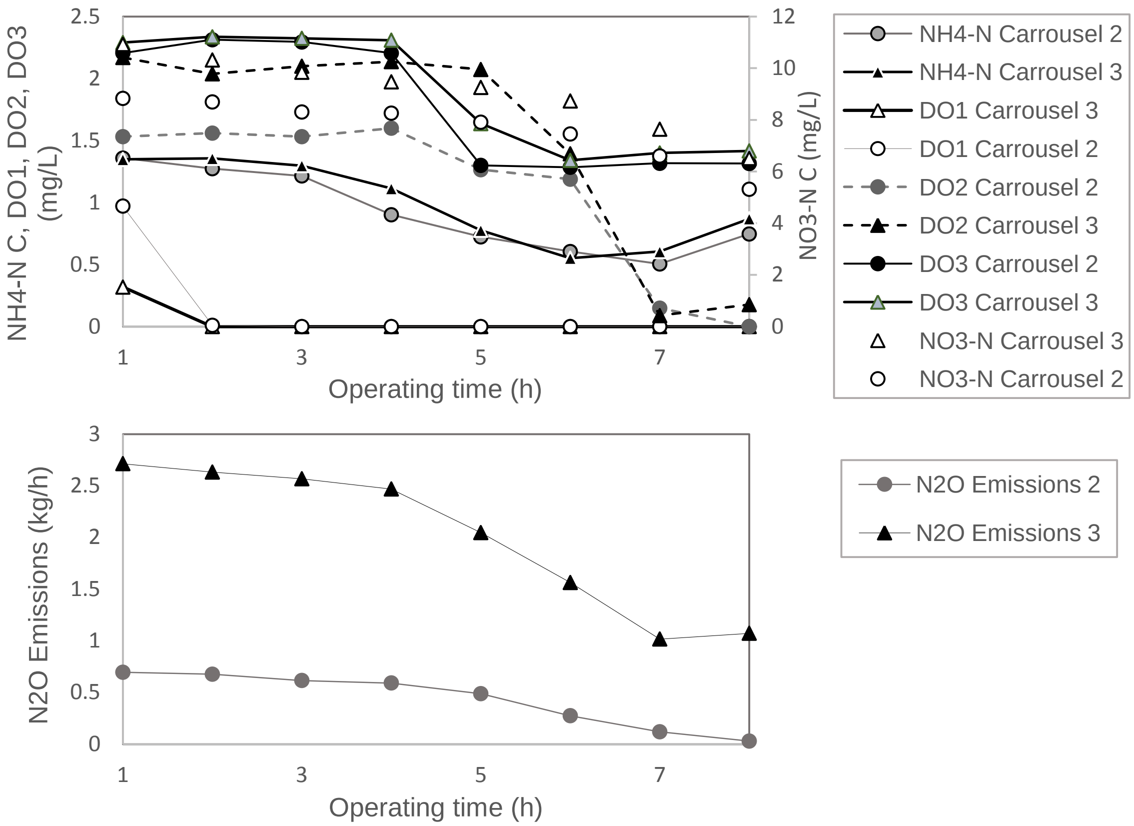


Figure 6: (Top): Variables monitored online for two separate occasions in sub-periods 2 and 3 (from 00:00 am until 8:00 am), (Bottom): The respective N2O emissions profiles

Figure

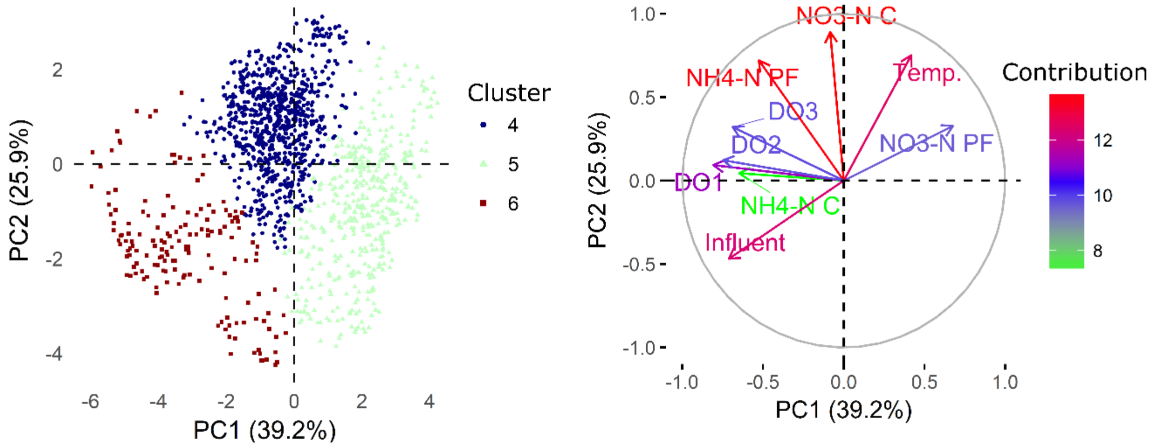


Figure 7: (left) Biplot of the first 2 PC scores, sub-period 2. The colored data-points represent the scores of the first two principal components. Groups 4, 5, and 6 represent sub-period 2, clusters. (right) Variable correlation plot. The arrows represent the direction and strength (variable coordinates = loading x component std) of the variables monitored in the system as projected into the 2-d plane. The contrib. legend represents the contribution (%) of the variables to the first two PCs. The arrows for each variable point to the direction of increase for that variable. The percentage given on each axis label represents the value of the total variance explained by that PC.

Figure

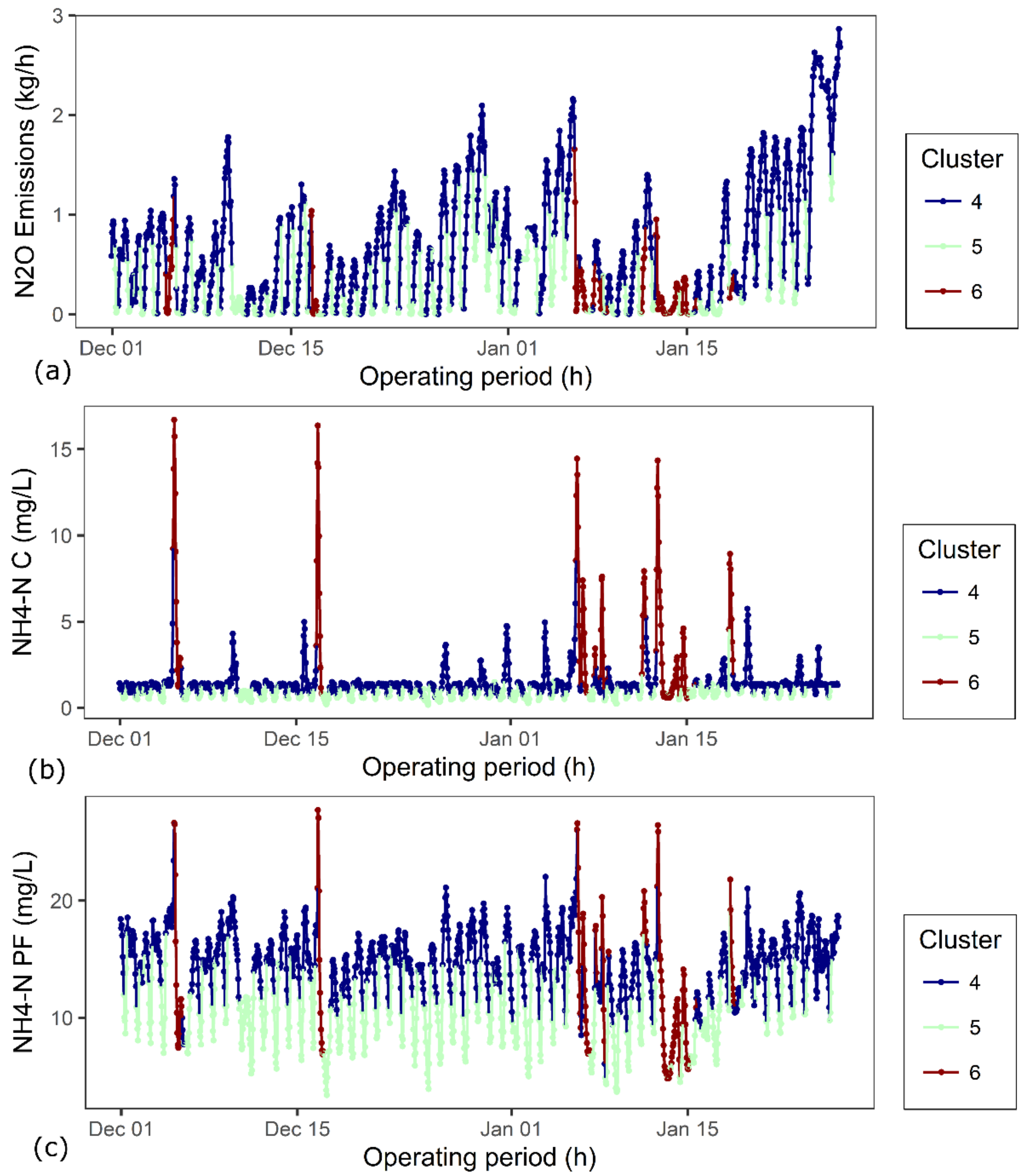


Figure 8: Profile of (a) N_2O emissions, (b) NH_4-N concentration in the Carrousel reactor and (c) NH_4-N concentration in the plug-flow reactor for sub-period 2; coloured points indicate the respective clusters

Figure

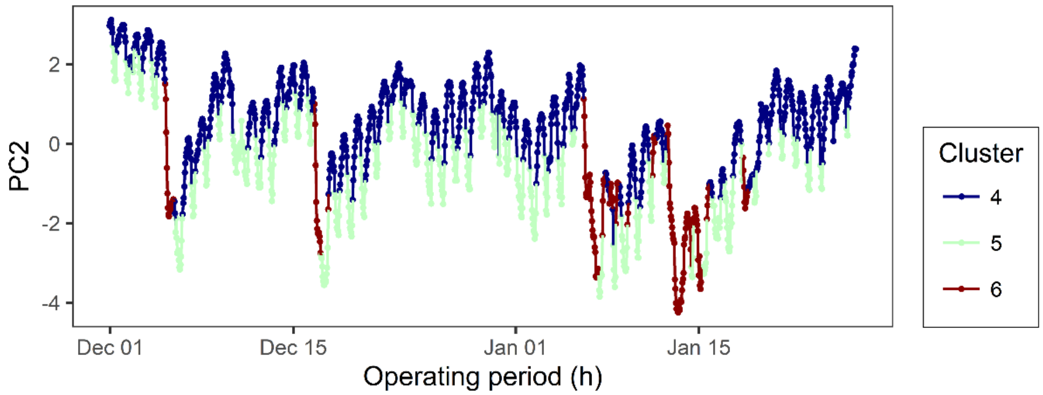


Figure 9: PC2 scores for sub-period 2

Figure

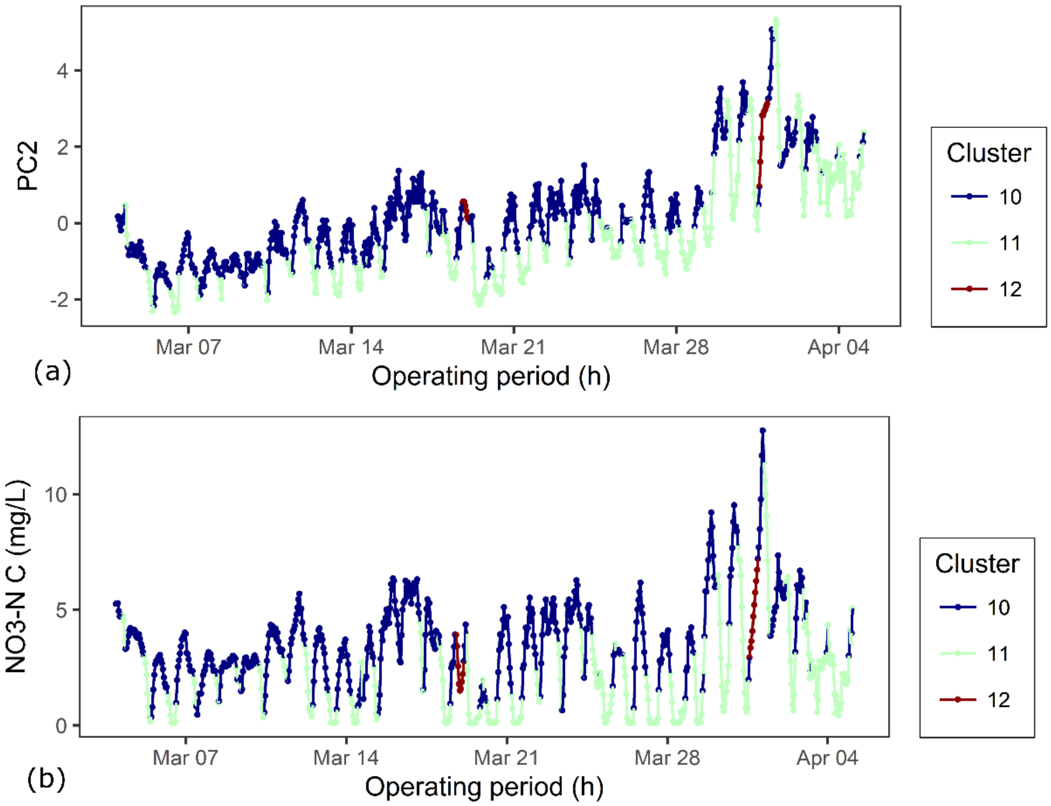
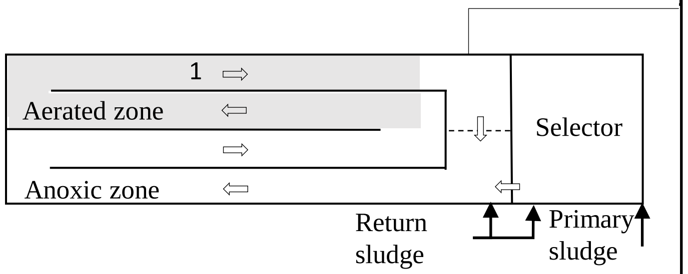
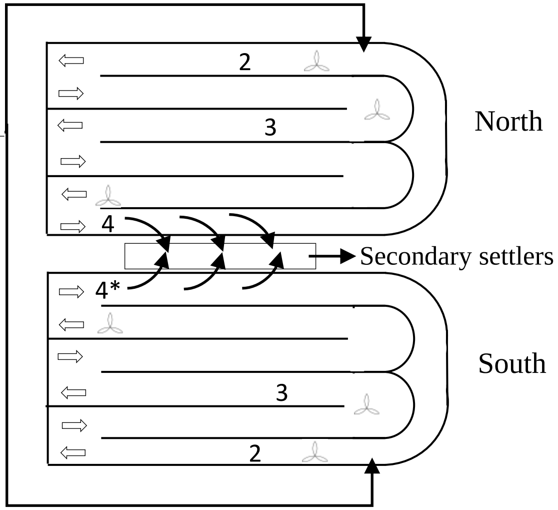


Figure 10: (a) PC2 scores for sub-period 4 and (b) NO_3-N concentration in the Carrousel reactor for sub-period 4.

Figure





Location	Probes
1	NH_4-N, NO_3-N, DO
2	DO1, TSS
3	DO2, Temperature
4	DO3, NH_4-N, NO_3-N, NO_2-N
4*	DO3, NH_4-N, NO_3-N

Figure

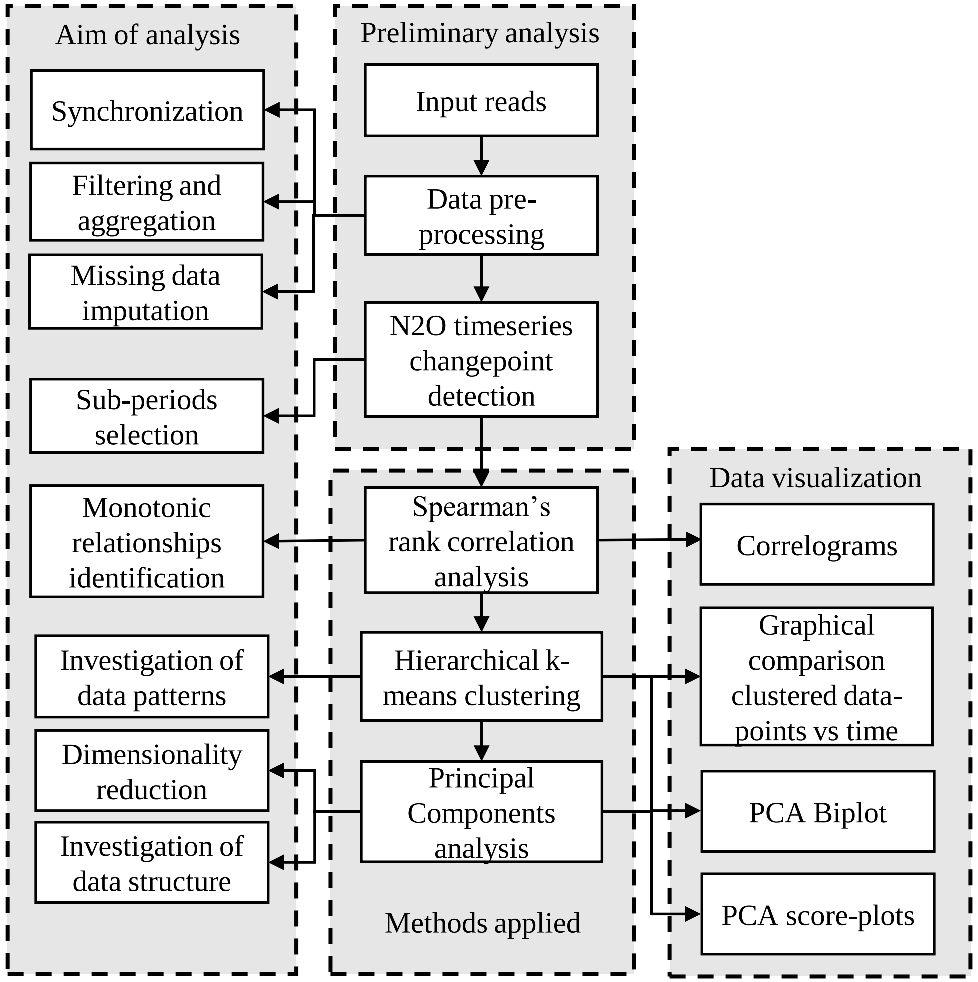
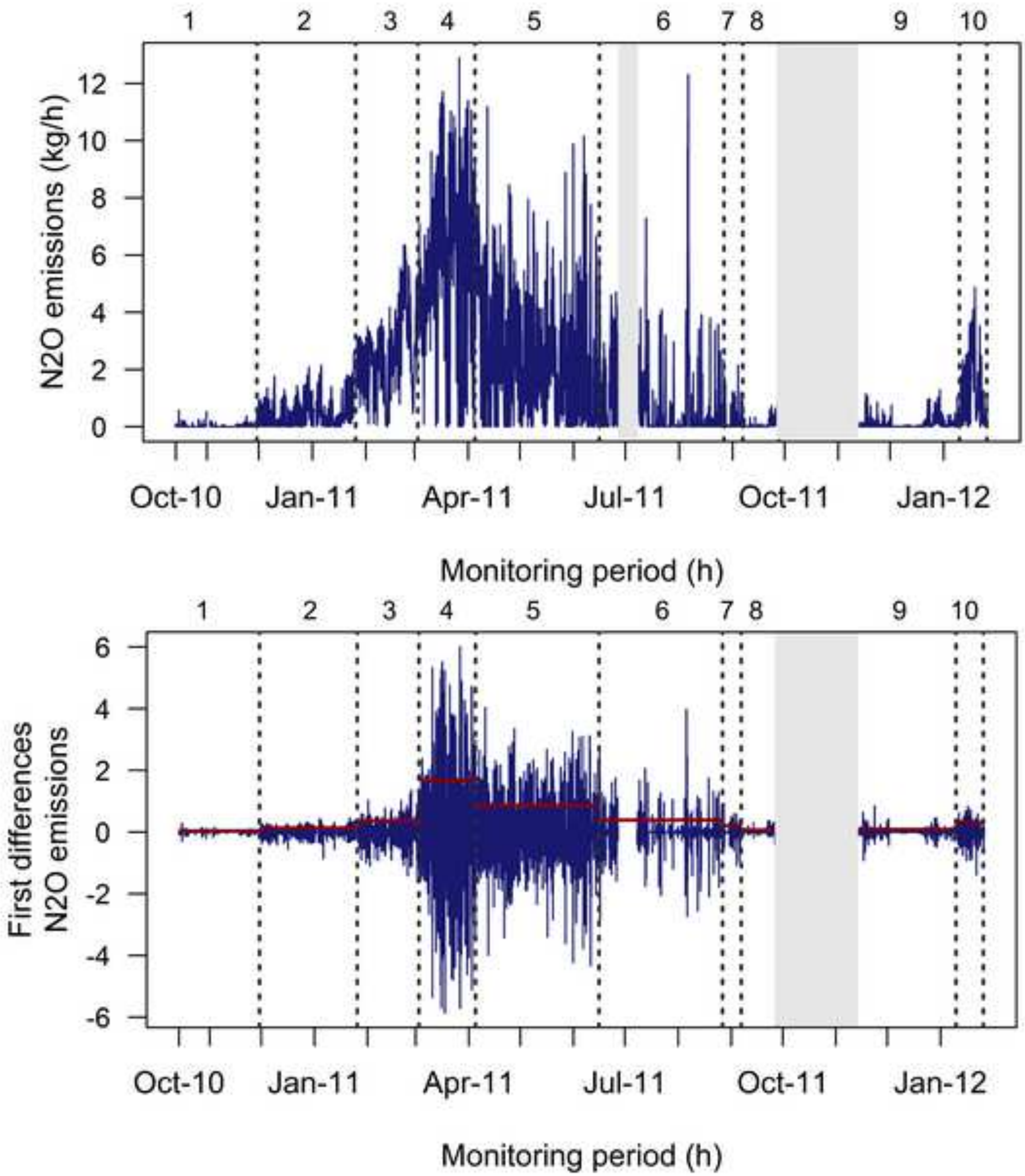
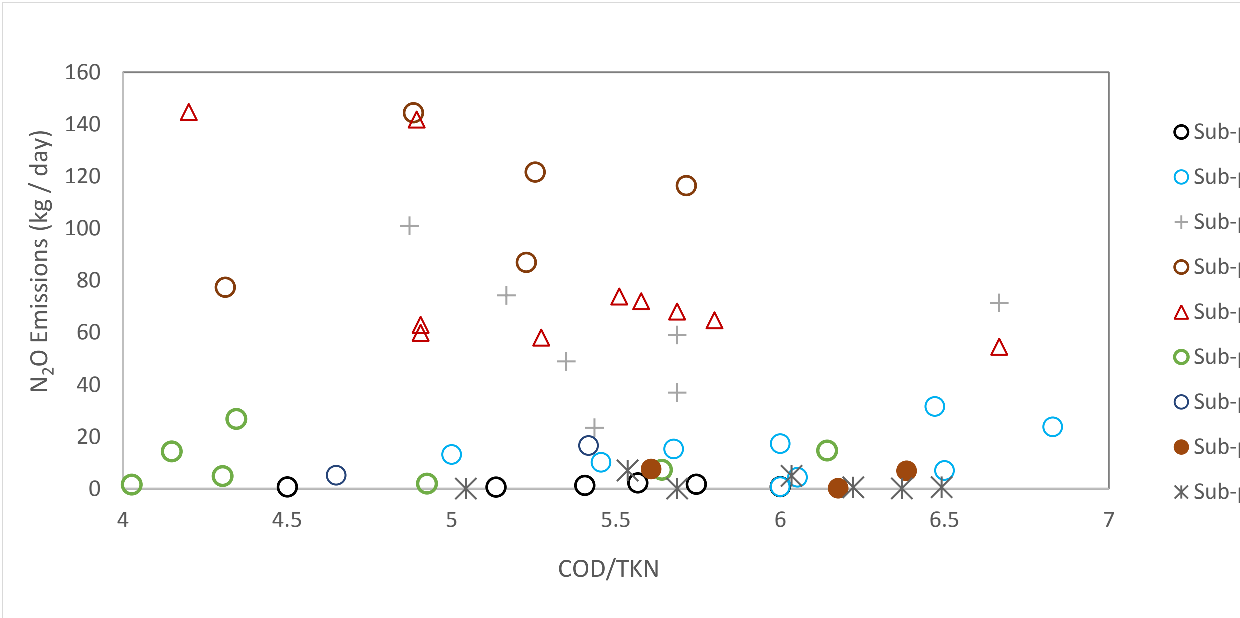

Aim of analysis
Synchronization
Filtering and aggregation
Missing data imputation
Sub-periods selection
Monotonic relationships identification
Investigation of data patterns
Dimensionality reduction
Investigation of data structure
Preliminary analysis
Input reads
Data pre-processing
N2O timeseries changepoint detection
Spearman's rank correlation analysis
Hierarchical k-means clustering
Principal Components analysis
Methods applied
Data visualization
Correlograms
Graphical comparison clustered data-points vs time
PCA Biplot
PCA score-plots

Figure
Click here to download high resolution image



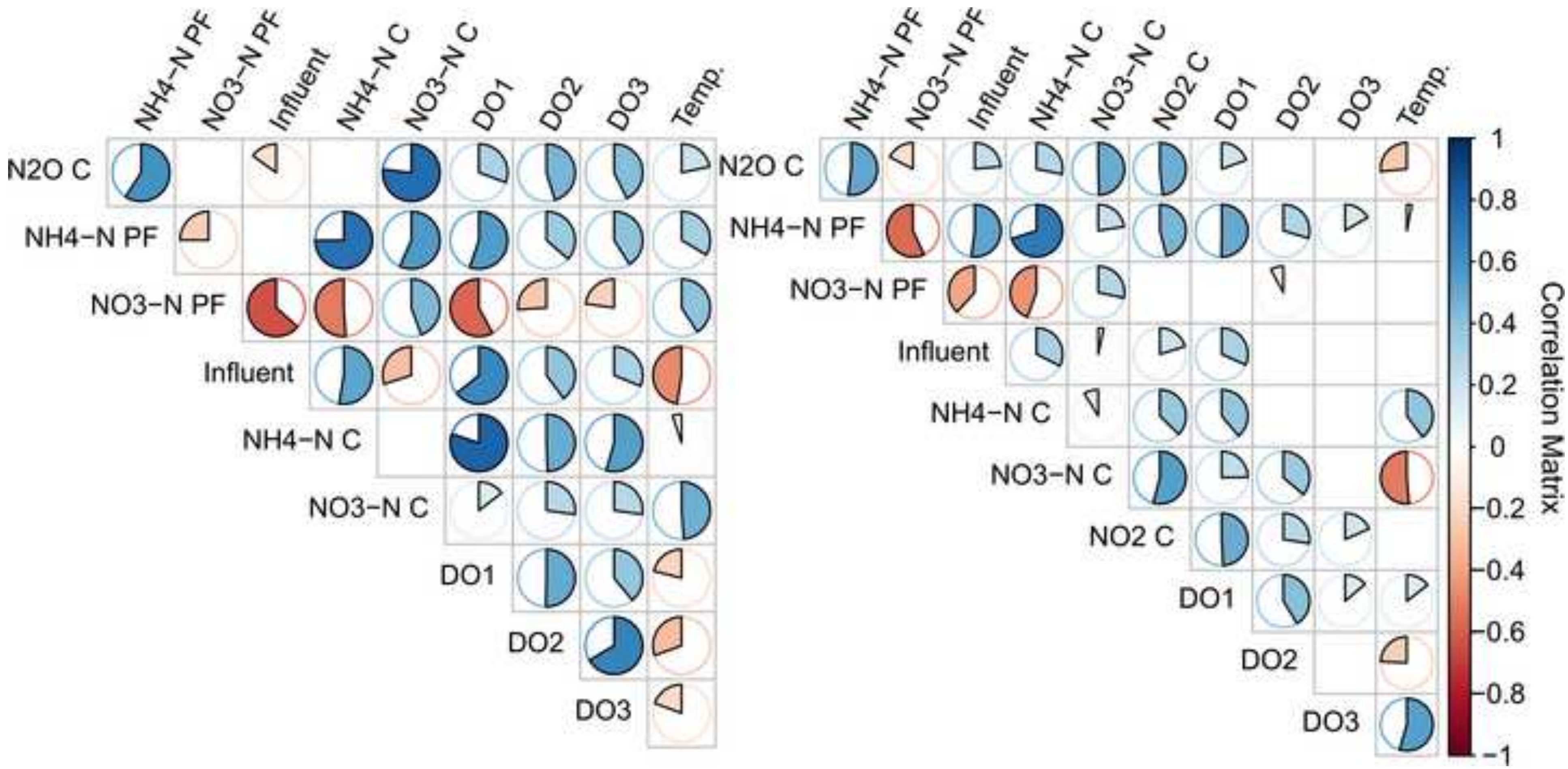
Figure



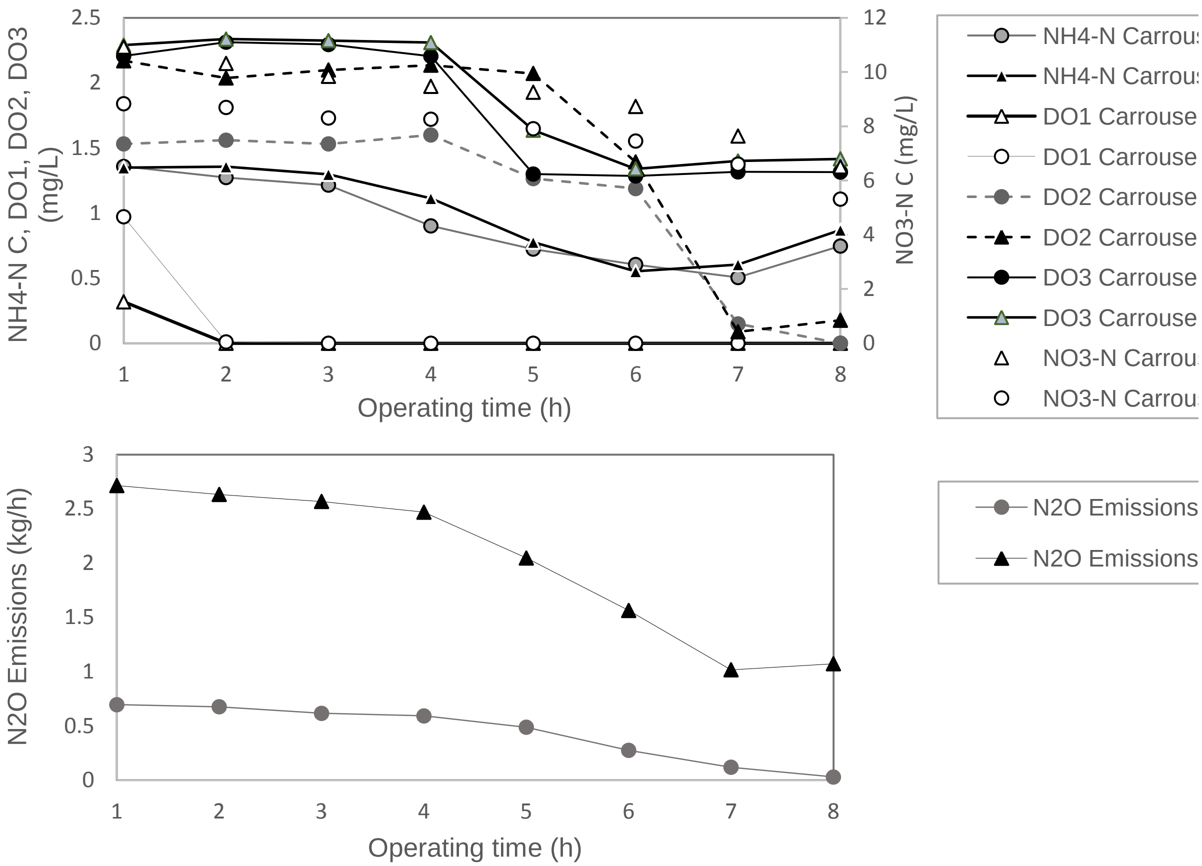
N2O Emissions (kg / day)
160
140
120
100
80
60
40
20
0
4
4.5
5
5.5
6
6.5
7
COD/TKN
Sub-p
Sub-p
Sub-p
Sub-p
Sub-p
Sub-p
Sub-p
Sub-p
Sub-p

Figure

Click here to download high resolution image

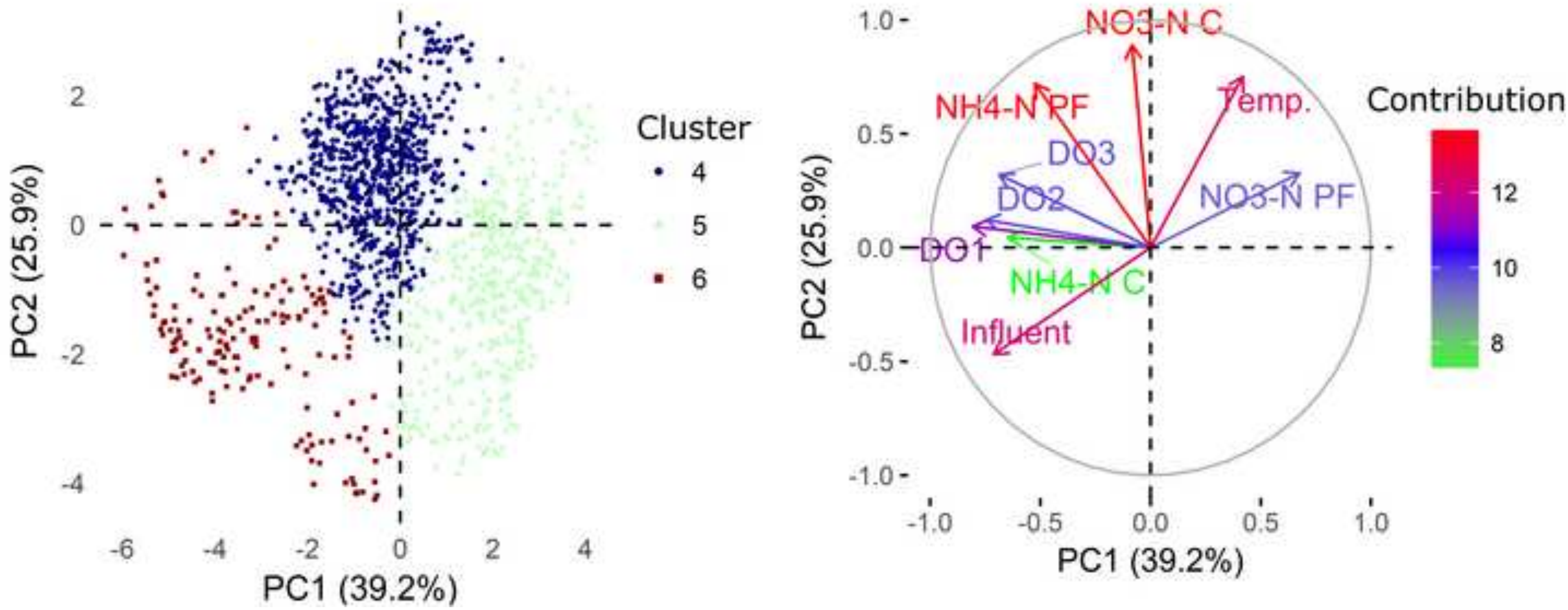


Figure

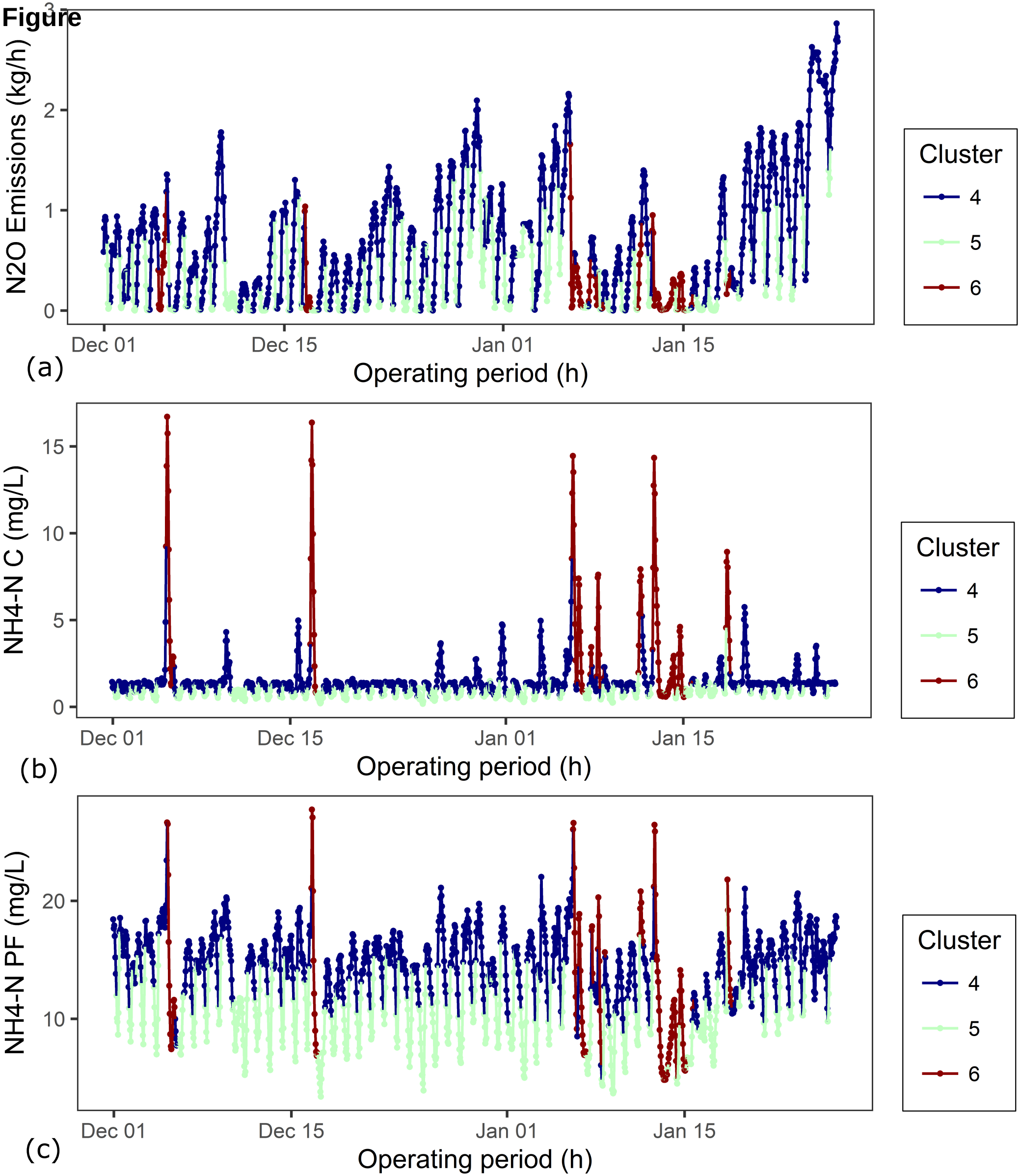


NH4-N C, DO1, DO2, DO3 (mg/L)
NO3-N C (mg/L)
Operating time (h)
NH4-N Carrous
NH4-N Carrous
DO1 Carrouse
DO1 Carrouse
DO2 Carrouse
DO2 Carrouse
DO3 Carrouse
DO3 Carrouse
NO3-N Carrou
NO3-N Carrou
N2O Emissions (kg/h)
Operating time (h)
N2O Emissions
N2O Emissions

Figure
Click here to download high resolution image



Figure



N2O Emissions (kg/h)
0
1
2
3
Dec 01
Dec 15
Jan 01
Jan 15
(a)
Operating period (h)
Cluster
4
5
6
NH4-N C (mg/L)
0
5
10
15
Dec 01
Dec 15
Jan 01
Jan 15
(b)
Operating period (h)
Cluster
4
5
6
NH4-N PF (mg/L)
10
20
Dec 01
Dec 15
Jan 01
Jan 15
(c)
Operating period (h)
Cluster
4
5
6

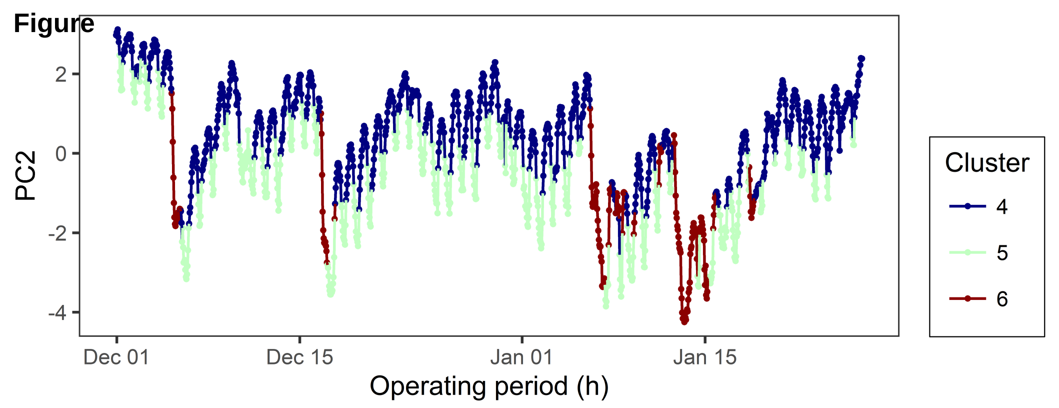
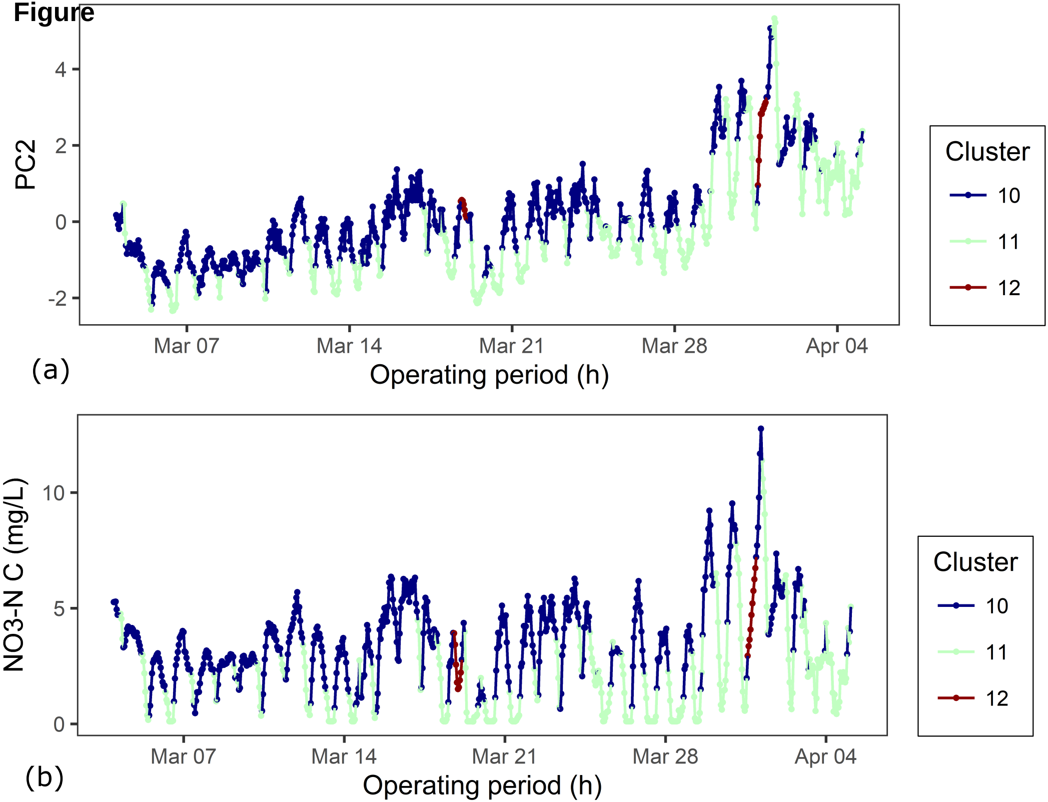


Figure
PC2
2
0
-2
-4
Dec 01
Dec 15
Jan 01
Jan 15
Operating period (h)
Cluster
4
5
6

Figure


PC2
4
2
0
-2
Mar 07
Mar 14
Mar 21
Mar 28
Apr 04
(a)
Operating period (h)
Cluster
10
11
12
NO3-N C (mg/L)
10
5
0
Mar 07
Mar 14
Mar 21
Mar 28
Apr 04
(b)
Operating period (h)
Cluster
10
11
12