Knowledge discovery and data mining to understand and optimise the environmental behavior of wastewater treatment processes

A thesis submitted for the degree of Doctor of Philosophy

by

Vasileia K. Vasilaki

Department of Civil and Environmental Engineering

College of Engineering Design and Physical Sciences

Brunel University London

Uxbridge, UB8 3PH, United Kingdom

July 2020

Acknowledgements

The completion of this thesis would not have been possible without the direct and indirect contribution and support of many people, who had an incremental role in my research journey.

Foremost, I would like to express my sincere gratitude to my supervisor Dr. Evina Katsou. Her enthusiasm, intellectual support, work ethics and confidence in my work have been the steppingstone for the completion of this work and my continuous academic development for the past years. She has always provided her insightful advice and was always able to guide me towards the right direction. She is a great scientist and mentor and source of inspiration for my future steps.

I would also like to deeply thank my second supervisor Dr Ali Mousavi for his invaluable contributions and guidance in the world of data-mining and process control and automation. His constant availability and stimulating discussions were a catalyst for this work and for this I am deeply grateful.

I am truly indebted to Dr Nicola Frison, Vinzenco Conca and all the team members from Verona who were always happy to share their data, knowledge and insights of the SCENA operation. Without their support the results would be much poorer.

I am deeply thankful towards Professor Eveline Volcke for her fruitful advice, for planning the visit at Kralingseveer WWTP and for her warm welcome and hospitality in Ghent. I would also like to thank Professor van Loodstrecht for supporting my research. I would also like to thank 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

The regular SMART-Plant meetings provided great opportunities to discuss new ideas and to get inspiration combining the different viewpoints of the group members. A special thanks to the co-ordinator, Dr Francesco Fatone.

Next, I would like to thank my friends and colleagues from Brunel University Peyo, Theoni, Eliza and YongHui who created a fruitful atmosphere and provided their advice and support.

The presence of my friends was decisive, especially in tough times that inevitably arose. I would like to thank them for the support and advice we exchanged all this time. A special

thanks to Ilias for his love and patience and for always cheering me up and giving me strength.

There are no words to describe my gratitude towards my family, especially my mother and my sister. Their unconditional support and optimism have made it possible for me to reach this point today.

Vasileia K Vasilaki

Brunel University London

Uxbridge, United Kingdom

November 2019

Abstract

Direct nitrous oxide (N_2O) emissions during the biological nitrogen removal (BNR) processes can significantly increase the carbon footprint of wastewater treatment plant (WWTP) operations. However, quantifying the emissions and understanding the long-term behaviour of N_2O fluxes in WWTPs remains challenging and costly. The aim of the current research is to combine wastewater domain knowledge with data-mining techniques to explain the long-term N_2O emissions' behaviour in full-scale biological reactors. A review of the recent full-scale N_2O monitoring campaigns is conducted resulting in the development of an emission factor (EF) database with information on configurations, control strategies and operational conditions. The analysis focused on mechanistic model development, molecular biology methods and on the current data management and analysis practices (i.e. visualization techniques, statistical analysis). Sensor and laboratory data acquired from the N_2O monitoring campaigns of mainstream and sidestream wastewater processes were used to develop, test and validate a methodological framework for knowledge discovery in wastewater databases. Abnormal events detection, structural changepoint detection, clustering, classification and regression algorithms are used in order to i) translate data into actionable information, ii) link N_2O emissions ranges with specific operational conditions, iii) identify and isolate re-occurring system disturbances that affect performance, iv) predict the range of N_2O emissions based on operational and environmental conditions and v) provide feedback to monitoring campaigns for the minimisation of sampling requirements. The analysis showed that the relationship of N_2O emissions with the operational variables fluctuates in long-term monitoring campaigns; this should be taken into consideration for the development of mitigation measures and during the investigation of triggering operational conditions. Additionally, findings indicate that structural changepoints of operational variables monitored online can be used to detect changes in the behaviour and range of N_2O emissions. Finally, data-driven models can reliably estimate N_2O behaviour in wastewater processes under given operational conditions. However, fluctuation of dependencies, system disturbances and process-specific characteristics should be taken into consideration.

Table of contents

List of Figures

List of Tables

List of Abbreviations

A/O	Anaerobic-Oxic
A^2/O	Anaerobic-Anoxic-Oxic
A_h	Area of Gas Hood
amoA	Ammonia monooxygenase
Anammox	Anoxic Ammonia Oxidation
ANN	Artificial Neural Networks
AOA	Ammonia Oxidizing Achaea
AOB	Ammonia Oxidizing Bacteria
AOR	Ammonia Oxidation Rate
A_r	Area of Biological Reactor
AS	Activated Sludge
ASM	Activated Sludge Models
BNR	Biological Nutrient Removal
BSM2	Benchmark Simulation Model 2
CAS	Conventional Activated Sludge
CH_4	Methane
CO_2	Carbon dioxide
COD	Chemical Oxygen Demand
CP	Changepoint
CRC	Carbon Reduction Commitment
DO	Dissolved Oxygen
ECD	Electron-Capture Detector
EF	Emission Factor
FTIR	Fourier Transform Infrared
GC	Gas Chromatography
GHG	Greenhouse Gases
HAO	Hydroxylamine Oxidoreductase
HDPE	High-Density Polyethylene
HRT	Hydraulic Retention Time
ICA	Independent Component Analysis
ICT	Information and Communication Technology
IPCC	Intergovernmental Panel on Climate Change
IWA	International Water Association
LCA	Life cycle assessment
MDS	Multidimensional Scaling
MLE	Modified Ludzack-Ettinger
MLSS	Mixed liquor volatile suspended solids
MLVSS	Mixed Liquor Volatile Suspended Solids
N	Nitrogen
N_2O	Nitrous oxide
N_2OR	Nitrous Oxide Reductase
NAS	New Activated Sludge

NH_2OH	Hydroxylamine
NH_3	Ammonia
NH_4^+-N	Ammonium Nitrogen
NO	Nitric Oxide
NO_2-N	Nitrite nitrogen
NO_3-N	Nitrate nitrogen
NOB	Nitrite Oxidizing Bacteria
NOR	Nitric Oxide Reductase
NUR	Nitrate Utilisation Rate
O_2	Oxygen
OD	Oxidation Ditch
ORP	Oxidation Reduction Potential
PC	Principal component
PCA	Principal component analysis
PE	Population Equivalent
PF	Plug Flow
PLC	Programmable Logic Controller
PLS	Partial least squares
PP	Polypropylene
PVC	Polyvinyl Chloride
qPCR	Real-time Polymerase Chain Reaction
r-A^2/O	Reversed A^2/O
RAS	Return Activated Sludge
Redox	Reduction-Oxidation
RT-qPCR	Reverse Transcription Polymerase Chain Reaction
SCENA	Short Cut Enhanced Nutrient Abatement Process
SBR	Sequence Batch Reactor
SEIFC	Surface Emission Isolation Flux Chamber
SND	Simultaneous Nitrification Denitrification
SP	Site Preference
SRT	Sludge Retention Time
SVM	Support Vector Machine
TKN	Total Kjeldahl Nitrogen
TN	Total Nitrogen
TSS	Total Suspended Solids
UK CCA	UK Climate Change Act
US EPA	United States Environmental Protection Agency
UWWTD	Urban Waste Water Treatment Directive
VFA	Volatile Fatty Acids
WaCCliM	Water and Wastewater Companies for Climate Mitigation project
WEF	Water Environmental Federation
WFD	Water Framework Directive
WMO-GAW	World Meteorological Organisation – Global Atmosphere Watch
WWTP	Wastewater Treatment Plant

1. Introduction

1.1 Research Motivation

Wastewater treatment is a growing segment of the water industry. It has been subjected to decades of continuous tightening of water quality regulations (Wang et al., 2015) increasing the demands from the sector to continuously adapt to new challenges and environmental targets. The introduction of the Urban Waste Water Treatment Directive (91/271/EEC) (UWWTD) and later the Water Framework Directive 2000/60/EC (WFD) have been the cornerstones instigating changes in the wastewater treatment practices across the EU. The directives aim to protect the European waters, and require the introduction of limits for the discharge of municipal and industrial wastewater contaminants, minimum treatment type, target actions against unsatisfactory intermittent discharges, while setting the framework for a holistic catchment management approach (Tippett, 2005). Specific deadlines were also defined urging member states to take action; the timeline of the agenda towards achieving the environmental targets and towards 'good status' for EU aquatic systems established by the WFD ended at 2015 with the possibility of postponement up to 2027 (two six-year phases) (European Commission, 2012a). Significant investments of the wastewater utilities have been made in the past years for the protection of the water bodies. For example, the UK water industry alone, reported in 2013, that the investment between 2010 and 2030 for capacity upgrade and conformity with the WFD is expected to be £27 billion (Severn Trent Plc, 2013). In Italy, municipal wastewater is the second most significant contributor to surface waters' pressures after agriculture, whereas compliance with the UWTTD is still incomplete; less than 72% of the collected wastewater is subjected to secondary treatment (COM, 2017; SWD, 2017).

Wastewater utilities are estimated to contribute by ~1% to 2% of the energy requirements in a country level (Haslinger et al., 2016; IEA, 2016). The US Environmental Protection Agency reported that the energy requirements for wastewater treatment in the US will increase by 20% in the following years; similar to what is expected in other developed countries (Curtis, 2010; Wang et al., 2012; EPA, 2006). The increase of the energy consumption in the form of electricity from the grid results in elevated Greenhouse Gas (GHG) emissions (Flores-Alsina et al., 2011; Maktabifard et al., 2018) putting more climate change pressures. In the UK wastewater utilities, the compliance with the WFD requirements is expected to result in

increase of the carbon dioxide (CO2) emissions by more than 110,000 tonnes annually only due to increase in operational energy requirements (Environment Agency, 2010). According to the Energy Efficiency Directive 2012/27/EU large water utilities (>250 employees, income exceeding €43 million or yearly trading volume >€50 million) are obliged to perform energy audits at their wastewater treatment facilities. In the UK, policy schemes, such as the Climate Change Act (UK CCA) and the Carbon Reduction Commitment (CRC) aiming at GHG emission reduction (targets >80% by 2050 with 1990 baseline) and energy efficiency, increase the pressure for water utilities. Severn Trent, a water utility in UK, paid £5.9 million in 2012, to the CRC due to carbon emissions from energy consumption in addition to the yearly electricity bill that was equal to £53 million (Severn Trent Plc, 2012). The water utilities have made efforts in order to cope with the increasing costs and climate change pressures.

At the same time current attitude in the wastewater sector is shifting towards a philosophy of considering wastewater as a resource and recovering materials and energy (Puchongkawarin et al., 2015). In the last few years increasing number of studies are focusing on the WWTPs and new technologies are developed with the ability to recover valuable resources, such as chemicals, nutrients, bioplastics, enzymes, metals and water (Batstone et al., 2015; Frison et al., 2015; Puyol et al., 2017; Sharma et al., 2014; Wang and Ren, 2014). According to Cordell et al. (2009), 20% of the phosphorous consumed is contained in wastewater, whereas the nitrogen loads in wastewater are equal to 10%-30% of nitrogen required in agriculture (Mulder, 2003). Therefore, the circular economy concept in the water sector is driven by economic, environmental and industrial incentives (Puyol et al., 2017).

Energy efficiency improvement and resource recovery are the two dominant objectives that are presently pursued in WWTPs to improve their sustainability (Mo and Zhang, 2013). However, research studies have shown that electricity consumption might not be the most significant contributor to the operational carbon footprint in the wastewater sector. Nitrous oxide (N_2O), is a potent GHG, 265 times stronger than CO_2 in terms of global warming potential (IPCC, 2013). N_2O can be generated at large amounts and stripped in the atmosphere during biological nutrients removal at WWTPs. N_2O can be formed in activated sludge systems, during the autotrophic oxidation of ammonia to nitrite/nitrate through the activity of ammonia oxidizing bacteria (AOB) under aerobic conditions (nitrification/nitritation). It is also an intermediate during the reduction of nitrate/nitrite to nitrogen gas (heterotrophic denitrification/denitritation) through the activity of heterotrophic denitrifying bacteria under

anoxic conditions. There is a wide variety of different biological nutrients removal (BNR) processes applied at wastewater facilities to treat the incoming wastewater (i.e. with different number of compartments/zones for nitrification and denitrification, recirculation flows, flow-patterns, continuous processes vs batch processes etc.). Key operational conditions in the BNR processes associated with elevated emissions include: i) low dissolved DO, NO_2^- or free nitrous acid (HNO_2) accumulation and changes in the NH_4^+ concentration in the nitrifying zones, ii) limitation of organic substrate (i.e. low chemical oxygen demand to N (COD:N) ratio), as well as, NO_2^- accumulation in the denitrifying zones, iii) alternation of anoxic/aerobic conditions and iv) abrupt changes in the processes and system shocks (Duan et al., 2017; Guo et al., 2017; Law et al., 2012; Massara et al., 2017). Studies have shown that the direct N_2O emissions of biological processes in WWTPs can contribute by up to by ~78% to the operational carbon footprint (Daelman et al., 2013b).

Significant N_2O emissions have been also reported from the biological treatment of high-strength wastewater streams. The anaerobic supernatant is a by-product from the treatment of the primary and secondary sludge via anaerobic digestion when the digestate is dewatered. This stream is small in volume (1-2% compared to the mainstream line), but very concentrated in nutrients and is conventionally recycled back to the primary treatment increasing the loads (and thus, the energy requirements and costs) of the mainstream biological treatment (i.e. contains 10-20% of the WWTP nitrogen load). For this purpose, biological technologies (i.e. partial-nitritation – anammox (anaerobic ammonium oxidation, nitritation/denitritation etc.)) have been developed, to treat sidestream high-strength streams in a cost and energy efficient way (Lackner et al., 2014; Zhou et al., 2018). In the sidestream biological processes, favourable condition for N_2O generation can prevail (i.e. NO_2^- accumulation, elevated NH_4^+ concentrations etc.). Studies have shown that biological processes treating high-strength streams can contribute by over 90% to the total direct N_2O emissions compared to the mainstream BNR processes (Schaubroeck et al., 2015)

The recent, mitigation roadmap to carbon neutrality in urban water published by Water and Wastewater Companies for Climate Mitigation (WaCCliM) project and the International Water Association (IWA) (Ballard et al., 2018), states that direct N_2O emissions in water utilities, should be considered for the carbon footprint assessment, reporting and mitigation. However, in practice, the quantification of direct N_2O emissions at WWTPs via monitoring campaigns is not a regulatory requirement. Therefore, wastewater utilities usually, estimate N_2O emissions

via theoretical methods i.e. based on the population equivalent of the WWTP (IPCC, 2006); the latter can significantly underestimate the actual emissions (Cadwallader Adam VanBriesen, 2017).

Several full-scale monitoring campaigns have been implemented in full-scale biological reactors to provide insights on the dynamics and triggering mechanisms for N_2O generation. However, benchmarking N_2O emission fluxes at wastewater treatment processes is still challenging. The exact generating mechanisms, operational and environmental conditions for its formation and long-term dynamics are still investigated (Peng et al., 2015). This is mainly because N_2O fluxes are characterised by significant spatial and temporal variability due to the different interacting biological processes that consume or produce N_2O and the variation of operational conditions (Daelman et al., 2015; Gruber et al., 2019). Mechanistic process-based models have also been developed over the past years aiming to integrate N_2O emissions generation of different processes in the design, operation and optimisation of biological processes (Domingo-Félez et al., 2017a; Mannina et al., 2016; Massara et al., 2017). However, their online integration for the reliable quantitative estimation of N_2O emissions is still a main challenge (Haimi et al., 2013; Mampaey et al., 2019).

Currently operational strategies at WWTPs do not consider the mitigation of direct GHG emissions. New objectives, environmental and sustainability targets in the water industry require novel approaches dynamically integrating new parameters (i.e. GHG emissions sensors, energy meters) into the process monitoring, control and decision making (Corominas et al., 2013; de Faria et al., 2015; Flores-Alsina et al., 2010; Longo et al., 2016; Sweetapple et al., 2014)

At present a significant amount of raw, heterogenous operational data are available from the daily operations of the WWTPs (Olsson et al., 2014). However, in practice data generated in WWTPs are underexploited (Newhart et al., 2019). The utilisation of the information hidden in the raw sensor signals can be the driving force of sustainable and efficient wastewater operations. Data-driven modelling and data-mining techniques have been applied in the wastewater sector in order to i) optimise monitoring and efficiency of wastewater processes (Mirin and Wahab, 2014) and improve process understanding (Moon et al., 2009), ii) detect process failures (Haimi et al., 2016; Liu et al., 2014; Maere et al., 2012) and erroneous sensor measurements (Corominas et al., 2011; Lee et al., 2004) and iii) predict significant operational variables in the biological systems (i.e. COD, nitrogen effluent concentrations etc.) (Lee et al.,

2002; Rustum et al., 2008). Advanced data management and knowledge extraction techniques from wastewater data can maximise the information coming from N_2O monitoring campaigns. The current PhD thesis develops and applies a framework for analysing the data and extracting useful conclusions for the behaviour of the system and the generated N_2O emissions.

1.2 Overview of research program

1.2.1 *Research questions addressed by this thesis*

The primary research questions (RQ) that will be addressed in this thesis are:

- What are the recent findings of full-scale N_2O monitoring campaigns in terms of EFs, triggering mechanisms and mitigation measures for different process groups? How can the design of monitoring protocols (i.e. frequency, sampling method) impact the results of monitoring studies in terms of the quantified N_2O EFs? What are the current techniques for data processing and analysis that are used to provide insights on N_2O dynamics? What are the research needs that can support the integration of N_2O emissions into process control (Chapter 2)?
- How can data mining techniques link the different variables monitored online in biological processes and assess the temporal system performance? Can the combined effect of operational variables on N_2O emissions be identified? (Chapter 3).
- Environmental and operational conditions at wastewater processes are not static and are characterised by temporal changes. What methods can be used to detect these changes? Can structural changes in the profiles of key operational variables (i.e. DO, NH_4^+) indicate changes in the N_2O dynamics? How can historical data from monitoring campaigns be used to develop data-driven risk-based models of the N_2O emissions' range (Chapter 4)?
- Is it possible to develop reliable data-driven models to predict N_2O behavior in wastewater systems? What is the role of domain knowledge when data-mining techniques are used in the wastewater sector (Chapter 5)?
- What type of information can be extracted from wastewater databases, when data-mining techniques are integrated into the analysis within a structured framework for knowledge discovery to support WWTP operation (Chapters 3-5)?

1.2.2 *Aim and objectives*

Research hypothesis: *Sensor data and laboratory analyses from wastewater treatment processes contain hidden information, that can be valorised to explain the long-term N_2O emissions' dynamics and triggering operational conditions.*

The aim of the current research is to develop a framework, coupling wastewater domain knowledge with data-mining techniques in order to enhance the understanding of the long-term N_2O emissions dynamics in full-scale biological reactors. A structured approach for knowledge discovery from wastewater databases is developed transforming readily available WWTP data from different and heterogeneous sources into actionable information. The purpose of the research is to build on the existing plant monitoring and information and communication technology (ICT) capabilities of selected wastewater treatment processes, and to investigate how methods to translate data into meaningful information using classical but relevant techniques, can be used to provide insights on the dynamic behaviour of N_2O emissions under different operational conditions. The outcome is expected to form the basis for control and reduction of the N_2O emissions from wastewater treatment processes.

The specific objectives of this thesis are to:

- Investigate the existing body of knowledge regarding past N_2O monitoring campaigns in full-scale wastewater treatment processes, evaluate current data analysis methods and visualization techniques. The outcome of this investigation will be used to provide insights on N_2O emissions' dynamics and link i) N_2O emission factors (EFs), ii) frequently reported N_2O generation mechanisms, iii) triggering operational conditions and iv) mitigation measures, with different wastewater process groups

- Investigate whether multivariate statistical techniques applied in long-term wastewater and N_2O emission datasets can provide insights on the operational conditions. The aim is to link with specific ranges of N_2O emissions, isolate sets of variables related to N_2O emissions and facilitate the detection of system disturbances and their influence on N_2O fluxes

- Develop a framework that classifies operational conditions of treatment processes with different behavior of the variables monitored online. The aim is to identify links with

specific ranges of N_2O fluxes, that can reduce the frequency and duration of the sampling requirements for reliable estimation of N_2O EFs

- Investigate whether data-mining techniques and emerging models can be used to accurately predict the pattern of N_2O emissions, in wastewater treatment processes.

1.2.3 *Methodological approach*

In the first part of the thesis, a review of the findings from recent full-scale N_2O monitoring campaigns in full-scale wastewater processes is conducted. The analysis covers the quantification and mitigation of emissions for different process groups, focusing on techniques that have been applied for the identification of dominant N_2O pathways and triggering operational conditions using data and N_2O emissions and other operational variables.

The second part of the research investigates how wastewater treatment domain knowledge can be coupled with data-mining techniques, to extract useful information from wastewater sensor data and laboratory analyses, with the view to maximise the insights on the long-term N_2O dynamics. The framework for knowledge discovery from WWTP databases is summarised in Figure 1.1 whereas Table 1.1 includes the specific steps undertaken in each phase of the framework. It has been adjusted from the framework proposed by Fayyad et al., (1996) for 'Knowledge discovery and data-mining' in databases of several sectors (i.e. businesses, manufacturing etc.).

As shown in Figure 1.1 and Table 1.1, the first phase of the knowledge discovery framework involves i) the investigation of existing knowledge on the behaviour of N_2O emissions in wastewater systems, ii) gaining insight into the operating principles, the processes' characteristics and the expected challenges associated with wastewater processes and iv) the definition of initial data-mining goals. Wastewater sector has unique characteristics that need to be considered during the analysis. Fundamental considerations include: i) the volume and quality of influent flow-rates at a WWTP are characterised by significant temporal changes and cannot be controlled by wastewater operators, ii) wastewater streams cannot be stored in large quantities at a WWTP or discharged untreated to the environment, iii) changes of variables occur at different time-scales; from seconds (i.e. conductivity) to even years (i.e. population increase) (Corominas et al., 2018).

Conventional and innovative wastewater treatment technologies have been analysed. The first dataset has been acquired from a 15-month long N_2O monitoring study in a mainstream secondary treatment system in Netherlands, consisting of a plug-flow followed by two parallel Carrousel reactors. The second dataset has been obtained from a sidestream SBR treating the anaerobic supernatant via short cut enhanced nutrient abatement process (SCENA). The SCENA process removes nitrogen from the anaerobic supernatant via nitritation-denitritation and phosphorus can be recovered in the form of phosphorus-rich sludge via enhanced biological phosphorus removal via nitrite with the alternation of anaerobic-aerobic-anoxic phases. Additionally, part of the mixed sludge is fermented for the production of volatile fatty acids (VFAs) that are subsequently dosed during the anoxic and anaerobic phases.

The second phase (Figure 1.1) involves the collection, exploration and initial understanding of the data. The analysis considers several parameters, such as the visual exploration of the data, the communication with the operators to maximise the understanding of process operation and data quality and the initial understanding of seasonal effects, abnormal events and dynamic system performance. This is one of the most time-consuming phases and guides towards the identification of data-mining methods that are more promising to test (i.e. linear vs non-linear methods). The inputs of the analysis consist mainly of the real-time monitored process variables (i.e. NH_4, NO_3, ORP, conductivity, temperature probes) due to their high resolution. However, laboratory analyses have also been considered. Regarding GHG emissions data different analysers have been used i) Low-cost commercial probes monitoring the dissolved N_2O concentration (SCENA system) and ii) conventional gas analysers coupled with gas chambers (SCENA system) or covered reactors (plug-flow and two Carousel reactors) for the collection and quantification of N_2O in the gaseous phase. The data obtained from the target biological processes have been assessed and explored in terms of frequency, duration, quality and quantity in order to obtain an initial understanding of the systems' operational behaviour and dependencies with the N_2O emissions.

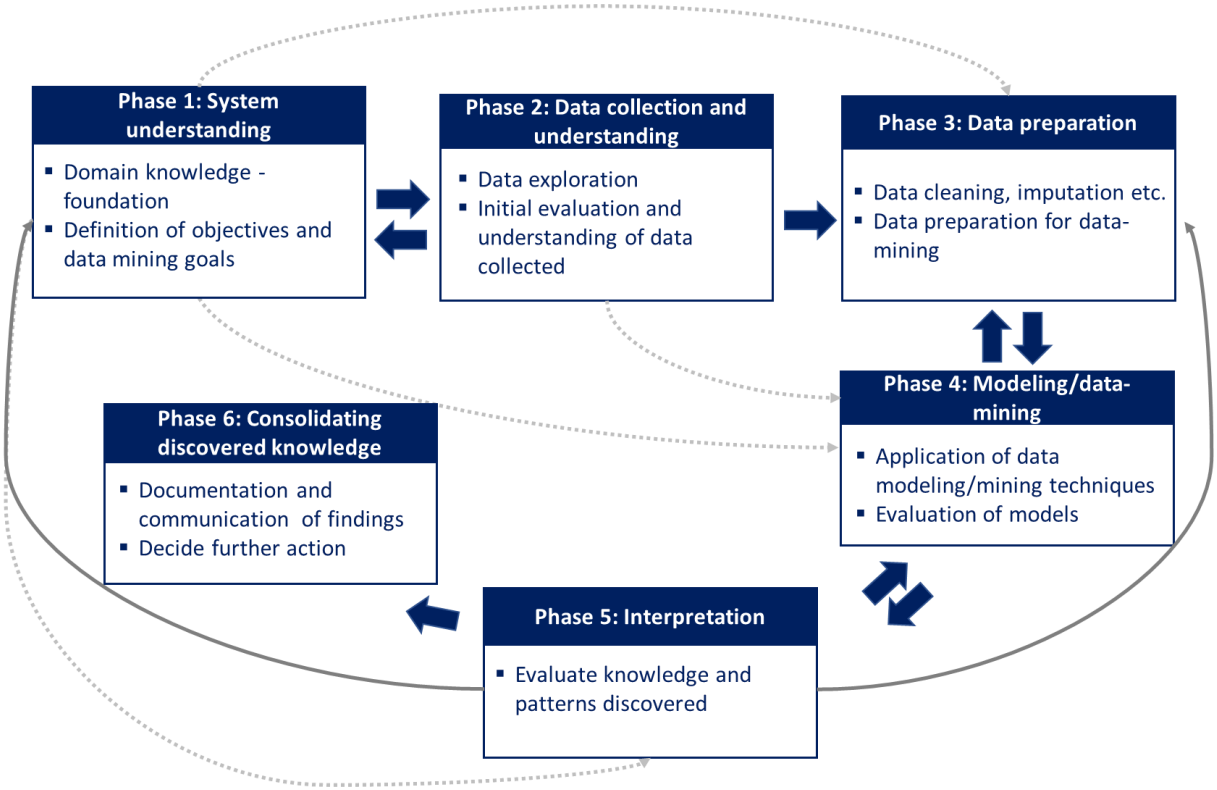


Figure 1.1: Structured Framework applied in the thesis for knowledge discovery in WWTPs adjusted by Fayyad et al., (1996)

Then data cleaning and appropriate transformations (i.e. dimensionality reduction, feature extraction, aggregation etc.) are applied based on phase 2 findings and the requirements of the selected data-mining approach.

Data-mining algorithms are applied, and the results are analysed and evaluated. The applied methodology depends on the application, the initial challenges, and the data exploration findings. Several methods, (i.e. clustering, classification, prediction) combined or standalone have been applied to analyse and assess the datasets. The description of the methods is provided in the relevant chapters. As shown in Figure 1.1, this procedure is not linear, and several feedback loops are required to leverage the knowledge discovered and adjust the data-preparation and mining phases. Whereas findings from the first phase of the analysis provide feedback to all subsequent phases (Figure 1.1). The final phase includes the consolidation, communication and reporting of the knowledge gained.

The tested methods can serve towards i) understanding of dependencies between N_2O emissions and operational conditions ii) identifying operational conditions with a high probability to result into high N_2O emissions iii) understanding operational conditions that can mitigate N_2O emissions.

Using the existing monitoring capabilities of biological processes and introducing GHG sensors the methodological framework translates the N_2O emissions into timely performance measures facilitating optimal process control. Historical information is used, to facilitate the interpretation of emissions data. The dependencies between the system variables monitored online or through laboratory analyses and the N_2O emissions are investigated. The analysis provides added value to the existing knowledge and sustainable operation of the wastewater treatment processes. The method can potentially lead to the identification of combinations of operating variables that have high probability to minimise GHG emissions in the target biological processes. Additionally, it can facilitate the visualization of set-points and operating parameters that optimise the operational carbon footprint of the processes.

Table 1.1: Detailed steps of the knowledge discovery framework applied

	Phase	Main steps
1.	**System understanding**	• Investigate current findings on N_2O emissions, triggering mechanisms, operational conditions in processes with similar operational principles • Evaluate data and investigate links with N_2O emissions • Define objectives and data mining goals • Develop a project plan
2.	**Data collection and understanding**	• Data exploration at different scales in time, visual exploration of the data enhanced by advanced visualization techniques • Communication with site operators for an initial understanding of data • Initial evaluation and understanding of abnormal/extreme events • Definition of treated effluent limits
3.	**Data preparation**	• Data cleaning, imputation and homogenisation • Identification and removal of outliers, abnormal/extreme events (if applicable in the following step) • Application of data transformation methods (i.e. feature selection, discretisation, standardisation etc.) • Application of other data preparation techniques
4.	**Modelling / data-mining**	• Data modelling/mining techniques (i.e. classification, clustering, regression etc.) selection • Generation of test design • Application of the method
5.	**Interpretation**	• Detect patterns / evaluate applied algorithm accuracy • Interpretation / visualization of patterns • Next step determination (i.e. loop back to phases 3 and 4)
6.	**Consolidating discovered knowledge**	• Documentation of results and actionable information • Communication of results • Incorporate model in the WWTP operation (if applicable)

1.2.4 *Thesis outline*

Figure1.2, summarises the main objectives of each chapter and how information extracted from each chapter is used in the subsequent analysis.

Chapter 2: Literature review - A decade of nitrous oxide (N_2O) monitoring in full-scale wastewater treatment processes

In this chapter, findings from full-scale N_2O monitoring campaigns of the past decade are investigated. The analysis classifies quantified N_2O EFs, triggering operational conditions and formation pathways and strategies proposed to minimise N_2O emitted for different process groups. Main reasons for EF discrepancies are investigated. Additionally, advantages and limitations of the techniques (i.e. data analysis, modelling, isotopic analysis) used to provide insights on N_2O generation and dynamics are analysed. Future work in the field of GHG monitoring and quantification at WWTP has been identified and reported.

Chapter 3: Relating N_2O emissions during biological nitrogen removal with operating conditions using multivariate statistical techniques

This chapter investigates whether 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 behaviour of N_2O emissions. It explains the combined effect of the operating variables monitored in wastewater treatment processes on N_2O emissions generation. A statistical methodological approach is developed applying changepoint detection techniques to identify changes in the N_2O fluxes behaviour combined with hierarchical k-means clustering and principal component analysis (PCA). It also provides insights on N_2O emissions patterns and generation pathways. The methods rely on process data coming from a 15-month long N_2O monitoring campaign at a covered reactor using a conventional gas analyser.

Chapter 4: Data-driven *versus* conventional N_2O EF quantification methods in wastewater; how can we quantify reliable annual EFs?

This chapter builds upon the work undertaken in chapter 3 and reverses the target of the changepoint detection algorithm. It investigates whether identification of structural changes in environmental and operational variables in wastewater treatment systems can provide insights in changepoints in the N_2O emissions range. Knowledge-based N_2O sampling can minimise GHG sampling requirements without compromising the reliability of N_2O emissions estimates.

Additionally, a classification model based on support vector machines (SVM) has been developed tested and validated using data from variables monitored online to predict the range of N_2O emission loads (i.e. low, medium, high). A methodological framework is developed that can i) identify the state of the system at a specific time-period and ii) predict the N_2O emissions behaviour based on historical data.

Chapter 5: A knowledge discovery framework to predict the N_2O emissions in the wastewater sector

In this Chapter the knowledge discovery framework is applied to a full-scale sidestream sequence batch reactor (SBR) treating the anaerobic supernatant via nitritation-denitritation and via-nitrite enhanced biological phosphorus removal process (anaerobic, aerobic and anoxic phases alternating). Abnormal events detection, classification and regression techniques are applied to identify main conditions triggering N_2O accumulation in the system and test data-driven regression and classification models on their ability to predict the dissolved N_2O behaviour and concentration for the different reactor phases. The proposed approach can predict the N_2O emissions utilizing sensor data (online) that are conventionally monitored in biological batch processes.

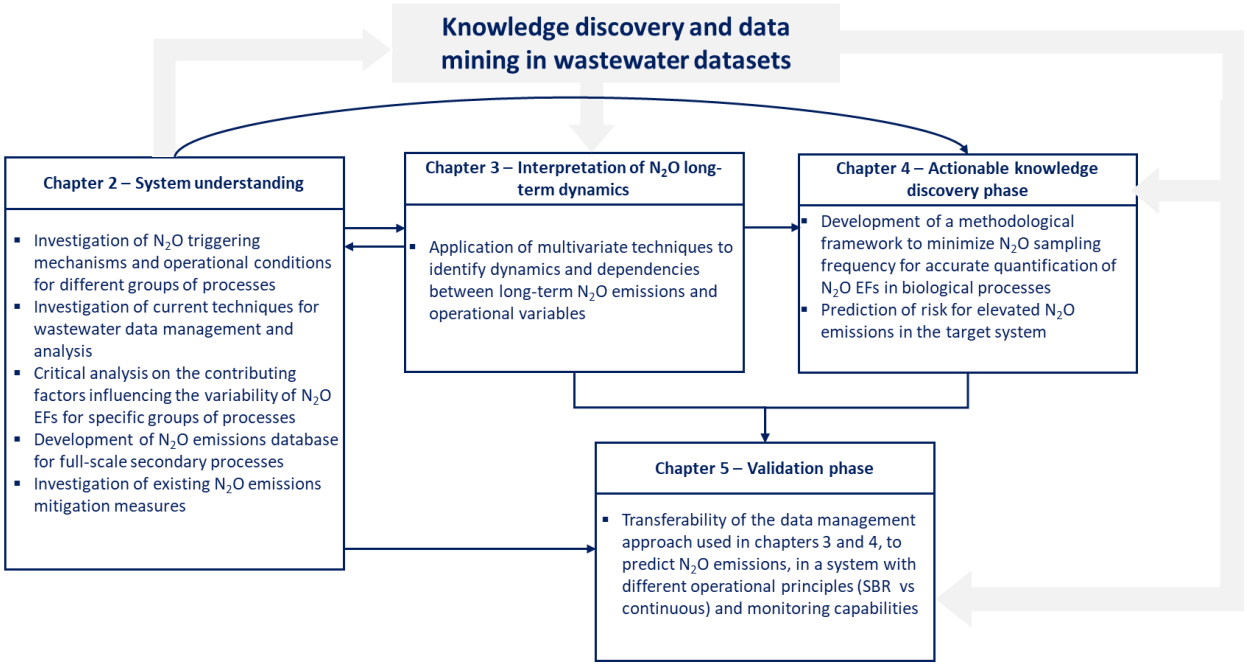


Figure 1.2: Main objectives of each chapter and how information extracted from each chapter is used in the subsequent analysis.

2. Literature Review

Wastewater treatment plants (WWTPs) produce substantial anthropogenic nitrous oxide (N_2O) amounts (Law et al., 2012). N_2O emissions from wastewater treatment processes have increased by 44% from 1990 to 2014 (US EPA, 2016). Moreover, they are responsible for up to 26% of the greenhouse gas (GHG) emissions of the whole water supply chain (including drinking water supply, wastewater collection and treatment, effluent discharge, sludge processing and disposal) (Lane et al., 2015). Therefore, the evaluation of the environmental impact of WWTPs is gaining increasing attention worldwide (Guo et al., 2017; Harris et al., 2015; Law et al., 2012; Mannina et al., 2016; Massara et al., 2017). N_2O has an approximately 300 times higher GHG effect than CO_2 (IPCC, 2013). Compared to other unregulated combined halocarbon emissions, reducing the anthropogenic N_2O emissions in the atmosphere plays the most significant role in preventing the ozone layer depletion (Daniel et al., 2010; Portmann et al., 2012).

Stricter environmental regulations are being imposed on WWTPs, while the need to control the energy consumption costs, pushes the plant managers towards the deployment of more energy and cost efficient operational strategies (Batstone et al., 2015; Ghoneim et al., 2016). The integration of additional sustainability metrics (e.g. GHG emissions, resource recovery) for the evaluation of the WWTP performance is also gaining attention (Flores-Alsina et al., 2014; Cornejo et al., 2016). Recent studies have revealed that the direct N_2O emissions of biological processes in WWTPs can increase the operational carbon footprint by ~78% (Daelman et al., 2013b) or even ~83% (Desloover et al., 2011). Moreover, the new technology adoption in WWTPs requires the consideration of trade-offs between direct and indirect GHG emissions to ensure that it will not result in increase of the overall carbon footprint. Since the GHG emissions have a major contribution to the overall environmental impact of WWTPs, they should be considered as part of the decision-making process for the improvement of the technological and operational plant performance (Sun et al., 2017; Frutos et al., 2018; Conthe et al., 2018). Therefore, research has been conducted in the past years to identify and suggest strategies leading to N_2O mitigation during the biological nutrient removal (BNR) processes (Desloover et al., 2012; Duan et al., 2017; Law et al., 2012). Different operating parameters, such as dissolved oxygen (DO), pH, ammonium (NH_4^+) and nitrite (NO_2^-) concentration, configuration types and environmental conditions (e.g. temperature), can affect the N_2O

production in WWTPs (Desloover et al., 2012; Kampschreur et al., 2009b; Massara et al., 2017). The exact triggering operational and environmental conditions that govern the N_2O generation are still under investigation by researchers and operators (Wan et al., 2019). Additionally, the exact mechanisms that determine the N_2O generation are not fully understood, thus hindering the establishment of mitigation measures (Duan et al., 2017). Consequently, the long-term dynamics of N_2O emissions in full-scale WWTPs cannot be fully explained, even for conventional processes (e.g. plug-flow (PF) reactors) (Daelman et al., 2015).

Different data, sampling techniques and analytical tools have been employed in existing full-scale N_2O studies. The objectives of the current review paper are to: i) evaluate recent findings from N_2O monitoring campaigns in terms of N_2O EFs, dominant pathways and mitigation measures for different groups of full-scale configurations and reactor types, ii) examine the discrepancies among the methods applied in different monitoring campaigns, and iii) evaluate the data provided by studies investigating N_2O emissions in WWTPs. The current study critically assesses how the findings of the research on N_2O emissions have been extrapolated in different full-scale process groups. Moreover, this work evaluates the methods that have been applied at full-scale systems for the N_2O quantification, control and mitigation. Finally, the research gaps that need to be addressed in future studies are highlighted.

2.1 The impetus for quantifying the N_2O emissions in WWTPs

2.1.1 *EF benchmarks*

EF quantification in conventional and advanced wastewater treatment processes is essential to assess and reduce the resulting environmental impact (Foley et al., 2015). Current methods for the theoretical calculation of N_2O EFs at WWTPs rely on fixed (Palut and Canziani, 2007) or country-specific EFs and similar over-simplified approaches (Singh and Maurya, 2016). These methods underestimate the actual emissions and are considered unreliable (Cadwallader and VanBriesen, 2017), since they are not representative for different process configurations, operational and environmental conditions. Therefore, a main target of the studies analysed in this paper was to: i) develop an N_2O emission database from different mainstream (Ahn et al., 2010b, 2010a)/sidestream (Kampschreur et al., 2008b; Weissenbacher et al., 2010) processes and innovative configurations (Desloover et al., 2011), and ii) identify N_2O EFs from processes

located in previously unreported regions with different environmental conditions (J. Wang et al., 2011).

2.1.2 *N_2O pathways and triggering operational conditions*

The majority of the full-scale N_2O monitoring campaigns were driven by the need to expand the knowledge on the N_2O generation in wastewater treatment systems. Several studies have highlighted that it is important to understand the dominant N_2O generation pathways in full-scale processes and investigate the effect of specific operational conditions triggering specific enzymatic reactions linked with elevated EFs in full-scale biological systems (Bollon et al., 2016a; Pan et al., 2016; Wang et al., 2016b).

The N_2O production during wastewater treatment involves several microbiological reactions during both autotrophic and heterotrophic processes that require either aerobic or anoxic conditions. Three main biological pathways have been identified in BNR systems; hydroxylamine (NH_2OH) oxidation, nitrifier denitrification and heterotrophic denitrification (Kampschreur et al., 2009b). The NH_2OH oxidation pathway is mainly catalysed by the autotrophic ammonia oxidizing bacteria (AOB) and ammonia-oxidizing archaea (AOA). However, several studies have shown that the contribution of AOA to N_2O emissions in wastewater is expected to be low (Hooper, 1968; Ritchie and Nicholas, 1972; Law et al., 2012). Caranto et al. (2016) have recently showed that N_2O can also be the main product of anaerobic NH_2OH oxidation catalysed by the cytochrome P460 in *N. europaea*. The latter can be considered as evidence of biological N_2O generation under limited DO and high NH_3 concentrations (i.e. Law et al., 2012), or as potential explanation for the high N_2O emissions observed during the transition from aerobic to anoxic conditions. White and Lehnert (2016) have also suggested that N_2O can be directly produced during the NH_2OH oxidation (mediated by the NH_2OH oxidoreductase (HAO) enzyme) under aerobic conditions, whereas the NO_2^- detected can result as by-product of the nitric oxide (NO) oxidation. The AOB can also reduce NO_2^- to NO (by the aid of *nirk*) and, subsequently, NO to N_2O (catalysed by *norB*) mainly under oxygen-limiting conditions via the other nitrification-related pathway (nitrifier denitrification) (Poth and Focht, 1985). During denitrification, the heterotrophic denitrifiers are responsible for the reduction of NO_3^-/NO_2^- to nitrogen gas (N_2). N_2O is an intermediate of denitrification (Schulthess and Gujer, 1996). With the NO reductase (NOR) as catalyst, NO is reduced to N_2O (Hochstein and Tomlinson, 1988). NO can also result as by-product of the

incomplete NH_2OH oxidation and then serve as a substrate during denitrification (Hooper and Terry, 1979). If the denitrification process continues undisturbedly, N_2O is reduced to N_2 in the final denitrification step (catalysed by the N_2O reductase (N_2OR)). Consequently, the heterotrophic denitrification process can act either as a sink or as a source of N_2O (Robertson and Tiedje, 1987). Under elevated NH_2OH and NO_2^- concentrations, abiotic yet biologically-driven N_2O pathways can also constitute important contributors to the N_2O emissions (Soler-Jofra et al., 2016; Terada et al., 2017; Harper et al., 2015). Inside nitritation reactors for example, the abiotic-biotic pathway of nitrosation is possible; NO_2^- can react with the biologically produced NH_2OH and form N_2O as end-product (Zhu-Barker et al., 2015).

Based on the existing knowledge on the N_2O production pathways, recent reviews on N_2O emissions from wastewater treatment processes have concluded that the key operational variables responsible for the N_2O generation include but are not limited to the following: i) low DO, NO_2^- or free nitrous acid (HNO_2) accumulation and changes in the NH_4^+ concentration in the nitrifying zones, ii) limitation of organic substrate (i.e. low chemical oxygen demand to N (COD:N) ratio), as well as, NO_2^- accumulation in the denitrifying zones, iii) alternation of anoxic/aerobic conditions and iv) abrupt changes in the processes and system shocks (Duan et al., 2017; Guo et al., 2017; Law et al., 2012; Massara et al., 2017).

Therefore, N_2O emissions can occur because of diverse contributing factors and enzymatic reactions. However, these parameters and reactions can occur simultaneously, dynamically and beyond operators' control in full-scale systems, whereas small changes (e.g. DO changes) can significantly affect the N_2O formation. Previous studies on full-scale monitoring campaigns intended to: i) identify the most important operating conditions (e.g. aeration rate, DO, NO_2^- concentration, pH, etc.) and correlate them with the N_2O generation (Brotto et al., 2015; Rodriguez-Caballero et al., 2014), ii) reveal the effects of seasonal variations on the N_2O formation (Yan et al., 2014), and iii) identify the key pathways for the N_2O production (Wang et al., 2016b).

2.1.3 *N_2O mitigation strategies*

Another key objective of the N_2O monitoring campaigns performed in the past years, particularly significant for the WWTP operators, was the development of operational strategies for the minimisation of the emissions (Desloover et al., 2012). Therefore, several authors have suggested N_2O mitigating measures based on the findings of full-scale N_2O monitoring

campaigns (i.e. Chen et al., 2016; Mampaey et al., 2016; Pan et al., 2016; Wang et al., 2016b). The proposed strategies to control N_2O emissions, are analysed in the following sections.

2.2 EF estimation using full-scale N_2O monitoring data

This section emphasizes the need to increase the comparability amongst different studies that report N_2O emissions. Moreover, it investigates potential trends in the EFs for certain groups of processes and summarises the data requirements for the EF assessment in WWTPs. In cases where the EFs were reported with respect to units other than the influent total nitrogen (TN) or the influent NH_4^+ content, appropriate conversions were made (where possible) (see the Appendix A – Table A.2).

Figure 2.1 shows the percentage of past full-scale N_2O monitoring campaigns with reference to the treatment configurations applied each time. All processes considered in Figure 2.1 are given in the Appendix A (Tables A.3 and A.4). Distinct mainstream process configurations include, modified Ludzack-Ettinger (MLE) reactors, conventional activated sludge (CAS) systems (only aerobic reactors), A^2/O (anaerobic/anoxic/aerobic) processes and A/O (anoxic/aerobic) reactors. Additionally, oxidation ditch (OD) reactor types and sequence batch reactor (SBR) types have been considered as distinct process groups. Sidestream processes that include partial-nitritation reactors, 1-step and 2-step partial-nitritation-anammox configurations) are considered a distinct process group in Figure 2.1. The processes that do not belong to the aforementioned process groups are categorized separately (i.e. Baresel et al., 2016; Mello et al., 2013; Wang et al., 2016a). Details for all the processes are provided in the Appendix A.

Most of the process-focused monitoring campaigns include, CAS systems (~12%), (MLE) configurations (~12%), A^2/O configurations (~10%), oxidation ditches (ODs) (~8%) and sidestream partial-nitritation reactors or anammox systems (~11%) (Figure 2.1). Overall, these studies refer to a wide range of configurations (i.e. Aboobakar et al., 2013; Brotto et al., 2015; Castro-Barros et al., 2015; Sun et al., 2017) that have been monitored mainly for short periods (i.e. Ahn et al., 2010b; Bellandi et al., 2018; Foley et al., 2010a; Pan et al., 2016) with varying methodology (e.g. different gaseous sampling and analytical measurement protocols) (Daelman et al., 2015; Ren et al., 2013).

EF comparability limitations and benchmarking amongst the various processes will be discussed in the following sections. However, as a general remark, the identification of potential emission patterns and the EF classification for specific groups of processes is still challenging, mainly due to differences in monitoring strategies, operational conditions and length of monitoring periods among the existing studies. Additionally, there is still little real-field data regarding N_2O emissions for several conventional and advanced biological processes (e.g. trickling filters, denitrifying packed bed reactors, biofilm or hybrid partial-nitritation anammox systems, etc.).

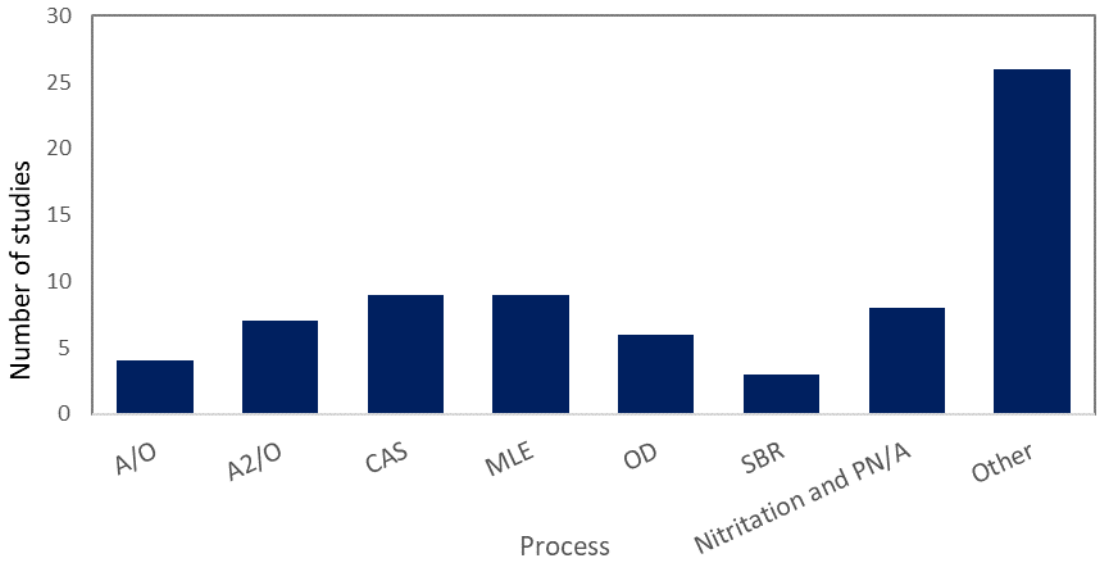


Figure 2.1: EF of secondary and sidestream treatment processes

The N_2O emissions of the full-scale wastewater treatment processes reported in past studies vary significantly; e.g. ranging from 0.0025% of the TN-load for a mainstream MLE reactor (Spinelli et al., 2018) to 5.6% of the TN-load for a mainstream aerobic/anoxic settling SBR reactor (Sun et al., 2013). Overall, the potential of N_2O emissions from sidestream reactors (ranging from 0.17% to 5.1% of the influent N-load – Appendix A, Table A.3) is considered higher compared to the mainstream BNR processes. The latter is mainly because the nitritation/nitrification occurring during sidestream treatment is linked with higher ammonia oxidation rate (AOR) and NO_2^- accumulation (Desloover et al., 2011; Gustavsson and la Cour Jansen, 2011; Kampschreur et al., 2008b). Figure 2.2 shows boxplots of the observed EFs (with respect to the influent N-load) based on the treatment step. The width of the shaded area surrounding the boxplots represents the data kernel density distribution of the EFs. Specific information for the mainstream and sidestream technologies included in Figure 2.2 can be

found in in the Appendix A (Tables A.1, A.3 and A.4). Average N_2O emissions for the studied mainstream processes is equal to ~0.87% of the N-load, whereas the majority of the quantified EFs are below 0.27% of the influent N-load according to Figure 2.2. On the other hand, Figure 2.2 shows that N_2O EFs resulting from the treatment of the anaerobic digestion supernatant (sidestream process) are highly concentrated just below the median (2% of the N-load). On average, ~2.1% of the N-load is emitted as N_2O in sidestream processes (Appendix A, Table A.4). According to a life cycle assessment (LCA) study quantifying the direct GHG emissions for a WWTP in Austria, the sidestream DEMON process contributed by over 90% to the total direct N_2O emissions compared to the mainstream BNR (Schaubroeck et al., 2015). However, examples of full-scale sidestream Anammox processes with EFs lower than 1% exist in the literature and demonstrate that the configuration and efficient operational strategies can mitigate a significant percentage of the N_2O produced (Joss et al., 2009; Weissenbacher et al., 2010).

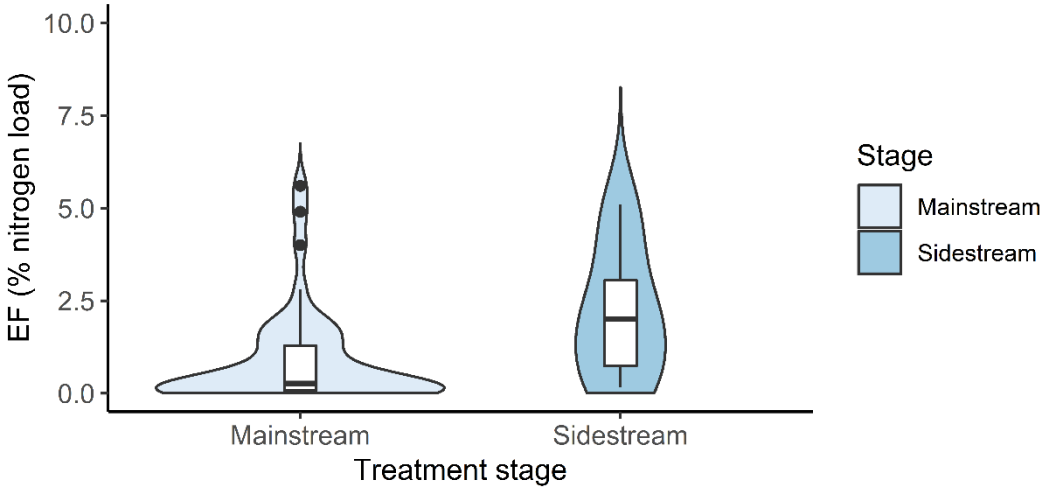


Figure 2.2: Boxplots of the of the reported EFs with respect to the stage of the treatment processes (i.e. mainstream or sidestream) using violin plot outlines. The rectangles represent the interquartile range. The median is denoted by the black horizontal line dividing the box in two parts. The dots represent the values exceeding 1.5 times the interquartile range. The upper and lower whiskers stand for values higher or lower the interquartile range, respectively (within 1.5 times the interquartile range above and below the 75th and 25th percentile, respectively). The violin plot outlines show the kernel probability density of the EF in mainstream and sidestream processes; the width of the shaded area represents the proportion of the data located there.

Figure 2.3 shows the EF boxplots of various processes applied in WWTPs. In total 51 systems were considered in Figure 2.3 (Appendix A, Tables A.1 and A.4). A general remark is that the N_2O EF for the majority of the different process groups shown in Figure 2.3 varies from 0.01 – 2% of the N-load. Discrepancies in the emission loads are observed in the majority of the different process groups and can be partially attributed to the different site-specific operational characteristics and control parameters. This indicates that apart from the reactor configuration, emission fluxes depend also on the operational/environmental conditions and preferred enzymatic pathways (Wan et al., 2019).

Mainstream SBRs are generally associated with higher N_2O emissions compared to the other process groups. EFs range between 2% of the influent TKN-load for an SBR operating under aerated feeding, aerobic, settling and decanting sequences (Foley et al., 2010) and 5.6% of the influent TN-load for an SBR operating under aerated feeding, aerobic and anoxic settling and decanting sequences (1h each). High N_2O fluxes in SBRs are attributed to sudden changes in the concentrations of NH_4^+ and NO_2^- within the cycle (compared to other configurations) or to accumulated dissolved N_2O during anoxic settling and decanting in the subsequent aerobic phase (Pijuan et al., 2014).

OD reactor types have been linked with relatively low N_2O emissions (average equal to 0.14% of the N-load), probably due to the strong dilution of the reactor concentrations (very high recycling rates) and less sensitivity to system shocks. One exception is the study of Daelman et al. (2015) who monitored a covered anaerobic/anoxic/oxic plug-flow reactor followed by two parallel Carrousel reactors for 1 year and found that the system had EF equal to 2.8% of the N-load. The authors argued that in the Carrousel reactor, the surface aerators led to zones with limited oxygen concentration to allow for complete nitrification (leading to nitrite accumulation), whereas the anoxic zones were also limited to allow for complete denitrification.

CAS systems shown in Figure 2.3 consist of aerobic reactors (1-step feed or multiple step-feed) without dedicated anoxic zones for denitrification. They are characterised by average EF equal to 0.27% of the N-load whereas the NH_4^+ removal, ranges between 38% to 53%. Peak loads and recirculation of the anaerobic supernatant can be responsible for the N_2O fluxes observed in CAS systems, whereas high aeration rates have been reported, enhancing N_2O stripping (Chen et al., 2016).

MLE configurations have a median EF equal to 0.857% of the N-load. MLE processes with high EFs (up to 4% of the N-load) have been reported by Foley et al., (2010). Low EFs in MLE configurations (i.e. 0.003%, 0.065%) have been observed in reactors with diluted influent concentrations due to groundwater infiltration (Bellandi et al. 2018), nitrification efficiency less than 73% (Ahn et al., 2010b) and low TN removal (~59%) due to COD/TN < 1.9 (Spinelli et al., 2018). Low EF in MLE reactors ranging from 0.003% to 0.065% of the NH_4^+ load have been also reported in the studies of Caivano et al. (2017) and Bellandi et al. (2018); however, conversion of these EF to % N-load was not possible and were not included in Figure 2.3.

N_2O emission fluxes in A^2/O configurations, are relatively low, in the majority of the studies, with median equal to 0.1% of the N-load. One exception is the study of Wang et al., (2016b); they monitored an A^2/O reactor once per month for 1 year and showed that EF varied from 0.1 to 3.4 % of the N-load between different months. The DO concentration and operating conditions varied significantly in the reactor (i.e. DO ranged from 0.6 to 6.8 mg/L).

Limited N_2O monitoring studies exist in full-scale sidestream processes. One-stage granular anammox reactors have an average EF of 1.1% of the N-load. The same two-stage suspended biomass partial-nitritation and anammox process has been monitored in two studies (Kampschreur et al., 2008b; Mampaey et al., 2016). In these studies, the average EF in the partial-nitritation SHARON reactor was ~2.8% of the N-load and was elevated compared to full-scale one-stage anammox reactors. N_2O fluxes quantification, in lab and pilot-scale single-stage granular anammox reactors have shown EFs ranging from 0.1 to 12.19% of N-load (Wan et al., 2019). Therefore, more studies are required to establish reliable ranges of EFs in sidestream processes.

Differences in the reported N_2O fluxes are also observed in studies that apply similar configurations and operational conditions (Appendix A, Tables A.3 and A.4). For example, Kampschreur et al., (2008) and Mampaey et al. (2016) monitored the N_2O emissions in the same two-step SHARON-Anammox reactor system and observed EFs that were equal to 1.7% and 3.8% of the N-load, respectively. Apart from slightly different DO setpoints (2 mg/L in the work of Mampaey et al. (2016) and 2.5 mg/L in the work of Kampschreur et al., (2008), the operational conditions (i.e. temperature, influent N-load, system treatment efficiency, hydraulic retention time (HRT), etc.; Table A.3) during the two monitoring periods were quite similar. The main identified difference was the increased anoxic liquid N_2O formation during the anaerobic period of the partial-nitritation reactor in the study of Mampaey et al. (2016). In

terms of monitoring protocols, Kampschreur et al. (2008) collected grab-samples for approximately 3 days, whereas continuous gas monitoring for 21 days was conducted by Mampaey et al. (2016). Moreover, the air infiltration in the covered reactor due to negative pressure was not considered in the study of Kampschreur et al. (2008).

The aforementioned examples emphasize the difficulty in the comparison of GHG emissions among various studies and development of EF databases for process groups. The benchmarking of GHG emissions of different plants can be hampered even if the same monitoring protocol is applied to monitor processes belonging to the same group. There are also cases where emissions have been measured according to different analytical procedures within the same system; fact that adds further difficulty in the comparisons. The relatively short monitoring periods and varying monitoring strategies can also influence the comparability and accuracy of the reported EFs. This will be discussed in detail in the following sections.

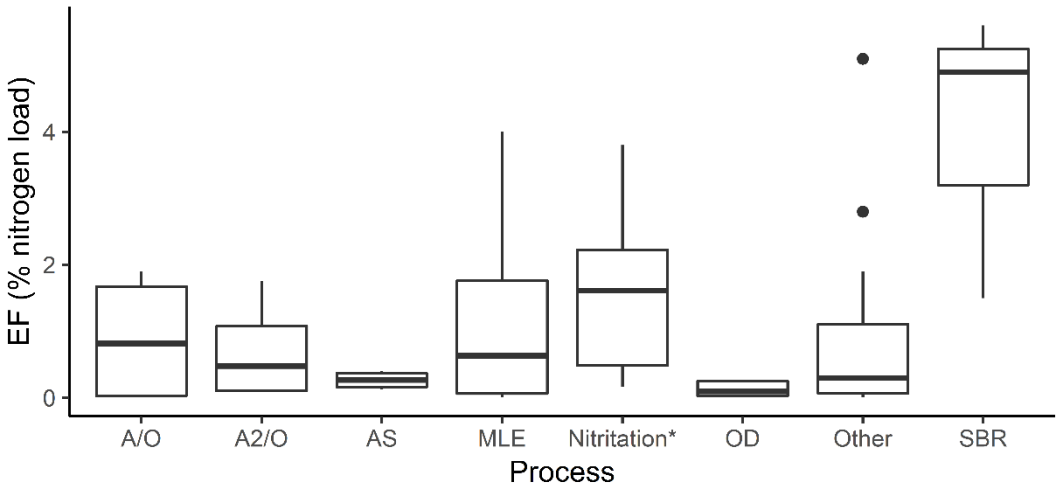


Figure 2.3: Boxplots visualizing the EF range for the different groups of mainstream processes. The rectangles represent the interquartile range. The median is denoted by the black horizontal line dividing the box in two parts. The dots represent values exceeding 1.5 times the interquartile range. The upper and lower whiskers represent values higher or lower the interquartile range, respectively (within 1.5 times the interquartile range above and below the 75th and 25th percentile, respectively).

2.2.1 *Duration of monitoring campaigns and seasonality*

Seasonal environmental variabilities, such as temperature, can influence the bacterial community structure in WWTPs (Flowers et al., 2013). The N_2O formation and emissions during the BNR are expected to have temporal variations. Temperature can significantly affect

the AOB specific growth rate during nitrification (Van Hulle et al., 2010). The higher temperature also decreases the N_2O solubility, thus intensifying the N_2O stripping to the atmosphere (Reino et al., 2017). On the other hand, Adouani et al. (2015) observed that the N_2O emissions increased up to 13%, 40% and 82% of the TN-removed at temperatures equal to 20°C, 10°C and 5°C, respectively, in a batch reactor fed with synthetic wastewater. The latter was attributed to the increased sensitivity of the N_2O reductase activities at lower temperatures compared to other denitrification enzymes and, therefore, to incomplete denitrification. Other seasonal variations (e.g. influent loading, wet and dry season) can also impact on the enzymatic reactions and affect the emissions. Vasilaki et al., (2018) observed peaks of N_2O emissions coinciding with precipitation events, at low temperatures, in an OD during a 15-month monitoring campaign. However, further investigation is required to understand potential seasonal effects on the N_2O emissions.

The monitoring periods of the full-scale campaigns for all the processes considered in this analysis are summarised in Appendix A, Tables A.3 and A.4. Figure 2.4 shows the EF for mainstream technologies based on the length of the monitoring period. Only studies that have reported EFs in terms of the influent N-load have been considered (Appendix A - Table A.1). The monitoring campaigns have been categorized into 3 distinct groups, based on their duration, i) short-term campaigns performed in a limited period of time (less than 1 month), ii) medium-term monitoring campaigns that last more than one month but have not captured all the temperature ranges observed in the system and ii) long-term monitoring campaigns that last at least 1 year. Both continuous and discontinuous monitoring studies have been included in the analysis. In the discontinuous N_2O monitoring studies, the gaseous N_2O fluxes have been grab-sampled (i.e. via gas-bags, closed chambers etc.) and subsequently quantified using analytical methods in the lab (offline monitoring) or intermittently sampled (i.e. with floating chambers for 1-2 days/month) but quantified continuously on-site (i.e. via GHG analysers) (online monitoring). Most discontinuous monitoring campaigns had monthly or bi-monthly sampling frequency. The only exception is the study of Ahn et al., (2010b) who monitored several systems in only two distinct seasons (warmest and coldest temperatures). Discontinuous studies with sampling extending over 1-month period, have been categorized as medium-term or long-term based on the duration of the study (less or more than 1 year). In continuous monitoring campaigns N_2O fluxes have been collected continuously (i.e. via chambers) and quantified online, on-site (i.e. via GHG analysers).

About 30% of the EFs shown in Fig 4. are based on monitoring periods lasting less than two days (Foley et al., 2010; Filali et al., 2013; Samuelsson et al., 2018). Annual EF variation has been investigated discontinuously (monthly or bimonthly sampling frequency) in approximately 10% of these systems (i.e. Sun et al., 2015; Wang et al., 2016b), whereas long-term (≥ 1 year) continuous monitoring campaigns have been performed only in two of the examined works (Daelman et al., 2015; Kosonen et al., 2016). SBRs have been excluded in order to avoid further biases in the results, since the reported average N_2O emissions are significantly higher than the average EF of other mainstream N-removal configurations. Sidestream technologies have been also excluded because they have not been monitored long-term to examine seasonal effects. The average EF is equal to 0.8 % (median 0.2 %) and 0.3 % (median 0.1 %) of the N-load for the systems monitored short-term and medium-term (but without capturing the whole spectrum of seasonality effects), respectively. The studies investigating seasonal trends of N_2O emissions reported an average EF of 1.5% (median 1.7%) of the N-load.

Daelman et al. (2013a) demonstrated that short-term campaigns, in the system investigated, are likely to produce unreliable EF estimates independently of the monitoring approach. Additionally, the authors found that short-term campaigns have a high probability to underestimate actual emissions. According to Figure 2.4, the highest gaseous N_2O loads belong to long-term continuous or discontinuous monitoring campaigns.

Long-term and medium-term campaigns have also shown a high variability of the N_2O emissions. Amongst the examined studies, Daelman et al. (2015) implemented the longest continuous real-field campaign that reinforced the existence of seasonal emission variability. Seasonality is also supported by the findings of several other studies (Brotto et al., 2015; Sun et al., 2013; Yan et al., 2014). Bollon et al. (2016a) and Bollon et al. (2016b) studied a sidestream nitrifying and post-denitrifying biofiltration system, respectively, by performing two monitoring campaigns; one in summer at 22.5°C and one in winter at temperatures lower than 14°C. Their results indicated a significant seasonal variation of the N_2O formation. The EF of the nitrifying filters was equal to 2.3% of the NH_4^+ removed during the summer campaign, and 4.9% of the NH_4^+ removed during the winter campaign. The dissolved N_2O concentration in the post-denitrifying biofilter effluent was equal to 1.3% in summer and 0.2% in winter with respect to the NO_3^- uptake.

Short-term monitoring periods are likely to miss underlying seasonal variations in the N_2O formation (or be affected by short-term process perturbations), and, consequently, complicate the direct cross-comparisons between different studies and their findings. For instance, the monitoring of an A^2/O process for nine months (grab-samples taken once per month) led to an average EF equal to 0.08% of the influent TN (Yan et al., 2014), whereas a similar A^2/O process monitored for two days (by taking grab-samples) presented an EF of 0.85% of the influent TKN (Foley et al., 2010). Wang et al., (2016b) showed that the EF from a PF A^2/O reactor was characterised by significant seasonality and varied from 0.01% to 3.5% of the influent TN; within the range of EFs reported in the studies by Yan et al., (2014) and Foley et al., (2010). It can be concluded that the EF differences between similar configurations shown in Figure 2.2 and Figure 2.3, are strongly affected from the seasonality of the emissions.

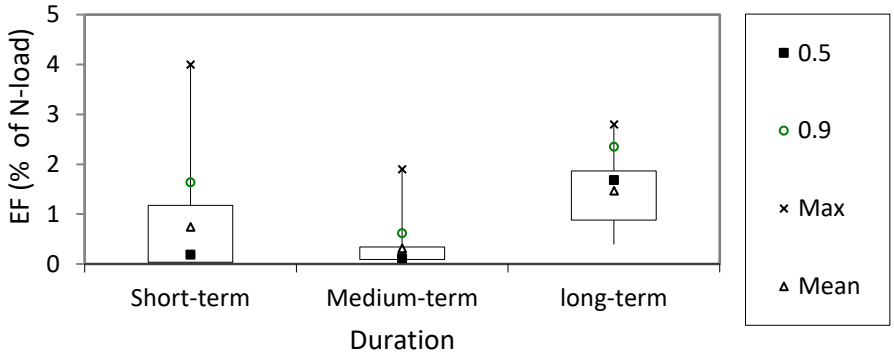


Figure 2.4: EF values with respect to the length of the monitoring period for mainstream treatment technologies.

2.2.2 *Monitoring and sampling methods*

Several authors have highlighted that the sampling methodology can influence the EF quantification (Daelman, 2013a; Aboobakar et al. 2013; Wang et al, 2016a; Kosonen et al., 2016; Hwang et al., 2016). Main sources of uncertainty of N_2O monitoring protocols are related to: i) the sampling methods, ii) the gas flux assessment methods, iii) the N_2O quantification methods, ii) the sampling location and v) the sampling frequency (continuous vs discontinuous monitoring) employed. A detailed analysis of the potential uncertainties in full-scale N_2O monitoring campaigns was out of the scope of this review. However, main sources of uncertainty are briefly discussed, to underline that further research is needed to reveal the influence of monitoring protocols and facilitate the benchmarking of EF values for different groups of processes.

- **Uncertainties of gas sampling methods**

The majority of the examined studies have selected chamber methodologies (~71%) to collect the GHGs. Several sources of uncertainty in chamber techniques similar to the techniques applied in the wastewater sector, have been identified in GHG monitoring campaigns of running waters (Duchemin et al., 1999; Lorke et al., 2015; Matthews et al., 2003; Vachon et al., 2010). The lack of consensus on the N_2O sampling methodologies via open, closed, static or dynamic chambers is highlighted in several recent reviews summarizing the different approaches (Denmead, 2008) and the recent developments (Hensen et al., 2013) of N_2O fluxes assessment in soils and running waters.

Additionally, there is a significant variability in the chamber techniques employed; there are differences in the chamber configuration (i.e. Desloover et al., 2011; Ren et al., 2013; Rodriguez-Caballero et al., 2015; Hwang et al., 2016), chamber area and material (i.e. Aboobakar et al., 2013; Sun et al., 2015), parameters monitored in the chamber (i.e. Pan et al., 2016; Rodriguez-Caballero et al., 2014, 2015) and gaseous flux calculation methodology (i.e. Ahn et al., 2010; Wang et al., 2016).

Dynamic floating chambers and especially the Surface Emission Isolation Flux Chamber (SEIFC) (Schmidt, 1994), have been widely used in WWTPs. The SEIFC is accepted by the United States Environmental Protection Agency (USEPA) as a device for monitoring gaseous emissions from liquid surfaces and proposed by the monitoring protocol developed by Chandran, (2011). Significant variations, though, especially in the parameters monitored in the SEIFC can be identified between the different studies. For example, Wang et al., (2011) considered that the off-gas emissions from the reactor surface were equivalent to the aerators flow-rate in the aerated zone of an A/A/O. According to van Loosdrecht et al., (2016) this information can only be used as an estimation of the advective flow. This is because there is a high uncertainty in links between the flow-rate of the aerators and the distribution of air inside the reactor. According to Duchemin et al. (1999) temperature is an important parameter; temperature changes affect pressure, gas solubility and diffusivity. However, N_2O concentration adjustments due to temperature variations have also not been considered in several studies.

Several custom-made both open and closed chambers have been also widely applied. For example, Wang et al., (2016a, 2016b) integrated a fan and a pressure bag in a closed chamber.

The latter enable the mitigation of uncertainties related with unbalance headspace air and pressure variabilities that can significantly affect the water level inside the chamber and consequently the area of the chamber. Gas bags have been widely applied either standalone for the aerobic periods of biological processes (Wang et al., 2011; Sun et al., 2013; Yan et al., 2014; Sun et al., 2015) or complementing a chamber (Ren et al., 2013), especially when the monitoring was not continuous. In an effort to simulate environmental conditions and mimic natural speed Desloover et al., (2011) have used a Lidvall hood (Lindvall et al., 1974) which follows the wind tunnel approach in a two-stage nitritation-Anammox reactor. A modified forced-draught chamber was used by Hwang et al., (2016) enabling the assessment of the flow-rate and the inlet and outlet N_2O concentration variation which again aiming to sustain the uncovered environmental conditions inside the chamber (International Fertilizer Industry Association, 2001).

Even the material of the chambers varies from wood (Bollon et al., 2016b) to PVC (Mello et al., 2013). Chamber material can have a significant effect on the microclimate inside the chamber by accumulating heat from the sun and the atmosphere. Another source of potential errors in the calculation of N_2O fluxes from soils has been attributed to the height (Senevirathna et al., 2006) or the effective area (Ah/Ar) of the chamber (Rochette and Eriksen-Hamel, 2008); however, in the examined studies both parameters are highly variable ranging 0.000083% (Li et al., 2016) for the anoxic area of an orbal oxidation ditch to 0.7% (Bollon et al., 2016a) for a biofiltration system. Consequently, studies to investigating the effect of the various parameters on N_2O sampling methods from the reactor's surfaces are essential.

Delre et al. (2017) and Yoshida et al. (2014) have argued that conventional chamber-based monitoring techniques do not consider leakages (e.g. from pipes) and can monitor only limited areas of the reactors in WWTPs. Therefore, tracer gas dispersion methods for monitoring plant-wide N_2O emissions at WWTPs have been tested either as standalone techniques (Delre et al., 2018, 2017; Yoshida et al., 2014) or combined with conventional methods (Samuelsson et al., 2018). Additionally, gas collection methods from covered reactors enable the gas sampling from the whole reactor area (i.e. Kampschreur et al., 2009; Stenström et al., 2014; Castro-Barros et al., 2015; Daelman et al., 2015), increasing the accuracy of assessment.

- **Analytical techniques uncertainties**

In the majority of the discontinuous N_2O monitoring studies, the gaseous N_2O fluxes are quantified using analytical methods such as gas chromatography (GC) coupled with an electron-capture detector (ECD) (Chen et al., 2016; Ren et al., 2013; Sun et al., 2015; Yan et al., 2014). Potential interference and enhancement of the N_2O measurements under the presence of contaminants and other gases (i.e. O2, CO_2) has been reported in the literature (Crill et al., 1995; Wang et al., 2010), however, only a few studies (Mello et al., 2013; Sun et al., 2013, 2015; Brotto et al., 2015) have introduced appropriate mixtures of tracer gases (i.e. Ar with 5-10% CH_4) to mitigate such effects as proposed by the WMO-GAW (2009). Additionally the analytical precision has been reported only in the study of Brotto et al., (2015). Even though more than 40% of the monitoring studies in full-scale WWTPs have applied off-line N_2O monitoring methods, a decreasing trend over the years indicates a gradual transaction to online strategies.

Infrared laser (i.e. Bollon et al., 2016b) and Fourier transform infrared (FTIR) spectroscopic (i.e. Castro-Barros et al., 2015; Kosonen et al., 2016) are commonly used commercial analysers for the online monitoring of gaseous emissions. The accuracy of commercial analysers is about 1% for emissions ranging from 2 ppmv - 500ppmv (Pan et al., 2016) with variable detection limits (i.e. from 50 ppb for the study of Bollon et al., (2016a)) and weekly calibration is often required. Even though, in a noticeable number of studies insufficient information are provided for the analyser in terms of accuracy, detection limits, range of detectable measurements, calibration requirements (i.e. Ahn et al., 2010a; Joss et al., 2009; Castro-Barros et al., 2015) commercial gas analysers can be considered a reliable technique for the requirements of N_2O monitoring in WWTPs (Rapson and Dacres, 2014).

Dissolved N_2O is normally monitored online with a modified Clark electrode probe (Unisense A/S, N_2O-R, Denmark) to provide insights on the mechanisms of N_2O formation (i.e. Mampaey et al., 2016; Pan et al., 2016). Alternative methods that have been recently applied include the salt-stripping method (offline) (Kosse et al., 2017), the gas stripping device (online) (Mampaey et al., 2015) and the extraction of N_2O from the liquid phase with a subsequent analysis in the gaseous phase by means of photoacoustic (PA) spectroscopy (Thaler et al., 2017).

- **Uncertainties of N_2O sampling location**

The spatial variability of nitrogen species encountered in reactors with plug-flow pattern, mandate a fundamentally different approach of monitoring protocols compared to completely mixed reactors due to the spatial variability of concentrations observed. For example, Aboobakar et al., (2013), separated 8 equally distributed different zones the aerated region of the reactor and used a gas hood to monitor the emissions in each zone for 1 week whereas Rodriguez-Caballero et al., (2014) based the location of the hoods on the separately controlled air diffusers in the reactor. Overall, the spatial and temporal variability of the N_2O concentrations that have been reported in plug-flow reactors (i.e. Pan et al., 2016), have shown that the extrapolation of reactor EFs from single-spot readings is accompanied by high uncertainties (Aboobakar et al., 2013). To mitigate sampling location uncertainties, Pan et al. (2016), used a Programmable Logic Controller (PLC) to simultaneously monitor the gaseous N_2O emissions from three different chambers located in the beginning middle and end of each aerobic zone in a step-feed plug-flow reactor.

- **Uncertainties of sampling strategy**

This section focuses on the sampling method (continuous and discontinuous). The main characteristics of the sampling strategies are shown in Appendix A (Table A.3 and A.4). In the discontinuous N_2O monitoring studies, the gaseous N_2O fluxes have been grab-sampled (i.e. via gas-bags, closed chambers etc.) and subsequently quantified using analytical methods in the lab (offline monitoring) or intermittently sampled (i.e. with floating chambers for 1-2 days/month) but quantified continuously on-site (i.e. via GHG analysers) (online monitoring). In continuous monitoring campaigns N_2O fluxes have been collected continuously (i.e. via chambers) and quantified online, on-site (i.e. via GHG analysers).

Figure 2.5 illustrates the boxplot of the average EF of mainstream processes for cases of continuous gaseous monitoring using a gas analyser versus the boxplot for studies with intermittent sampling campaigns. Only medium-term and long-term studies have been considered in the analysis (Appendix A – Table A.1). Mainstream processes monitored discontinuously exhibited an average EF of 0.44% of the N-load (median EF was 0.2% of the N-load), whereas processes monitored continuously with gas analysers had an average EF equal to 1.2% of the N-load (median EF is 1.1%). The majority of the process monitored

intermittently (once or twice per month) have collected grab-samples of N_2O fluxes (Appendix A, Table A.4). Offline grab sampling is often characterised by time limitations; usually the sampling occurs during WWTP operating times and provides discrete measurements (e.g. Wang et al., 2011) that are unable to capture the whole spectrum of diurnal variabilities (Daelman et al., 2015; Wang et al., 2016a). Additionally, the temporal variability of N_2O emissions is highly dynamic (Daelman et al., 2015) and strongly affected by operational conditions. Therefore, low-frequency, long-term sampling might not capture adequately the whole range of N_2O emissions induced by short-term changes in operational conditions and pollutant concentrations. Overall, the differences can be attributed to the case-specific nature of the EF, as well as to the restrictions concerning the duration and frequency of the discontinuous campaigns (difficulty in performing grab-samples for longer periods - whole days, night time, weekends, etc.).

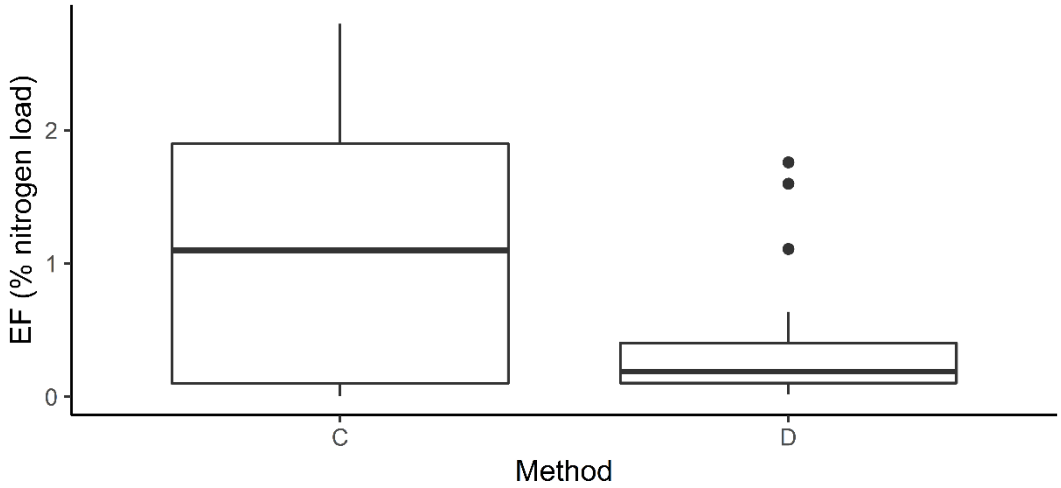


Figure 2.5: Boxplots of the average EF with respect to the method of gaseous sampling for medium-term and long-term studies (C: discontinuous monitoring, C: continuous monitoring). The rectangles represent the interquartile range. The median is represented by the black horizontal line dividing the box in two parts. The dots represent values exceeding 1.5 times the interquartile range. The upper and lower whiskers represent values higher or lower the interquartile range, respectively (within 1.5 times the interquartile range above and below the 75^{th} and 25^{th} percentile, respectively).

2.2.3 *Towards benchmarking of EFs: progress and limitations*

The amount of quantified emissions is highly affected by a variety of parameters (e.g. process type, WWTP characteristics, monitoring strategy, duration of monitoring campaign, etc.).

Therefore, estimating the N_2O EFs in WWTPs with either offline or online monitoring campaigns remains challenging. In this section, efforts have been focused on the classification of EFs for different groups of processes based on the NH_4^+ removal efficiency and the influent flow-rate. Additionally, an N_2O monitoring framework for the development of comparable EFs for the wastewater sector is also discussed.

Figure 2.6 shows the EF (with respect to the influent N-content) for mainstream processes and the achieved NH_4^+ removal efficiency (%). The different processes are represented by different colours. A detailed list of the examined studies is provided in the Appendix A (Table A.4). Triangles show that seasonal effects have been investigated in the respective process, whereas circles represent short-term studies that have not investigated seasonality. The size of the data points represents the size of the WWTP. No specific trend was identified between the observed EFs and NH_4^+ removal for the different mainstream processes. The N_2O emissions for most of the processes in smaller WWTPs (influent flow-rate<200,000 $m^{3/d}$) were less than 0.5% of the N-load, independently on the process type and nitrification efficiency. In addition, most of the processes with N_2O emissions less than 0.1% of the N-load referred to studies performing short-term and medium-term N_2O monitoring campaigns.

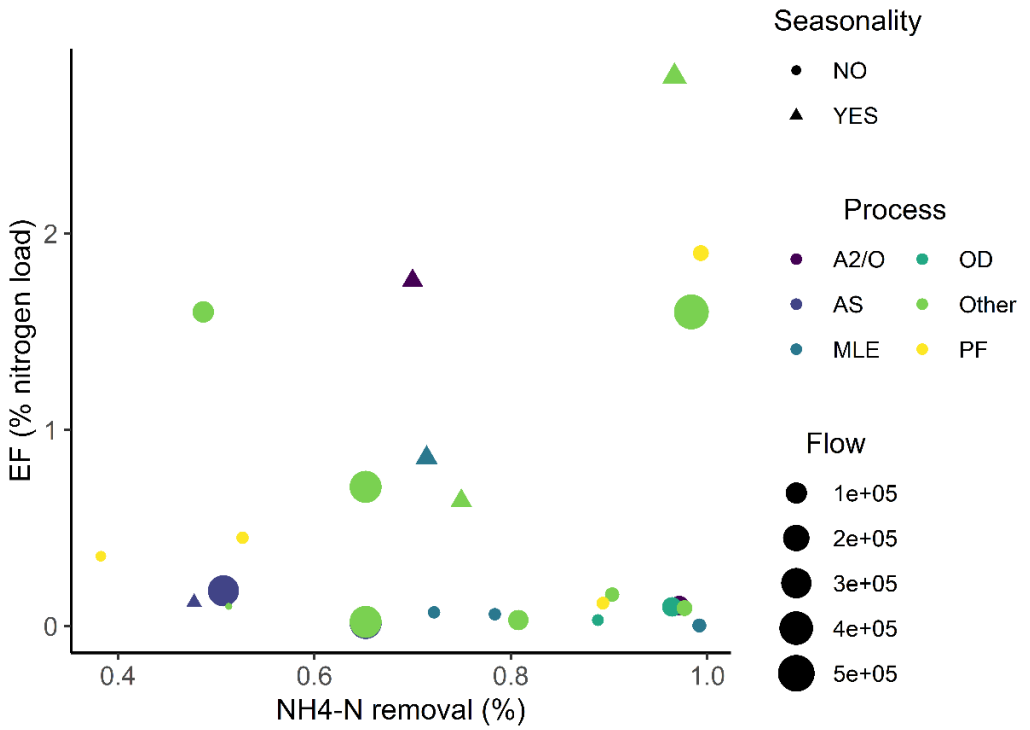


Figure 2.6: EF of the mainstream technologies with respect to the achieved NH_4^+ removal. The different colours represent different processes, whereas the different shapes show if

seasonal effects have been investigated. The size of the data points depicts the size of the WWTP in terms of influent flow-rate.

In-depth comparisons require more details on configurations, control strategies and operational conditions. Currently, there are still no specific guidelines to standardise the reporting of operational, process and monitoring strategy information from existing studies. The Appendix A summarises process-based information from past full-scale N_2O monitoring campaigns. Overall, ~70% of the mainstream studies have reported the EFs in terms of N-load. Additionally, influent and effluent NH_4^+ concentrations are available for ~35% of the systems analysed. Information on the water temperature during the monitoring campaign has not been provided for half of the studies. Limited studies provided information on the control strategy of the system (i.e. DO set-point) and other operational parameters of the processes (i.e. HRT, SRT). Given the variability of EFs amongst similar process groups (Figs. 2-6), the identification of EF patterns needs to consider process-specific operational and environmental information.

The monitoring strategies require sampling protocols that are case-specific (e.g. the choice of appropriate sampling locations). Specific protocols for the design of monitoring strategies can be found in the study of van Loosdrecht et al. (2016). Elemental mass balances can also be used to confirm the validity of the measurements (Castro-Barros et al., 2015; Mampaey et al., 2016) independently of the applied monitoring protocol. Grab-sampling the gaseous fluxes often lacks the acquisition of weekend or night-time samples, thus failing to depict the diurnal variability of the emissions. Moreover, short-term monitoring studies are frequently unable to accurately capture the temporal N_2O dynamics (Daelman et al., 2015; Kosonen et al., 2016). Daelman et al. (2013a) concluded that the accurate quantification of the average N_2O emissions requires long-term online or grab-sampling monitoring campaigns that consider the seasonal variations of temperature.

The analysis of historical process and plant data can be also useful, linking the emissions with specific and reoccurring operational and environmental conditions (i.e. dry vs wet weather, temperature) for short-term and long-term monitoring campaigns. Relationships established during short-term monitoring campaigns can be linked with the periodic operational and environmental process conditions and cannot be generalised to understand the long-term N_2O dynamics (Vasilaki et al., 2018). The latter is important due to the restrictions on the duration of the monitoring campaigns due to the entailed costs. It is also essential to identify and report

process perturbations that can affect the N_2O emissions even on a long-term basis (Vasilaki et al, 2018).

The current analysis shows that N_2O fluxes must be reported together with the configuration type, the seasonal operating conditions (e.g. pH, temperature, influent TN, effluent TN, PE, wastewater volume, COD, mixed liquor suspended solids (MLSS), SRT, HRT, recycle ratios, etc.).

2.3 N_2O monitoring campaigns and N_2O dynamics

Several methods that have been applied in N_2O monitoring campaigns can increase the understanding of the N_2O generation in full-scale processes. These include: i) techniques for the translation of WWTP operational and N_2O data into information (e.g. graphical representation of variables, feature extraction techniques, multivariate analysis, etc.), ii) mechanistic models simulating the N_2O dynamics, and iii) techniques for unveiling the relative contribution of different pathways to the emissions (e.g. isotopic analysis, real-time polymerase chain reaction (qPCR) etc.).

The effect of several parameters that are significant for N_2O generation (i.e. DO, COD/N, pH, temperature) has been extensively investigated (Kampschreur et al., 2009b; Law et al., 2012; Massara et al., 2017) based on lab-scale, pilot scale and full-scale mainstream and sidestream processes. This section aims to complement these studies; the main findings of the full-scale technologies are categorized for different process groups focusing on the techniques that have been applied and their contribution to the understanding of the behaviour of N_2O emissions.

2.3.1 *Overview of the techniques*

- **Techniques translating WWTP operational and N_2O data into information**

More than 40 physical, physico-chemical and biochemical variables (e.g. temperature, flow-rates, reduction-oxidation (redox) potential, DO, N-compounds and organic matter concentrations, alkalinity, etc.) can be monitored online to evaluate process performance (Vanrolleghem and Lee, 2003). Online monitored variables when combined with laboratory analyses can provide useful insight into the N_2O patterns and behaviour. Techniques that have been applied to translate WWTP data into information in full-scale N_2O monitoring campaigns

include: i) graphical representations and simple feature extraction methods and ii) statistical analysis and data mining methods.

In most of the studies, the online and laboratory data utilisation has been limited to the investigation and graphical representation of the significant parameters' profiles (i.e. DO, NO_2^-, NH_4^+, aeration flow-rate) in combination with the response of the N_2O emission behaviour. Additionally, descriptive statistics (i.e. central tendency, dispersion, position, etc.) of the process variables and N_2O emissions is commonly analysed and reported.

Correlation analysis and linear multivariate regression models are the main statistical techniques that have been used to reveal the N_2O emissions' dependencies with operational variables in full-scale systems (i.e. Brotto et al. 2015; Bollon et al. 2016a; Aboobakar et al. 2013). Dimensionality reduction techniques (e.g. principal component analysis (PCA), independent component analysis (ICA)), clustering (e.g. hierarchical, k-means), linear and non-linear supervised learning techniques (e.g. partial least squares (PLS), artificial neural networks (ANN)) and support vector machines (SVM)) are also powerful tools utilised to transform the WWTP data into knowledge (Haimi et al., 2013; Corominas et al., 2018). However, advanced information extraction methods have rarely been used to analyse data from N_2O monitoring campaigns. Recently, Sun et al. (2017) constructed a back-propagation ANN to simulate N_2O emissions in an anaerobic-oxic (A/O) process; thus demonstrating the feasibility and simplicity of predicting N_2O emissions with data-driven models.

- **Experimental studies in full-scale systems – modification of operational conditions**

Several N_2O monitoring campaigns in full-scale sidestream processes have tested different operational conditions to investigate their impact on the emissions (Bollon et al., 2016b; Castro-Barros et al., 2015; Mampaey et al., 2016). The majority of the studies have focused on inducing changes in the duration and flow-rate of aeration compared to the baseline control strategy of the examined reactors.

2.3.2 *Techniques unveiling the relative contribution of different pathways to the emissions*

Isotopic and molecular biology analysis are emerging techniques that can provide insights into the N_2O generation pathways. Molecular biology methods (e.g. quantitative reverse

transcription polymerase chain reaction RT-qPCR, FISH, etc.) can quantify the microbiological structure driving the N-cycle and the bacterial population able to reduce N_2O at WWTPs under various environmental and operational conditions (Castellano-Hinojosa et al., 2018; Song et al., 2014). Isotope techniques have only recently been implemented at full-scale systems to distinguish the respective contribution among the N_2O pathways and increase the understanding of the pathways that are responsible for N_2O formation (Townsend-Small et al., 2011; Tumendelger et al., 2014). A recent critical evaluation of natural abundance and labelled isotopes for N_2O studies can be found in the study of Duan et al. (2017).

- **Mechanistic models**

The mechanistic models are a popular tool for the prediction of the N_2O generation and emission during the BNR in WWTPs. Based on different assumptions, a variety of one- and multiple-pathway models have been suggested. Their structure is based either on the widely accepted ASM layout (as suggested by Henze et al. 1987, 2000), or on the more recent electron carrier concept that describes the N_2O production via the mechanism of the relevant complex oxidation-reduction reactions taking place during wastewater treatment (e.g. Ni et al., 2014). All models, though, consider the effect of changing operational parameters (e.g. DO, NO_2^- levels, aeration regime, etc.) on the N_2O generation. A comprehensive evaluation of the different modelling approaches, underlying assumptions, kinetics, stoichiometric parameters, calibration and validation procedures of several single-pathway and two-pathway AOB models, heterotrophic denitrification pathway models, and integrated N_2O models describing the three major microbiological pathways is provided by Ni and Yuan (2015). The authors have provided guidelines for the selection of modelling approach under different DO and NO_2^- concentrations based on the structural assumptions of the models. The debate on the model that best describes and decouples the major N_2O formation pathways is still ongoing with several extensions and variations of the original approaches developed recently (Ding et al., 2017; Domingo-Félez and Smets, 2016; Massara et al., 2018).

2.3.3 *Process-based insights based on the applied techniques*

Robust documentation of the dominant pathways among the different process configurations is still missing (Ma et al., 2017). This section discusses correlations between N_2O emissions and operational variables and dominant N_2O pathways that have been identified for different

full-scale process groups. Table 2.1 provides a summary of the dominant N_2O pathways that have been reported for different wastewater treatment processes based on the techniques that have been applied in the monitoring campaigns. Studies that have not discussed possible N_2O pathways have not been considered.

Table 2.1: Main findings of past studies that result in the identification of the most contributive N_2O production pathway (where possible). a: Visualization of significant profiles & descriptive analysis, b: Modified operation mode, c: Statistical analysis and data mining, d: Mechanistic model development, e: Isotopic analysis, f: real-time qPCR

Source	**Process**	NH_2OH **oxidation**	**Nitrifier denitrification**	**Heterotrophic denitrification**
Castro-Barros et al., 2015[a, b]	One-stage PNA granular Sidestream	N_2O emissions elevated during shifts from low to high aeration (NH_4^+ accumulation, high AOR) → main pathway	Not a main pathway	N_2O emissions elevated during shifts from low to high aeration (NH_4^+ accumulation, high AOR) → potential contributor
Mampaey et al., 2016[a, b]	One-stage SHARON granular Reactor Sidestream	Not a main pathway	Presence of NH_2OH in anoxic periods; lower DO resulting in increased N_2O emissions	N_2O formation during anoxic periods under the presence of NO_2^- & small amounts of organic substrate
Kampschreur et al., 2008[a]	Two-reactor partial-nitritation-anammox process Sidestream	Excluded during anoxic conditions in the nitritation reactor (despite significant N_2O formation)	• Anammox reactor: absence of O_2 • Nitritation reactor: emissions not affected by the influent composition (therefore C/N ratio)	Not a main pathway
Stenström et al., 2014[a]	Nitrification-denitrification SBR, Sidestream	Not discussed	Considerable N_2O formation under DO=0.5 mg/L & NO_2^- accumulation (>20 mg/L)	N_2O accumulation during denitrification (under conditions of low COD:N and high NO_2^-); quickly stripped off to the atmosphere as soon as aeration resumed

Wang et al., 2016b[a, c, f]	A^2/O with plug-flow pattern	Not a dominant pathway (low NH_4^+ concentrations)	• Coexistence of NO_2^-, NH_4^+ & O_2-limiting conditions • Correlation between NO_2^- and N_2O emissions • Strong responses between NO and N_2O emissions and the relative abundance of AOB.	Not a dominant pathway (no peaks observed after anoxic zones)
Wang et al., 2011[a]	A^2/O	Not a main pathway	• Rapidly increased N_2O emissions due to DO limitation (DO<2.5 mg/L); maximum N_2O emission at DO=0.75 mg/L • Increase in NO_2^- concentration (from 0.2 to 0.6 mg/L) during nitrification leading to increase in N_2O fluxes	Not a main pathway
Toyoda et al., 2011[a, f]	A^2/O	SP Isotopic analysis: • ~50% contribution in the beginning of aerobic tank	SP Isotopic analysis: • ~50% contribution in the beginning of aerobic tank • Dominant pathway from middle to the end of aerobic tank	N_2O was produced during denitrification
Aboobakar et al., 2013[a, c]	A/O plug-flow reactor	Not dominant pathway	Considered dominant in zones with DO<1.5 mg/L	Considered dominant in zones with depleted NH_4^+, DO fluctuations, NO_3^- availability & lack of NO_2^-
Sun et al., 2017[a]	A/O	Higher DO: certain NO_2^- amount potentially utilised to	Low-DO condition (i.e. <1 mg/L) usually observed at the beginning of the oxic zone	Not a main pathway

		oxidize NH_3 to NH_2OH, thus leading to the N_2O production		
Kosonen et al., 2016[a, c] & Blomberg et al., 2018[d]	A/O bioreactor	Only this AOB pathway modelled due to the existing DO & NO_2^- conditions, N_2O production mainly in the aerated zones, N_2O consumption in the anoxic zones→ main pathway	• Increasing the number of nitrifying zones resulting in higher overall N_2O emissions (N_2O production possibly via nitrifier denitrification) • Given that anoxic-aerobic volume controlled by the NH_4^+ concentration, unclear if increased emissions caused by increased NH_4^+ concentration or increased number of nitrifying zones	Not main pathway
Pan et al., 2016 & Ni et al. 2015 [a, d]	2 step-feeding, anoxic/oxic/anoxic/oxic plug-flow reactor	N_2O emissions increasing with the AOR increase (2nd step of the plug-flow reactor)	N_2O emissions increasing with the AOR increase (2nd step of the plug-flow reactor)	Not a main pathway
Rodriguez-Caballero et al., 2014[a, b]	Anoxic/oxic/short anoxic/oxic plug-flow reactor	Not a main pathway	• N_2O peaks when transitioning from anoxic to aerobic conditions; O_2 limitation considered as enhancing the activation of the nitrifier denitrification pathway • N_2O emissions increasing with potential shock loads; the AOB likely to activate their denitrification pathway after shock loads of toxic compounds	Not a main pathway

Castellano-Hinojosa et al., 2018[c, e]	Two sequential bioreactors (anoxic and aerated)	Not a main pathway	• Strong positive correlation between AOB abundance & N_2O emission; hence, more possible pathway under anoxic conditions • 0.5<DO<1 mg/L: enough O_2 provided to the AOB for the oxygenation of NH_3 to NH_2OH but not for aerobic respiration; NO_2^- potentially used as alternative electron acceptor to complete nitrification	Not a main pathway
Tumendelger et al., 2014[a, f]	CAS	SP Isotopic analysis: • Up to 90% contribution at DO ~2.5 mg/L • ~50% contribution at DO ~1.5 mg/L	SP Isotopic analysis: • Dominated at DO <1.5 mg/L • ~50% contribution at DO ~1.5 mg/L	Not discussed
Daelman et al., 2015[a, c]	Carrousel reactor	Carrousel: emissions coinciding with aerated periods (AOR governed by DO); the relationship between the AOR & the N_2O production usually explained by referring to the NH_2OH pathway; however, not considered dominant	• Carrousel: emissions correlated with the NO_2^- concentration peaks • Prevalence of low-DO zones	Carrousel: reactor lacking sufficient anoxic space to allow the completion of denitrification
Ni et al., 2013[d]	OD with surface aerators	Main pathway since high NH_4^+ concentrations were observed without	Not a main pathway	Not a main pathway

		simultaneous NO_2^- increase in the aerated zones/phases		
Ni et al., 2013[d]	Feeding and aeration (90 min)/ settling (35 min) / decanting (55 min) SBR	Main pathway since high NH_4^+ concentrations were observed without simultaneous NO_2^- increase in the aerated zones/phases	Not a main pathway	Not a main pathway
Sun et al., 2013[a, c]	Feeding (synchronous aeration)/aeration/settling/decanting SBR (1 hour each)	Not a main pathway	• Low DO during nitrification significantly affecting the N_2O production → main pathway	Correlation between N_2O emission and influent COD/N → contributor
Rodriguez-Caballero et al., 2015[a, b]	Reaction phase (~130 min) /settling (~65min) and decanting (~65 min) SBR (anoxic/aerobic alternations – 3 cycle types)	Not a main pathway	• Certain NO_2^- accumulation under aerobic conditions • N_2O generation continuing after aeration stop	N_2O generation continuing after aeration stop
Wang et al., 2016a [a, c]	Full-scale biological aerated filter (BAF) for secondary nitrification	Low influent NH_3 concentration (<6 mg/L) → not a main pathway	• Significant linear correlation between N_2O & NO EFs in different seasons • Nitrifier denitrification suggested as possible pathway in accordance with the fact that the influent NO_2^- found as key factor regarding the N_2O & NO production	• Significant linear correlation between N_2O & NO EFs in different seasons • Minor possibility of heterotrophic denitrification contribution during the denitrification of NO to N_2O

Bollon et al., 2016[a]	Nitrifying biofiltration (Biostyr® filters)	Not discussed	Not discussed	• Rapid increase in the net N_2O production rate at BOD:N< 3 • Intensity increased with the duration of carbon-limiting conditions

Overall, the majority of the studies investigating N_2O dominant pathways in mainstream full-scale systems with descriptive statistics and visual inspection (i.e. via univariate/bivariate graphs) of significant variables (e.g. NH_4^+, DO concentrations, influent flow-rate and N_2O emissions) have not considered NH_2OH to be a significant pathway for N_2O generation, regardless the configuration (Table 2.1).

Specifically, in 1-step feed A/O and A^2/O configurations and processes with anoxic/aerobic alternations and plug-flow pattern, studies have observed i) spatial N_2O peaks in the transitions from anoxic to aerobic zones (i.e. Rodriguez-Caballero et al., 2014; Sun et al., 2017) under low DO concentrations (<1 mg/L), ii) temporal increase in N_2O emissions that coincide with elevation of NO_2^- and NH_4^+ concentrations (i.e. peak loads) under O_2-limiting conditions (J. Wang et al., 2011; Wang et al., 2016b) and iii) N_2O emission peaks coinciding with elevated NO_2^-concentrations in aerobic zones (Sun et al., 2017). Therefore, the nitrifier denitrification pathway has been suggested to be dominant. This is also supported by Wang et al. (2016b) who studied the relative abundancy of the AOB and the denitrifying bacteria under different seasonal conditions for an A^2/O process with a plug-flow pattern (DO 0.6 – 6.8 mg/L and N-concentration up to 30 mg/L) to provide insights on the N_2O generation pathways. The authors quantified the expression of functional genes harboring the NH_3 monooxygenase (*amoA*) (i.e. enzyme catalyzing the first step of nitrification) for the AOB, as well as of the *nosZ* harboring the N_2O reductase (i.e. enzyme catalyzing the reduction of N_2O to N_2) for the denitrifiers by RT-qPCR. Wang et al. (2016b), applied also correlation analysis; the N_2O emissions (ranging from 0.01 to 3.4% of the TN-load) were mainly dependent both on the NO_2^- concentrations and the relative AOB abundances, hence indicating that nitrifier-denitrification was the dominant pathway in the aerobic zones of the reactor. On the contrary, the obtained results revealed that the emissions were not affected by the relative abundancies of the denitrifiers. However, it must be noted that the functional gene levels are not always representative of the activity of the corresponding enzyme (Henderson et al. 2010). Similarly, Castellano-Hinojosa et al. (2018) quantified the 16SrRNA, *amoA* and *nosZ* genes of the total bacterial and archaeal population in three full-scale predenitrification-nitrification systems. They employed multivariate analysis (i.e. non-metric multidimensional scaling (MDS) and similarity analysis based on Euclidean distance, to identify the environmental variables that are best linked to the patterns of community structure via BIO-ENV procedure) to link the bacterial structure with the N_2O emissions and environmental/operational variables. They found a strong positive correlation between the AOB and the emissions in the anoxic compartments, where the N_2O release was

higher. Therefore, the authors suggested nitrifier denitrification as the dominant pathway. On the other hand, the emissions were negatively correlated with the AOA abundance and the N_2O reducers. It was concluded that the elevated NO_2^- concentrations, the low temperatures and short SRTs mainly influenced the abundance of the bacterial community encoding the *nosZ* gene and contributed to the N_2O accumulation.

NH_2OH oxidation has not been considered as a dominant N_2O pathway in the majority of A^2/O and A/O process groups. However, Toyoda et al. (2011), applied site-preference (SP) isotopic analysis in a A^2/O configuration and found that NH_2OH oxidation and nitrifier denitrification pathways contributed almost equally to the N_2O formation in the beginning of the aerobic tank. The DO concentrations in the reactor, though, were not reported. Additionally, Blomberg et al. (2018) developed an ASM3-type NH_2OH-heterotrophic denitrification N_2O model with a K_La-based approach for N_2O stripping (K_La_{N2O}: mass transfer coefficient for N_2O). The model was developed for the full-scale underground WWTP of Viikinmäki (Kosonen et al., 2016) that is divided into six zones (i.e. one anoxic pre-denitrifying zone, two alternating switch zones, three aerated nitrifying zones). High DO concentrations in the aerated zones (i.e. 1.5-3.8 mg/L) and low NO_2^- concentrations (i.e. 0.1 and 0.7 mg/L), were observed and the model adequately fitted the observed dissolved N_2O profiles. However, under the applied stripping modelling, the model overestimated the EF; showing that the stripping modelling approach must be improved.

Ni et al. (2015) considered all N_2O production pathways in a two-step plug-flow type reactor (anoxic/aerobic/anoxic/aerobic), in an attempt to explain the difference between the EFs of each step (1^{st} step: 0.7% of influent-N, 2^{nd} step: 3.5% of influent-N) using real data obtained from the study of Pan et al. (2016), for a full-scale step-feed plug-flow reactor. The N_2O production was mainly attributed to the heterotrophic denitrification taking place in the anoxic zone of the 2^{nd} step that was receiving 70% less biomass compared to the 1^{st} step (Table 2.1). This model has been successfully applied for the explanation of the observed EF difference and identification of dominant pathways since it (i) includes all the possible production pathways, (ii) has considered the design and operating features of the WWTP, and (iii) was calibrated/validated using data from the plant operation.

It must be noted that, in an anoxic-aerobic plug-flow reactor, the application of zone-based stepwise multiple regression showed that the effect of the N-load, DO and temperature in N_2O emissions varied within the reactor (Abookabar, 2013). Therefore, the dominant pathways can

potentially vary in the aforementioned studies based on the location of the sampling between the different studies.

In OD reactors N_2O fluxes have been mainly linked with i) stripping of dissolved N_2O that is generated in anoxic zones (Sun et al., 2015; Yan et al., 2014) and ii) NO_2^- accumulation in low DO zones, indicating nitrifier denitrification and heterotrophic denitrification (Daelman et al., 2015) as dominant pathways. For instance, in an OD reactor, strong positive correlation (Pearson's coefficient) was identified between daily N_2O emissions and daily NO_2^- peaks (0.7) (487 days monitoring campaign) (Daelman et al., 2015). The authors proposed to use NO_2^- peaks as a diagnostic method for the prediction of N_2O peaks. In the same system Vasilaki et al. (2018) applied changepoint detection techniques combined with hierarchical k-means clustering and PCA, to reveal the N_2O emission patterns and generation pathways and identified changes in the N_2O fluxes. The study concluded that the N_2O dependencies with other operational variables (i.e. NH_4^+, NO_3^-, DO) are dynamic and affected by the seasonal variations. The preferred N_2O pathways were also found to be dependent on time and operational conditions.

Additionally, a full-scale OD was modelled by Ni et al. (2013) considering NH_2OH oxidation pathway (via the electron carriers approach and assuming the NH_2OH/NO model and no inhibition of AOB NO reduction by DO) and heterotrophic denitrification pathway (based on Hiatt and Grady (2008) and electron competition between denitrification steps). Even though, the operational control of the OD is not explicitly described (i.e. aeration/DO set points), more than 90% of the N_2O emissions were observed in aerated zones with DO > 2 mg/L. The model was calibrated using 3-days data from an intensive monitoring campaign and validated based on 1-day data with different influent conditions. The developed model linked the higher N_2O generation with NH_4-N concentration peaks (up to ~9 mg/L in the calibration and ~4 mg/L in the validation dataset) within the aerated zones (OD) and the NH_2OH oxidation pathway. However, in the study of Ni et al. (2013), the respective contribution among the two AOB pathways was not explored whereas short-term data were used to validate the model. The AOB denitrification model developed by Mampaey et al., (2013) was applied using the same dataset from the OD, for calibration and validation (Spérandio et al., 2016). The model could adequately follow the trends of N_2O behavior after calibration of the anoxic reduction function (high value of 0.63 was required). NO_2^- accumulation and the resulting nitrifier dentification

contribution to the emissions cannot be excluded in full-scale WWTPs. Future enhanced versions of models must consider this fact.

Overall, as shown in section 2.2, N_2O fluxes in CAS systems are generally low. Tumendelger et al. (2014) applied SP isotopic analysis and observed that the NH_2OH oxidation pathway was responsible for up to 90% of the N_2O formation under high DO (~2.5 mg/L at the middle and end of the aerobic tank) in a conventional AS system (>44.2% of NH_4^+ was nitrified and removed in gaseous form probably due to unintentional zones with low DO). Nitrifier denitrification and NH_2OH oxidation were almost equally contributing to DO levels around 1.5 mg L^1, whereas nitrifier denitrification dominated at DOs below 1.5 mg/L.

In sidestream reactors, elevated N_2O emissions during shifts from low to high aeration (NH_4^+ accumulation, high AOR), have been attributed to NH_2OH pathway (Castro-Barros et al., 2015). Aeration intensity and profiles have been determined as significant control parameters for the N_2O generation (Harris et al., 2015; Rathnayake et al., 2015). The N_2O dynamics under different aeration intensities is likely to depend on the reactor configuration. For example, Mampaey et al. (2016) and Stenström et al. (2014) observed higher emissions when lower DO was applied in a partial nitritation-anammox system and a sidestream nitrification-denitrification SBR, respectively. On the other hand, (Kampschreur et al., 2009a) could not identify a relationship between the N_2O increase and the higher aeration flowrate during a prolonged aeration experiment in a single-stage nitritation-anammox reactor. Hence, the influence of the aeration regime on the N_2O generation is variable, depending on the reactor configuration. In Anammox reactors, N_2O formation during anoxic periods has been mainly attributed to the nitrifier denitrification pathway and partially to heterotrophic denitrification (e.g. under conditions of limited substrate provision; Castro-Barros et al., 2015; Mampaey et al., 2016). However, a recent study performed by Ma et al. (2017) demonstrated that N_2O formation via NH_2OH oxidation can also occur at low DO (~1 mg/L) probably catalysed by cytochrome P460. This finding contradicts previous experimental and model-based works according to which the NH_2OH oxidation pathway dominates solely at higher DO concentrations (e.g. Brotto et al., 2015) and is linked with the AOR (Peng et al., 2014).

Studies investigating dominant N_2O pathways for several groups of full-scale processes are still missing (Table 2.1) and further research is required for a robust mapping of the dominant pathways in different process groups. As a general remark, in most processes with elevated N_2O emissions (>0.85% of the N-load), independently of the configuration, elevated NO_2^-

concentrations were also observed (Daelman et al., 2015; Foley et al., 2010; Rodriguez-Caballero et al., 2015; Wang et al., 2016b)

2.3.4 *Limitations and future research*

Several authors have underlined the difficulties in determining the respective contribution of each N_2O generation pathway during full-scale monitoring campaigns (Aboobakar et al., 2013; Wang et al., 2016b). None of the techniques analysed, in the previous sections can be applied standalone to explain the ambiguities surrounding the mechanisms and operational conditions that enhance the N_2O formation during wastewater treatment. This section, summarises the main limitations of the techniques applied to explain N_2O fluxes behavior in wastewater systems and describes how these techniques combined, can maximise the outcome of future monitoring campaigns.

Two major drawbacks have been identified when studies rely exclusively on simple descriptive statistics and graphical representations of operational variables to extract insights on N_2O emissions behaviour. Firstly, this approach, considers independently several parameters that affect the N_2O generation. Thus, it becomes difficult to quantify the combined effect of several variables in full-scale systems via simple univariate or bi-variate graphical representations. For instance, Castro-Barros et al. (2015) observed higher N_2O emissions and formation in the transition from the anoxic to the aerated periods in a one-stage granular partial nitritation-anammox reactor. However, this increase cannot be solely attributed to the DO change, since the AOR, the NH_4^+ and the NO_2^- concentrations were also elevated during the transition. Secondly, graphs representing the behaviour of process variables in relation to N_2O emissions usually cover only a limited period (often shorter than the monitoring campaign duration) or visualize average data in the majority of the reported studies. There are limitations in the effective visualization and dependencies extraction in long-term temporal multivariate datasets (i.e. overcrowded and cluttered visualization) (Shurkhovetskyy et al., 2018). For example, Aboobakar et al. (2013) monitored the N_2O emissions of a PF reactor for 56 days; they reported the diurnal profile of the normalised average emissions, the average ammonia diurnal profile and the average daily DO concentration, over the duration of the monitoring campaign.

The N_2O emission behaviour is characterised by temporal variations. Additionally, for a specific process, the dependencies of the emissions with operational variables are also expected to fluctuate under different environmental/operational conditions based on the preferred N_2O

production pathway. Vasilaki et al. (2018) showed that the dependencies between N_2O emissions and operational variables fluctuated in a Carrousel reactor that was monitored for 15 months. Therefore, employment of advanced visualization techniques capturing the dynamic behaviour of the operational variables from the whole duration of the monitoring campaigns (e.g. data abstraction, principal component-based analysis, clustering - Shurkhovetskyy et al., 2018; Aigner et al., 2008) can facilitate an accurate and deep understanding of the long-term N_2O behaviour in both sidestream and mainstream treatment processes. There is an interchangeable link among operational/environmental conditions, N_2O production pathways and emission rates. The combination of data mining methods can also be applied to identify distinct and different patterns and operational conditions in order to determine the contribution among the N_2O production pathways. It can reveal the links and relationships among the conditions triggering the N_2O emissions, the dominant pathways and the emission rates. However, few data-driven monitoring and control approaches have been validated in full-scale applications in the wastewater sector (Haimi et al., 2013). Additionally, there is still little guidance for the selection of the most appropriate techniques for particular wastewater applications (Hadjimichael et al., 2016). Multivariate statistical analysis techniques have only recently been applied to translate data from the monitoring campaigns into useful information regarding the N_2O production (Vasilaki et al., 2018). Hence, structured approaches and data-driven extraction techniques need to be developed to process the incoming data from WWTPs (Corominas et al., 2018) and acquire information concerning the N_2O emission patterns.

Given that previously unreported pathways (Harris et al., 2015), or alternative conditions under which the already known pathways are activated have been recently identified in literature, isotopic and molecular biology analysis can be applied to support insights into the N_2O generation pathways. In a recent review on the isotopic methods for the identification of the respective contributions of the different N_2O pathways, Duan et al. (2017) concluded, though, that there are still uncertainties regarding the accuracy of the SP methods (i.e. not standard SP signature values, unknown N_2O production pathways, etc.). Therefore, the authors suggest to complement the isotopic methods with other approaches, such as the mRNA-based transcription analysis (Ishii et al., 2014).

Mechanistic models considering all the possible N_2O production pathways are powerful tools to describe the operation of full-scale WWTPs, N_2O emissions generation pathways and guide towards mitigation measures. However, there are still, several challenges in the practical

application, calibration and validation of mechanistic N_2O models in full-scale systems. Parameter uncertainty still plays a significant role in explicitly differentiating the contribution of different N_2O pathways via modelling studies; for instance, different AOB pathway models (i.e. Spérandio et al., 2016) and models with different contributions of denitrification N_2O-producing pathways (i.e. Domingo-Félez et al., 2017) have been adequately fitted to describe N_2O emissions behaviour in the same systems. Inclusion of all major N_2O production pathways results in complex and overparameterized models impairing reliable calibration and validation. Additionally, short-term calibration and validation of models under specific operational conditions (i.e. dry weather) limits their accuracy when the system varies significantly (Guo and Vanrolleghem, 2014).

Quantifying the contribution of the N_2O production pathways in addition to the triggering mechanisms in biological processes remains a challenge and still requires extensive research (Guo et al., 2017). As a general remark, standardisation of monitoring and reporting, long-term N_2O monitoring campaigns along with combined multivariate analyses of the provided data and mechanistic model development are required to increase the understanding and effectively control the N_2O emissions at WWTPs.

2.4 Monitoring campaigns and mitigation strategies

Mitigation measures have been developed and proposed mainly as outcome of studies that targeted at the: i) testing of different aeration/feeding control strategies in full-scale sidestream technologies (e.g. Castro-Barros et al., 2015; Mampaey et al., 2016; Rodriguez-Caballero et al., 2015), ii) development of mechanistic models simulating changing operational conditions (e.g. Ni et al., 2015), and iii) establishment of non-linear regression models (e.g. ANNs) to predict the behaviour of N_2O emissions (e.g. Sun et al., 2017).

2.4.1 *Mitigation measures and full-scale monitoring campaigns*

Table 2.2 summarises the main N_2O mitigation strategies that have been proposed for full-scale systems along with the methodological approaches that facilitated the identification of these measures. As shown in Table 2.2, there is no standardised methodology for the establishment of N_2O mitigation strategies in full-scale systems.

Several studies have modified the aeration intensity, DO and cycle duration to investigate the effect on N_2O emissions within full-scale systems (Castro-Barros et al., 2015; Kampschreur et al., 2009a; Mampaey et al., 2016; Rodriguez-Caballero et al., 2015). For instance, Mampaey et al. (2016) achieved a reduction in the N_2O emissions by 56% when the cycles in a one-stage granular SHARON reactor were shortened by 1 h. Rodriguez-Caballero et al. (2015) tested different operational conditions in a full-scale SBR. They have suggested an optimum control strategy for the minimisation of N_2O emissions based on the application of short aerobic-anoxic cycles (20-min aerobic phase and short duration of anoxic stage). Therefore, testing different operational modes is regarded as one of the most effective ways to identify measures for the emission mitigation (Table 2.2).

Ni et al. (2015) developed a mechanistic model utilizing the data from a two-step PF reactor (Pan et al., 2016) showing that the biomass specific N-loading rate was responsible for the elevated N_2O emissions observed in the 2nd step of the process. Different operational conditions were tested with the model demonstrating that lower N_2O emissions (<1% of the N-load) can be achieved if 30% of the total return activated sludge (RAS) stream is recirculated to the 2nd step of the PF reactor (Table 2.2). However, it is unknown whether the suggested mitigation strategy was actually demonstrated in the system. Ahn et al. (2010b) identified dependencies between the WWTP operating conditions and the N_2O emissions via multiple linear regression. According to their findings, intermittent aeration or over-aeration must be avoided in aerobic reactors. Castellano-Hinojosa et al. (2018) linked the population of AOB, AOA and N_2O-reducers with the changes in the operational and environmental variables in four conventional AS systems with pre-denitrification zones. They observed that N_2O emissions mainly occurred due to incomplete denitrification, thus underlining the importance of ensuring the completion of the process for the emissions mitigation. Overall, the main techniques for mitigating the N_2O emissions in wastewater treatment processes include: i) the application of the optimal aeration intensity and DO concentration, ii) preventing NH_4^+ concentration peaks (e.g. via equalization tanks), iii) the avoidance of NO_2^- accumulation through proper control, and iv) the supply of additional carbon source (when required) to ensure complete denitrification in the anoxic reactors (Table 2.2).

However, studies applying and evaluating mitigation measures for long-term applications are still missing. Additionally, there is a gap between the data coming from the monitoring campaigns and their processing in order to establish a mitigation strategy. Moreover, several

monitoring campaigns do not conclude on the development of mitigation strategies (Filali et al., 2013; Stenström et al., 2014; Yan et al., 2014). The long-term implementation and evaluation of the proposed mitigation strategies is still an issue. Therefore, the establishment of standardised methodological approaches for the identification of N_2O mitigation strategies is required.

Table 2.2: Methods and main findings of studies resulting in mitigation measures.

Source	Process	Method	Main findings	Mitigation Measures
Castro-Barros et al., 2015	One-stage PNA granular Sidestream	• Calculation of dissolved N_2O based on Mampaey et al. (2016) • Modified operation mode: Prolonged anoxic & aeration periods • Visualization of significant profiles (i.e. DO) & descriptive analysis	• Smoother aeration transitions during normal reactor operation connected with lower N_2O emissions; comparison with experiments • Prolonged anoxic periods leading to increased N_2O emissions	• Optimise the aeration regime • Ensure smooth shifts in the aeration pattern • Optimise for short aeration intervals
Mampaey et al., 2016	One-stage SHARON granular Reactor Sidestream	• Modified operation mode: prolonged anoxic & aeration periods, lower DO experiments, shorter SBR cycles • Calculation of dissolved N_2O based on Mampaey et al. (2016) • Visualization of significant profiles (i.e. DO) & descriptive analysis	• Nitritation reactor: N_2O formation higher during anoxic periods • Splitting the anoxic period: average anoxic N_2O formation rate decreased • Shorter cycles reducing the N_2O EF by 56% at the expense of higher NO_3^- concentrations	• Preferably operate under shorter cycles • Apply continuous aeration in nitritation reactor; this requiring optimisation • Alternatively operate under lower DO setpoint
Kampschreur et al., 2009	One-stage PNA granular	• Modify operation: varying aeration rate • Visualization of significant profiles & descriptive analysis	• Over-aeration significantly impacting on N_2O emissions	• Ensure sufficient aeration control

Kampschreur et al., 2008	Two-reactor partial-nitritation-anammox process Sidestream	• Visualization of significant profiles (N-compounds) & descriptive analysis	• Nitritation reactor: N_2O accumulation during the non-aerated phase • Anammox reactor: NO_2^- accumulation potentially increasing N_2O emissions	• Avoid anoxic phases in nitritation reactor (i.e. smaller reactors to ensure sufficient HRT) • Control the aeration in the nitritation reactor • Operate a one-reactor nitritation-anammox system; potentially emitting less N_2O due to limited NO_2^-accumulation
Ahn et al., 2010	Multiple processes (i.e. MLE, step-feed BNR, OD)	• Multiple linear regression for several processes	• Investigate possible links between WWTP operating conditions & N_2O emission fluxes • Aerobic zones: N_2O fluxes correlated with location-specific pH, AS mixed liquor temperature, DO, NH_4^+ & NO_2^- concentrations & interactive combinations • Anoxic zones: N_2O fluxes correlated with location-specific sCOD, pH, AS mixed-liquor temperature, DO, NO_2^- & NO_3^- concentrations & interactive combinations	• BNR processes: Avoid high NH_4^+ & NO_2^- concentrations, DO & transients • Aerobic processes: avoid incomplete/intermittent nitrification & over-aeration • Rely on more uniform spatial DO profiles to promote SND • Minimise peak N-flow (flow equalisation)
Ni et al., 2013	OD with surface aerators & Feeding and aeration (90 min)/ settling (35 min) /	• Mechanistic model development	• Modelling of two full-scale municipal WWTPs (i.e. an OD & an SBR) • OD: decrease in the NH_4^+ concentration without simultaneous NO_2^- increase in the aerated zones	• Use the developed model to accurately simulate the emissions from the surface aerator zone in OD WWTPs, thus potentially correcting the N_2O emission underestimation in full-scale WWTPs where the floating chamber method is not valid

	decanting (55 min) SBR		• SBR: NH_4^+ accumulation leading to a high AOR during the aerobic SBR phases &, finally, to the increased production of intermediates (e.g. NH_2OH)	
Li et al., 2016	Reversed A^2/O and OD	• Observations & literature	• N_2O generated & emitted more in summer than in winter • Microbial population & aeration strategy as key factors of N_2O generation & emission	• Avoid incomplete/intermittent nitrification & over-aeration during the aerobic processes to achieve lower N_2O emissions • Apply uniform spatial DO profiles to promote SND that probably leads to less N_2O emissions • Perform flow equalization to control the peaking factor of the influent N-loading to the AS • Ensure a sufficiently long SRT to prevent NO_2^- accumulation during nitrification • Avoid the COD limitation of the denitrification process by minimizing the pre-sedimentation of organic carbon in the influent & dosing additional organic carbon
Pan et al., 2016 & Ni et al., 2015	2 step-feeding, anoxic/aerobic/anoxic/aerobic plug-flow reactor	• Mechanistic model development	• Step-feeding resulting in incomplete denitrification & affecting the AOR in nitrification, hence increasing the total N_2O emission	• Decrease the N_2O EF to the lowest value of <1% if 30% of the total RAS returns to the 2^{nd} step

Wang et al., 2016b	A^2/O with plug-flow pattern	• Investigation of AOB abundances	• N_2O emitted mainly from the oxic zone, with the emitting levels increasing greatly from the beginning of the oxic zone towards the zone end • NO_2^- accumulation directly triggering N_2O production • Both diurnal & seasonal N_2O emission levels fluctuating strongly • Other factors influencing the N_2O emission: low DO/temperature	• Increase DO availability for both AOB & NOB • Improve the AOB living conditions • Apply a step-stage aeration mode with varying aeration intensities (location-specific emission patterns for a plug-flow process) • Ensure a better mixing via a higher horizontal flow-rate combined with an appropriate vertical airflow flux; the large cross-section widths reduced using partition walls to elevate flow velocities under a constant A^2/O tank working volume
Sun et al., 2013	Feeding (synchronous aeration)/aeration/settling/decanting (1 hour each) SBR	• Visualization of significant profiles (DO & N_2O) & descriptive analysis • Multiple linear regression analysis to investigate relationship of N_2O emissions & environmental factors • Bimonthly sampling to examine changes in the relationship between N_2O emissions &environmental factors (long-term: 12 months)	• N_2O flux from different treatment units/periods following a descending order: feeding period, aeration period, settling period, swirl grit tank, decanting period & wastewater distribution tank • Feeding & aeration periods accounting for >99% of N_2O emissions • Low DO during nitrification majorly influencing N_2O production	• Increase the aeration rate during the feeding period & decrease it to a proper level for nitrification in the aerobic stage • Supply external carbon source during denitrification/change the operational SBR mode (from feeding under synchronous aeration to feeding with anoxic stirring) to ensure enough COD provision/better utilisation of influent COD for denitrification

Rodriguez-Caballero et al., 2015	Reaction phase (~130 min) /settling (~65min) and decanting (~65 min) SBR (anoxic/aerobic alternations – 3 cycle types) SBR	• During the experimental campaign, 3 different cycle configurations implemented as part of the normal SBR operation • Testing of a modified cycle configuration for the N_2O mitigation • Observe dissolved & gaseous N_2O profiles vs time	• N_2O emissions accounting for >60% of the total carbon footprint of the WWTP • Cycles with long aerated phases showing the largest N_2O emissions, with a consequent increase in the carbon footprint • Transient NH_4^+ & NO_2^- concentrations & transition from anoxic to aerobic possibly involved in the increased N_2O production	• Apply intermittent aeration to reduce the NO_2^- accumulation • Adopt a cycle configuration with short aerated periods • Allow the system to consume N_2O through denitrification
Spinelli et al., 2018	MLE	• Event-based sensitivity analysis • Box-plots of diurnal behaviour of significant variables & N_2O	• Lower COD:N resulting in higher N_2O emissions due to disturbed denitrification • Daily N_2O peaks occurring under conditions of higher aeration flow-rate (more intense stripping)	• Equalization of the influent flow-rate
Townsend-Small et al., 2011	MLE	• Isotopic composition of N_2O	• Both nitrification & denitrification contributing to the N_2O emissions within the same WWTP • BNR significantly increasing urban N_2O emissions	• Apply engineering processes for the selection of bacteria capable of reducing NO_3^- without releasing significant N_2O amounts
Castellano-Hinojosa et al., 2018	Two sequential bioreactors (anoxic and oxic)	• Simultaneously link the abundance of AOB, AOA & N_2O-reducers with the changes of the operational/environmental variables	• N_2O emissions strongly correlated with increased abundances of AOB & lower counts of N_2O-reducers • Unlikely significant contribution of AOA to N_2O generation since their abundance correlated negatively to N_2O emissions	• Avoid NO_2^- accumulation, low temperatures & excess DO in the anoxic bioreactors to enable complete heterotrophic denitrification & hinder nitrifier denitrification

			• AOB abundance favoured by higher NO_3^- & NO_2^-concentrations in the AS	
Sun et al., 2017	Biological tank with an anoxic & an oxic zone A/O	• Construction & performance evaluation of BP-ANN model	• DO having a significant influence on the N_2O production • BP-ANN model suitable for the prediction of N_2O emissions in other WWTPs with different configurations (e.g. A^2/O, SBR & nitrification-anammox), if influent/environmental parameters & N_2O emission data can be investigated through full-, pilot- or lab-scale experiments	• Apply proper control of DO during both nitrification & denitrification • Apply the BP-ANN model as a convenient & effective method for the prediction of N_2O emissions in an A/O WWTP
Blomberg et al., 2018 based on Kosonen et al., 2016	A/O bioreactor	• Mechanistic model development	• Model describing the full-scale underground WWTP of Viikinmäki • AOB pathways: only NH_2OH oxidation included due to the dynamic & relatively high DO concentrations (1.5-3.8 mg/L) in the aerated zones & the low NO_2^- concentrations (0.1 & 0.7 mg/L) • N_2O production mainly in the aerated zones, minor N_2O consumption & minor stripping effect in the anoxic zones • Applied stripping model: EF overestimation	• Improve the stripping modelling approach • Consider the nitrifier denitrification contribution in future model versions
Chen et al., 2016	CAS	• Fugacity model • Lab-scale in situ experiments	• Compared to other parameters (e.g. sludge concentration/retention	• Reduce the aeration rate

			time), the adjustment of the aeration rate effectively mitigated the GHG emission in the AS without significantly affecting the treated water quality • N_2O as main contributor to the total GHG emission (i.e. 57-91% of total GHG emission) • Lowering the aeration rate in the AS by 75% enabled decreasing the mass flux of N_2O by up to 53% • Most important benefit of changing the aeration rate: lower energy consumption during the WWTP operation (fractional contribution of pumping to the total emission from the WWTP=46-93% within the range of the aeration rate tested)	
Ribeiro et al., 2017	Extended aeration CAS	• Different aeration rates tested	• Nitrification as the main driving force behind N_2O emission peaks • Air flow-rate variations possibly influencing the N_2O emissions; high N_2O emissions under conditions of over-aeration or incomplete nitrification along with NO_2^- accumulation	• Add an anoxic zone & recirculation to a non-BNR system for nitrification; otherwise, high N_2O emissions expected in case of increased DO • Control the DO; dynamic changes in DO concentrations reported as being responsible for N_2O emission peaks in SND BNR systems • Control the TN (denitrification) • Avoid the concurrence of decreased DO & NO_2^- accumulation

2.4.2 *N_2O mitigation strategies: progress and limitations*

Statistical and/or mechanistic modelling as well as observatory analysis of the N_2O emissions' behaviour have been commonly used either as standalone or in combined analyses for the development of mitigation strategies. However, the suggested mitigation schemes have not yet reached commercial applications at full-scale wastewater treatment processes. There is a gap between the identification of appropriate N_2O mitigation measures and their integration into the control of WWTPs. Future studies shall focus on the development, implementation and integration of the mitigation strategies into the existing control strategy of wastewater treatment processes. Special attention must be paid to trade-offs between GHG emissions, energy consumption, system performance and compliance with the legislative requirements in order to support evidence-based multi-objective optimisation of the WWTPs operation.

The investigation of direct GHG emissions at full-scale wastewater systems is important for the minimisation of the environmental footprint of WWTPs and the integration of the sustainability dimension into wastewater treatment process control.

2.4.3 *Integrating the N_2O emission monitoring into the WWTP operation*

Multivariable and multi-objective approaches have been proposed for the optimisation of the WWTPs performance to enable the correlation of energy consumption or operational costs with system performance (Qiao and Zhang, 2018; Zhang et al., 2014). Sweetapple et al. (2014) used a modified version of the benchmark simulation model 2 (BSM2) (Jeppsson et al., 2006) to identify control strategies for the simultaneous minimisation of GHG emissions, operational costs and pollutant loads. The BSM2 models have also been combined with LCA to evaluate the sustainability of different operating strategies (Flores-Alsina et al., 2010; Arnell et al., 2017). However, studies utilizing long-term real-field WWTP data are still scarce.

Figure 2.7 summarises the research priorities in terms of real-field N_2O monitoring campaigns that can act as foundation for the integration of N_2O emissions into WWTP monitoring and control. Long-term monitoring campaigns that capture the seasonal variability, studies on the uncertainties of the sampling strategies and reporting of operational, environmental and sampling data can ensure the robustness and comparability of the monitoring results. The development of methodological approaches for the translation of WWTP data into information can facilitate the understanding of the N_2O emission behavior and relationship with different

operational conditions. Combination of mechanistic models and/or data mining techniques with methods for the N_2O quantification can validate the models and provide insights into the dominant pathways under changing operational conditions. Finally, research studies implementing and translating the N_2O mitigation measures into control strategies are essential.

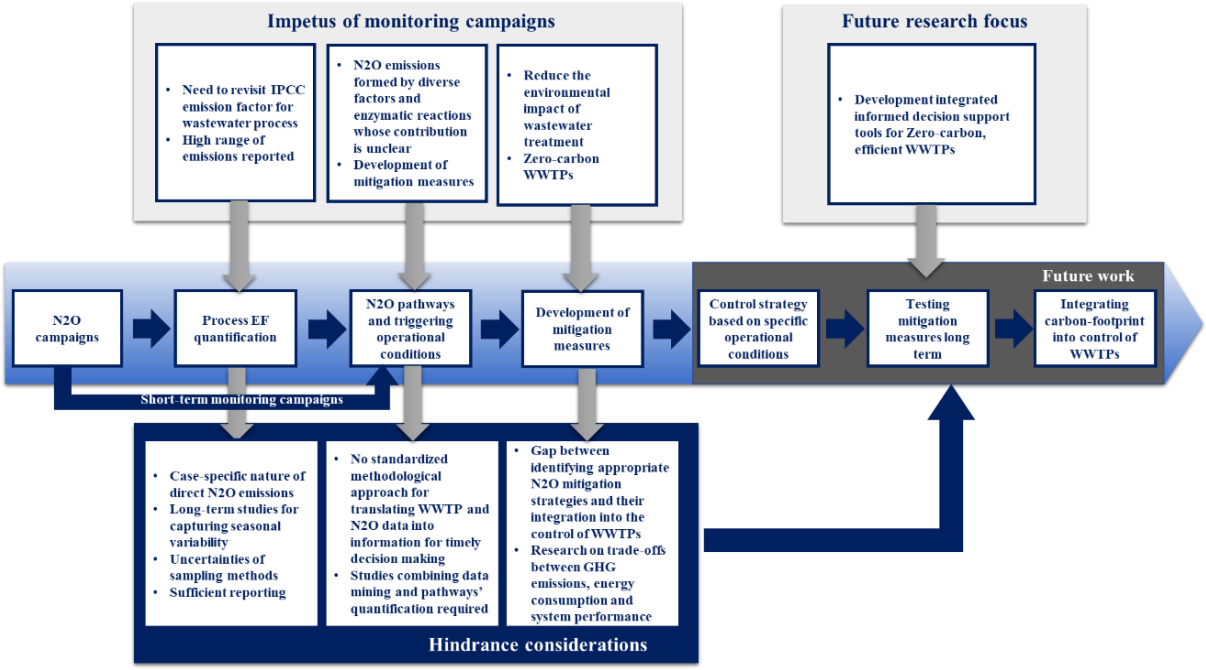


Figure 2.7: Monitoring N_2O emissions in full-scale wastewater treatment systems - research priorities

2.5 Summary of main findings

A number of full-scale N_2O monitoring campaigns in WWTP were studied. The processes were classified so that the ranges of N_2O emission factors, dominant N_2O pathways and triggering operational conditions and the mitigation measures for different process groups could be set. The key conclusions of the current review are:

- There is a wide range of EFs within similar groups of wastewater treatment processes. The emission factor ranges between 2% and 5.6% of the influent N-load in mainstream SBR, while OD reactor types, exhibit have a low EF, ~0.14% of the N-load. Long-term continuous or discontinuous monitoring campaigns are characterised by higher EF compared to short-term campaigns. The studies investigating seasonal behaviour of N_2O emissions, have an average EF equal to 1.7% of the N-load, whereas monitoring

campaigns lasting less than a month have an average EF equal to 0.7%. Long-term campaigns show a high variability of the N_2O emissions. There is no specific EF correlation with the NH_4^+ removal in the mainstream processes or in specific groups of processes. Most of the processes in smaller WWTPs (i.e. flow-rate<200,000 m^3/d) had EFs less than 0.5% of the N-load, independently of the process type and nitrification efficiency. This study concluded that efficient operational strategies can mitigate the generated N_2O for different configurations and groups of processes.

- It is difficult to compare the results of the N_2O monitoring campaigns because of: i) the differences in the duration of the monitoring campaigns (e.g. short-term campaigns ignoring seasonal variations), ii) the uncertainty in the gaseous sampling methods and analytical measurements, and iii) the insufficient reporting regarding the reactor control strategy, operational and environmental conditions, etc.
- Simple feature extraction and graphical representations of selected process variables and N_2O emissions are employed to explain the N_2O triggering mechanisms. Given that a combination of several parameters affects the N_2O generation, multivariate statistical analysis techniques can be a useful alternative for analyzing data and understanding the N_2O emission behavior. The combined application of mechanistic models and statistical techniques can lead to better design of mitigation strategies.
- Isotopic and molecular biology analyses are emerging techniques that can qualitatively and quantitatively assess the N_2O generation pathways. Data mining methods can be deployed to identify patterns of operational conditions and N_2O emissions. This can be complemented with techniques for the determination of the N_2O production pathways.
- Studies testing and validating the long-term full-scale N_2O mitigation measures are still missing.
- Future research should focus on: i) long-term N_2O monitoring campaigns, ii) the uncertainties of different sampling protocols, iii) the application of data mining approaches, machine learning and mechanistic models for the development of effective

and adaptive models to be integrated into WWTP operation and control and iv) the development, implementation and integration of the mitigation strategies into the existing control strategies of WWTPs.

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

3.1 Introduction

The demand to increase efficiency and reduce the environmental footprint of the municipal wastewater treatment plants (WWTPs) in terms of, greenhouse gases (GHG) and energy consumption reduction, is posing new challenges for the water industry (Flores-Alsina et al., 2014). The climate change pressures, prompt the quantification and minimisation 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 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., 2009).

With the potential contribution of 265-298 times more than CO_2 for a 100-year time horizon to global warming (IPCC, 2019), 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 6% of the global anthropogenic N_2O emissions (Palut and Canziani, 2007). N_2O is generated mainly during the autotrophic nitrification and heterotrophic denitrification (Kampschreur et al., 2008a) and can contribute up to 78% (Daelman et al., 2013b) of the operating carbon footprint of a wastewater treatment facility. Recent studies have focused on the understanding, quantification, control and minimisation 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). 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 could not be fully explained (Daelman et al., 2015; Kosonen et al., 2016).

In an effort to identify interconnections between operating and environmental variables as well as N_2O formation, statistical techniques have been used for the analysis of data from full-scale monitoring campaigns. Aboobakar et al., 2013 applied linear multiple regression analysis to investigate dependencies between variables monitored online in a plug-flow reactor implementing biological nitrogen removal and N_2O emissions. Their study showed a variable impact of nitrogen load, temperature and dissolved oxygen (DO) on the N_2O emissions in each reactor compartment, while the effect of DO was more intense in oxygen abundant zones. Multi-regression analysis of one year of data 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. However in the same study, bi-monthly grab-sampling provided limited data-points for a comprehensive analysis of the dynamic behavior of N_2O emissions and operating variables. Brotto et al., (2015) used Spearman's rank correlation to explain the behavior of N_2O emissions in an activated sludge system. 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 have not considered 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.

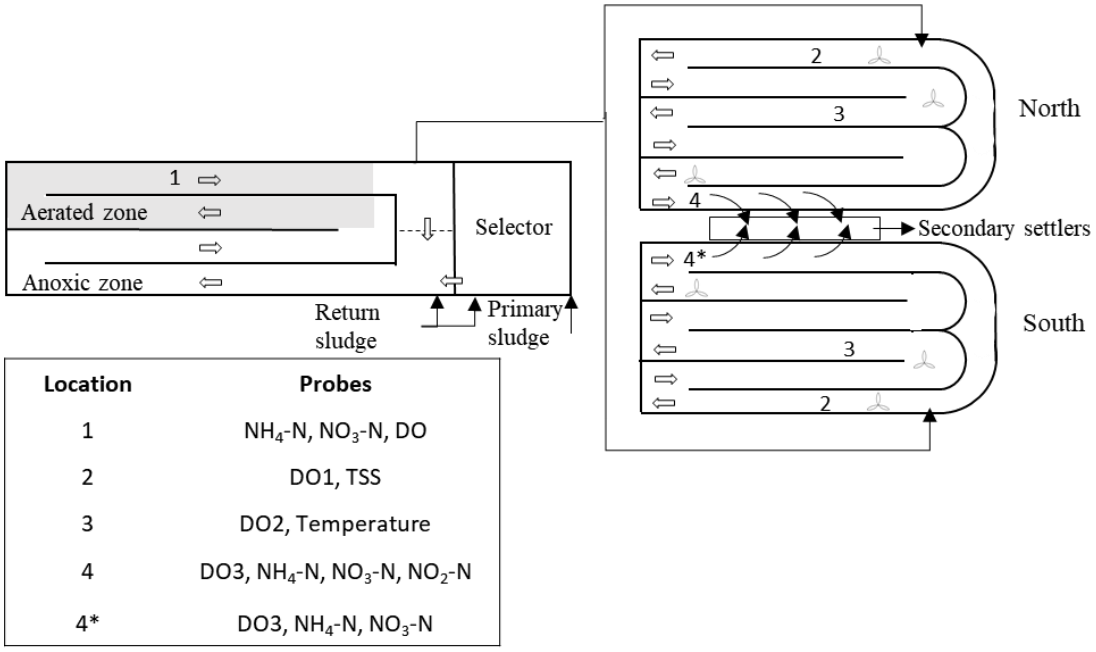
Multivariate analysis has been proven to be a suitable method for the identification of patterns and hidden relationships within WWTP data (Rosén and Lennox, 2001) and can be applied to provide insights on the combined effect of operational variables on N_2O emissions in full-scale systems. Chemometric 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 process (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 studies.

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 behaviour of N_2O emissions and explain the combined effect of the operating variables monitored in wastewater treatment processed 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 behaviour combined with hierarchical k-means clustering and PCA, to provide insights on N_2O emissions patterns and generation pathways.

3.2 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 treats 80.000 m^3 d^{-1} of domestic wastewater from a combined sewer system. The Carrousel reactors are characterised by alternating anoxic/oxic zones; aeration is performed through surface aerators which are manipulated by the loop controlling the ammonium concentration in the effluent. Aerator 1 operates under on/off pattern (when ammonium is higher than 1.2 mg/l), while aerators 2 and 3 operate always and peak when ammonium is higher than 0.6 and 0.9 mg/l, respectively. The average efficiency of the system in terms of total nitrogen (TN) removal was 81 ±10% and in terms of COD removal equal to 87 ±5% during the monitoring period. 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 3.1).



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

The Carrousel reactors are equipped with NO_3-N, NH_4-N, total suspended solids (TSS), temperature probes, and 3 DO probes (DO1, DO2, DO3) (locations 2, 3, 4, Figure 3.1). The Northern Carrousel reactor is also equipped with a nitrite probe. All the reactors are covered, and the off-gas is collected in ducts and pumped to a Servomex gas analyser, where N_2O is measured. Table B.1 lists all the variables monitored online (Appendix B). 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 quality with ~6 days frequency, were also available and considered in the analysis. Figure 3.1, summarises the methodological framework applied in the database.

3.3 Methodological Framework for data analysis

The monitoring period was divided into 10 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 can be easily

interpreted. The methodological framework is extensively described in the sub-sections which follow.

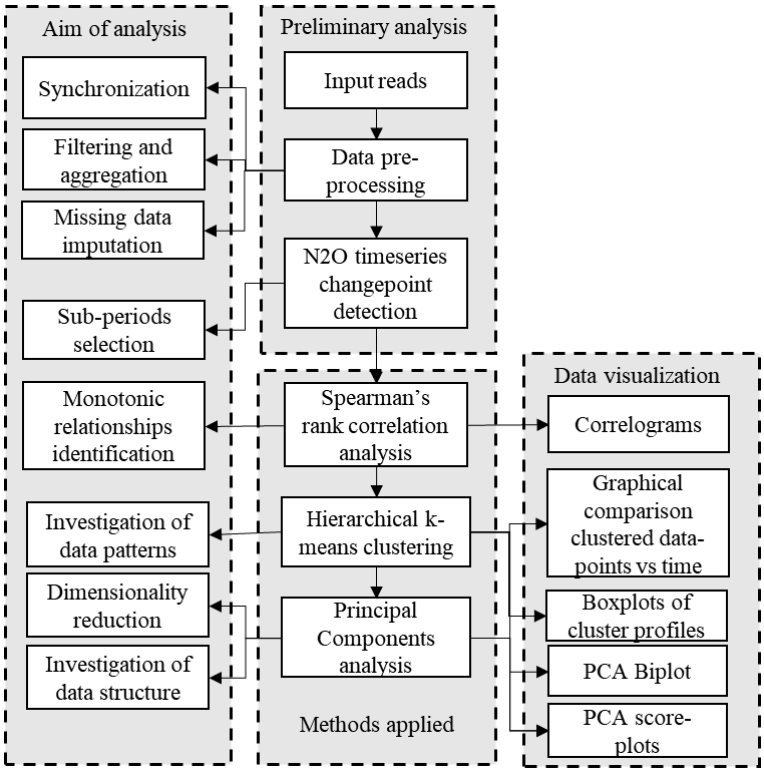


Figure 3.2: Methodology followed in the current study for data processing and visualization

The data-driven approach enabled the utilisation 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 Appendix B (Table B.2).

3.3.1 *Preliminary data processing*

The preliminary data analysis includes: (i) data synchronisation under the same time-stamp, and ii) removal of duplicate and unreliable measurements (multiple readings at the same time stamp for the same sensor). Additionally, 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.

3.3.2 *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.

3.3.3 *Application of 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 serves 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.

3.3.4 *Application of 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 recognised clustering algorithm (Haimi at al., 2013). K-means clustering was applied in order to categorize the data in groups of similar observations and investigate the patterns of N_2O emission fluxes and Euclidean distance is used as a distance function. K-means algorithm begins with the selection of k random centroids of the same dimension with the original data. All the data 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 are minimised. However, the randomly selected initial centroids can result into locally optimised 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 are normalized before the analysis and the number of clusters is selected in order to maximise the average silhouette value (Rousseeuw, 1987).

The profile of the N_2O emissions was highly variable during the monitoring campaign. Binary segmentation enabled the identification of the sub-periods that are characterised by different N_2O emissions' profile. Hierarchical k-means clustering was applied to the Carrousel reactor data matrix considering the different sub-periods in order to investigate whether different temporal patterns of the operating variables were responsible for the different behaviour of N_2O emissions during the monitoring campaign. Additionally, hierarchical k-means clustering enabled the i) detection of frequency and persistence of extreme ranges of operating variables, and ii) comparison of the operational modes between the plug-low and Carrousel reactor. Ammonium and nitrate probes in the plug-flow were included in the analysis, since they can provide indirect feedback in terms of the Carrousel reactor influent and additional information for the operational behaviour of the system. However, the analysis was repeated excluding plug-flow variables (NH_4-N and NO_3-N). As shown in Figure 3.2 graphical comparison of the clustered data-points vs time and boxplots of the variables in each identified cluster are displayed in the results' section.

3.3.5 *Application of 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 (Principal components) explaining the maximum observed variability. The principal components are linear combinations of the data variables. The loadings of the variables in each principal component can map their relationship with the respective principal component. 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, $x_j = (x_1, \dots x_i)T$ 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*:*

$$X = TP^T + E \tag{3.1}$$

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

The biplot of the first 2 PCs is used in order to visualize the combined behavior of significant variables that affect the system. The biplot enables 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, that enable the identification of occasions in which the behavior of the system changes, are also displayed in the results section. PCA was applied to the data matrix of the Carrousel reactor excluding N_2O emissions timeseries in order to i) identify the most significant variables that affect the system, (ii) analyse the structure of the sensor data, iii) investigate if changes in the relationship of the system coincide with changes in the N_2O emissions profile, and iv) validate the results from hierarchical clustering. N_2O emissions timeseries were excluded from the PCA in order to investigate the relationship between the PC scores and N_2O emissions and examine which PCs are more significantly linked to the behaviour of N_2O emissions.

3.4 Results and discussion

3.4.1 *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 are linked with lower N_2O emissions.

Table 3.1 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 characterised by very low N_2O emissions. Gradual decrease was observed until November 2011 and negligible emissions again until January 2011 (Figure 3.3).

Table 3.1:Variables monitored in the Northern Carrousel reactor (average ± std) (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			

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. The changepoints detected are displayed by the vertical dotted lines in Figure 3.3 (bottom). 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.

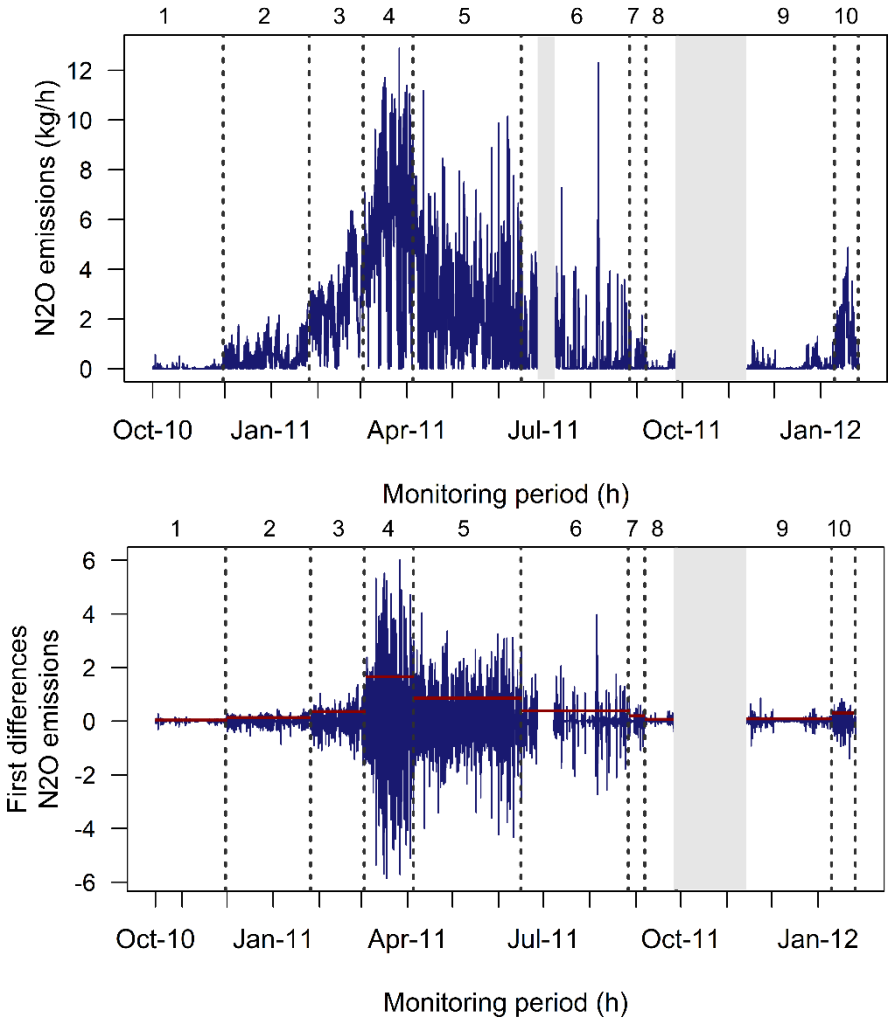


Figure 3.3 (top): N_2O emissions profile in the Northern Carrousel reactor (grey area: periods with missing N_2O data) (bottom): First difference of the N_2O emissions timeseries (blue line) showing the sub-periods identified by the application of binary segmentation (grey area: periods with missing N_2O data, blue dotted lines: changepoints identified by the algorithm, red horizontal lines: standard deviation in each sub-period)

The average COD concentration in the influent of the plug-flow reactor (effluent of primary sedimentation) is 239 ± 80 mg COD / L during the 15-month monitoring period. The average influent and effluent concentrations of COD, TKN, BOD and TP for each sub-period are given in the Appendix B (Table B.3). A slight increase of COD concentration (by 27% compared to average) is observed in sub-period 5 that can be attributed to the reduction of precipitation events and to the 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) is by 17% less compared to the respective one in sub-period 1. TKN and TP loadings are also reduced by 11%

and 12% in sub-period 4 compared to sub-period 1 respectively. The COD:TKN:TP ratio remains almost stable ranging between 1:0.17:0.02 (sub-period 2) and 1:0.20:0.03 (sub-period 4).

Figure 3.4 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 coincided with low N_2O emissions. Although the COD/N ratio of the stream entering the Carrousel tank was not measured in this study, the laboratory analyses indicate that limitation of COD cannot be considered the sole contributor of N_2O emissions via heterotrophic denitrification in sub-periods 4 and 5.

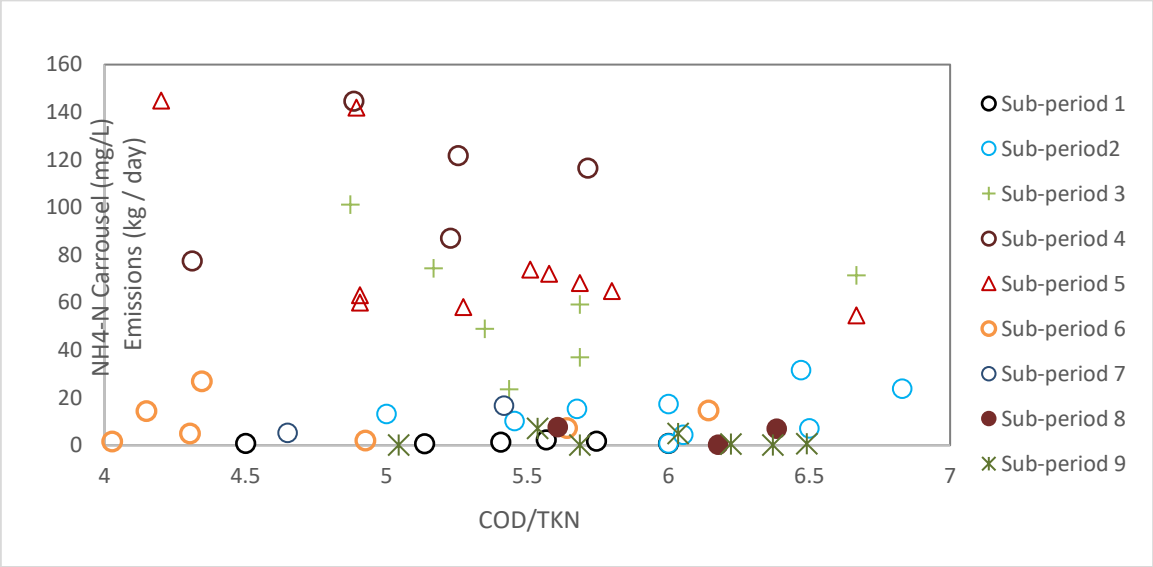


Figure 3.4: COD/TKN (offline data) for each sub-period

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 was from 73 % (sub-periods 1 and 9) to 92% (sub-period 5) and from 67% (sub-period 7) to 87% (sub-period 4). The pH of the effluent is steady (~ 8) and does not show seasonal variability that could influence the generation of N_2O emissions. A significant variation is observed for all variables monitored by looking at the complete database. Table 3.2 summarises the average values and standard deviations of the 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 mgN/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).

Analysing the whole profile, the emissions tend to be low at higher temperatures (sub-periods 6, 7, and 8). Higher emissions are also observed, though, at temperature higher than 18 °C and low nitrite concentrations (i.e., sub-period 5). According to Ahn et al. (2010b) 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 are also observed when ammonium concentration is lower than 13 mg/l and nitrate is higher than 2.5 mg/l. in the plug-flow reactor. The probe is 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.

The analysis of the variables' ranges for the N_2O emission profiles provides limited insight on the dependencies between the system variables which will be further analysed in the following sections.

Table 3.2: Average values and standard deviations of the main variables for the 10 sub-periods (C: =Carrousel reactor, N: Northern, PF: plug-flow reactor).

Var.	N_2O (kg/h)		NO_3-N C (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

3.4.2 *Spearman's rank correlation analysis*

This section includes the results from the Spearman's rank correlation for the Carrousel reactor. The application of Spearman's rank correlation coefficient in the data obtained during the whole monitoring period could not identify significant correlations between the N_2O emissions and the operating variables. The lack of monotonic univariate dependencies can 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. Considering the sub-periods, fluctuating correlation coefficients between N_2O emissions and Carrousel reactor variables were identified (Appendix B, Figures B.1-B.2). 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 between nitrite and N_2O emissions ranges from 0.78 (sub-period 7) to 0.51 (sub-period 9). As a general remark, nitrite is correlated with N_2O emissions in sub-periods 4, 6 and 7, while lower correlation was observed during sub-periods 5 (Figure 3.5), 8 and 9. N_2O emissions and NO_3-N concentration in the Carrousel reactor exhibited positive correlation - higher than 0.7 for sub-periods 2 (Figure 3.5), 4 and 10 (the temperature was lower than 13 °C in all cases). Both NO_3-N and N_2O follow similar diurnal patterns, whereas peaks in nitrate concentration coincide 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 utilisation rate in these sub-periods is affected by the low temperatures (Elefsiniotis and Li, 2006). Additionally, when N_2O is positively correlated with DO1 (> 0.5), medium to significant correlation with ammonium concentration in the Carrousel reactor is also observed (sub-periods 1, 6 and 7). Stripping of the already formed N_2O can be a potential explanation. The surface aerator in the location of DO1 probe operates based on the loop

controlling the ammonium concentration in the effluent; thus, peaks trigger the surface aerators to start.

The correlation coefficient between any two of the system variables did not remain stable in the examined sub-periods. Figure 3.5 shows an example of a correlogram for sub-periods 2 and 5. These sub-periods are characterised by low and high ranges of N_2O emissions and temperature respectively (Table 3.2). In sub-period 2, the average NO_3-N concentration in the plug-flow reactor is equal to 2.5 mg/l (Table 3.2) and correlates negatively with the influent flow-rate (~ -0.63) (Figure 3.5). In sub-period 5 the behavior of nitrate concentration (average equal to 2.1 mg/l) is mainly correlated negatively with ammonium concentration in the same reactor. The ammonium concentration in the Carrousel reactor is positively correlated with DO1 only in sub-period 2. NH_4-N concentration in the plug-flow reactor is correlated with the influent flow-rate only in sub-periods 4 and 5. However, the profiles of these two variables show that in the majority of sub-periods, abrupt and rapid increase of influent flow-rate (i.e., precipitation events) coincide with increase of the NH_4-N. However, the NH_4-N concentration tends to dissipate 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 is 0.26. However, when days with significant precipitation events (and thus high influent flow-rate) are omitted, the correlation coefficient is 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 is 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 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 B.3, Appendix B). The correlograms for all the monitoring sub-period are provided in the Appendix B (Figures B.1-B.2).

Spearman's rank correlation indicated structural changes in the dependencies between the system variables. Therefore, the fluctuating structural dependencies have 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, TSS) 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.

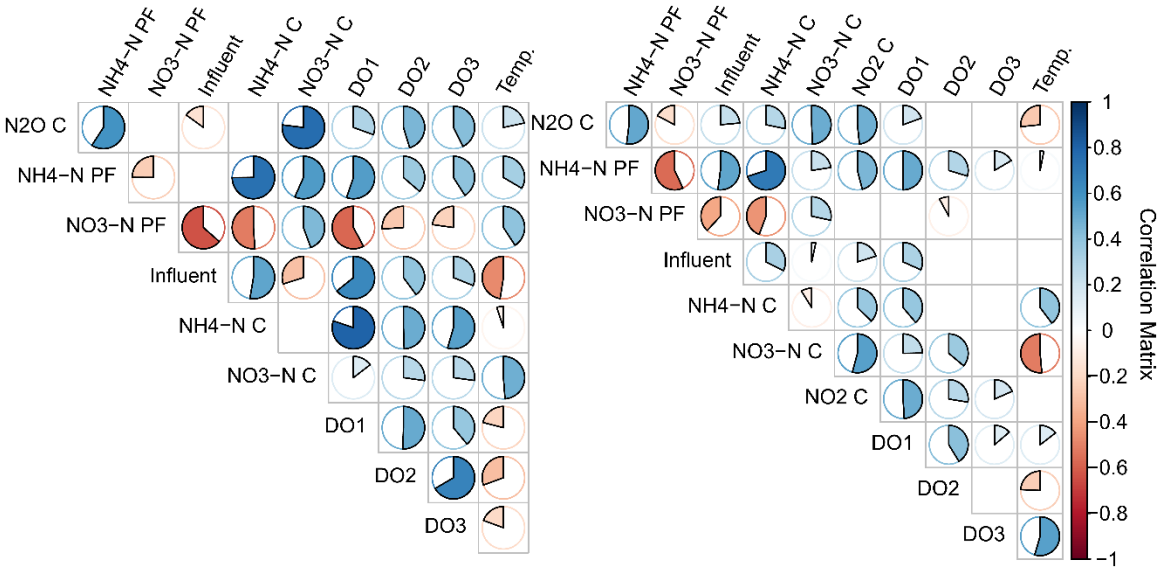


Figure 3.5: Spearman's rank correlation coefficient for sensor signals in Northern Carrousel reactor. (Left): Sub-period 2. (Right): Sub-period 5. (Red: negative correlation, blue: positive correlation, the coloured part of the circles is proportional to the correlation coefficient, only results with p-value < 0.01 are shown)

3.4.3 *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 in order to evaluate the effect of these parameters on the clusters identified. 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 is differentiated in the majority of the clusters. In all the examined sub-periods, two major clusters were identified since these are characterised 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 represent 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 characterised by low average N_2O emissions (i.e., 1, 2, 7, 8 and 9), clusters with increased N_2O emissions (yet relatively low) are 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 are observed when the average NO_3-N concentration is consistently lower than 1 mg/l in the plug-flow reactor and NO_3-N concentration is lower than 4 mg/l in the Carrousel reactor. Table 3.3 compares the clustered average values for all the variables in sub-periods 2 (average N_2O emissions equal to 0.6 kg/h - Table 3.2) and 4 (average N_2O emissions equal to 5.6 kg/h – Table 3.2). Sub-period 4 is characterised 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 Appendix B (Table B.5).

In clusters 2 and 16 the averages of operating variables are similar yet the N_2O emissions are different (0.01 and 0.51). Similarly, in clusters 1 and 4 and 7, the averages of operating variables are similar yet the N_2O emissions are different (0.09,0.87 and 3.22 respectively). A corollary to this also exists. In clusters 1 and 2 the averages of operating variables are different but the N_2O emissions are similar (0.09 and 0.01). Similarly, in clusters 5 and 6 the averages of operating variables are different but the N_2O emissions are 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 3.6 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 show a similar behaviour (in terms of range and trends). N_2O emissions profiles have also the same trend, however, their range strongly depends on the initial N_2O fluxes emitted at 00:00 am. The influent flow-rates, NH_4-N and NO_3-N concentrations in the plug-flow reactor are also similar in these two occasions. For the time-periods shown, the average N_2O fluxes are equal to 0.44 and 2.01 kg/h for occasion 1 and 2 respectively.

Table 3.3: 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

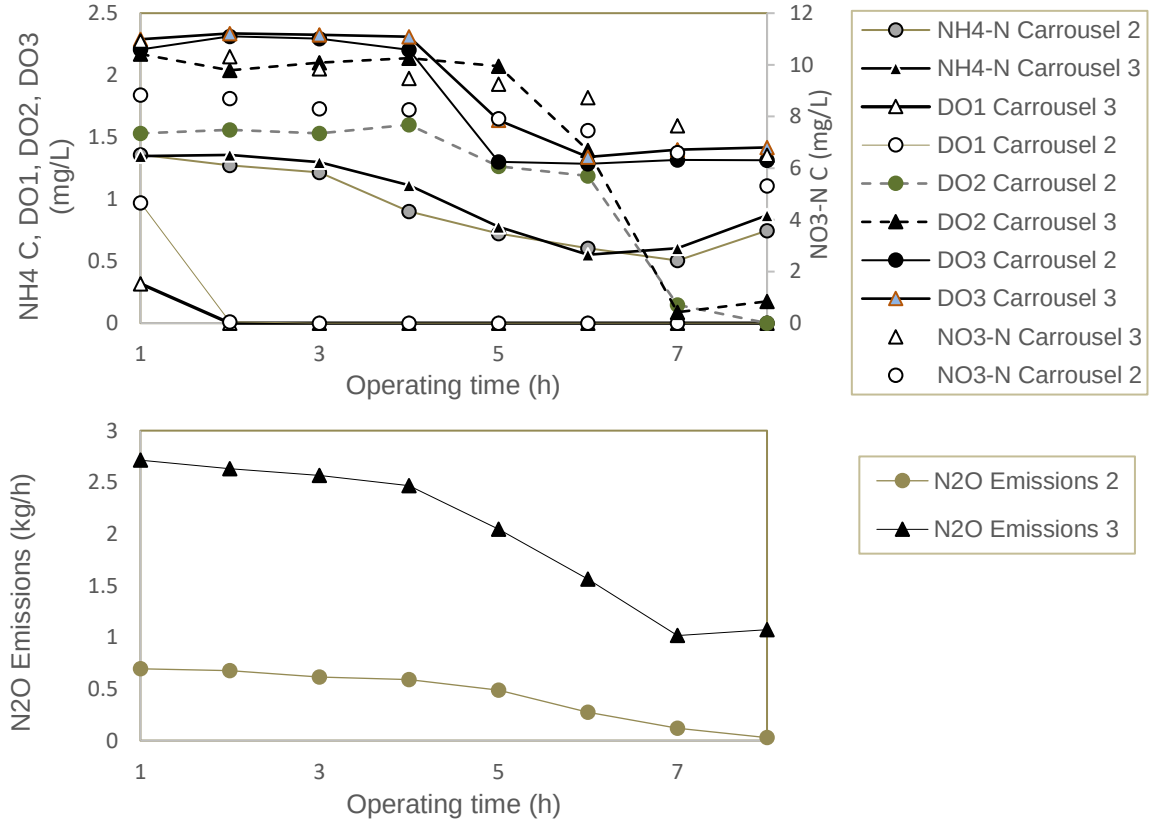


Figure 3.6: Variables monitored online for two separate occasions in sub-periods 2 and 3 (from 00:00 am until 8:00 am), (Bottom): The respective N_2O emissions profiles

3.4.4 *Principal component analysis in the Carrousel reactor*

PCA was applied in order to transform the original correlated measured variables to uncorrelated variables (Principal components) in order to 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 Appendix B (Tables B.6-B.13, Figures B.4-B.28). 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 shown in Table 3.4. The loadings of each component are an indication of the variation in the variables explained by a specific component. Influent flow-rate, ammonium in the Carrousel reactor (NH_4-N C) and the three DO (DO1, DO2 and DO3) concentrations have the highest negative loadings in PC1. This means that the first principal component increases 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 aerators operate based on the loop controlling the ammonium concentration in the effluent. PC2 links 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 has high negative loading, while DO2 and DO3 levels have 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 does not present a similar trend. Overall, in sub-period 2 the average DO concentration recorded in the Carrousel reactor from the 3 DO probes is increased compared to sub-period 1.

Table 3.4: 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

The biplot of the first 2 PCs is used in order to visualize the combined behavior of significant variables that affect the system. Data points assigned to cluster 6 (Figure 3.7), have negative scores in PC2 and PC1. Therefore, ammonium concentration in the Carrousel reactor and influent flow-rate are higher than average, while the nitrate in the system is lower than average. Figure 3.8 shows the profile of N_2O emissions and NH_4-N in the Carrousel reactor for sub-period 2. The coloured points in the diagram represent the identified clusters. Peaks in emissions coincide with peaks in the NH_4-N C profile, whereas peaks in NH_4-N C coincide with precipitation events (cluster 6).

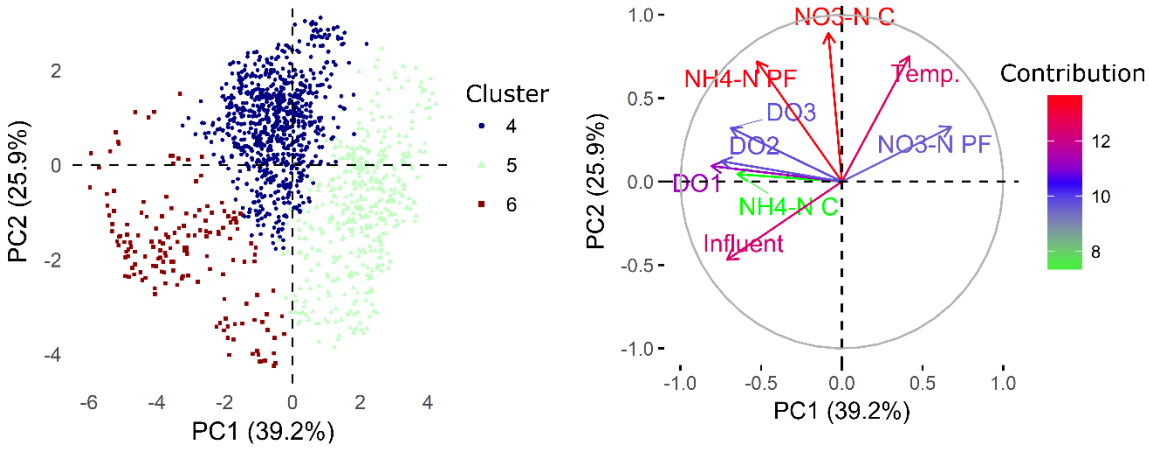


Figure 3.7: (left) Biplot of the first 2 PCs, sub-period 2. The coloured data-points represent the scores of the first two principal components. Groups 4, 5, and 6 represent sub-period 2, clusters. (right) Variable correlation plot. The arrows represent the direction and strength of the variables monitored 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 length of the arrow shows the quality of representation of the variables on the biplot (variable

coordinates = loading x component std). The percentage given on each axis label represents the value of the total variance explained by that axis.

The scores of the data-points in cluster 5 are mainly positive in PC1 and negative in PC2. PC2 increases with the increase of NH_4-N concentration in the plug-flow reactor (Table 3.4). Given that PC2 has average equal to 0 (data are standardised), 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 where the arrow of NH_4-N concentration in the plug-flow reactor points to the direction of increasing concentrations of NH_4-N in the plug-flow.

Therefore, data-points belonging to cluster 5 are characterised by higher than average ammonium concentration in the plug-flow reactor. Similarly, NO_3-N concentration in the plug-flow reactor has relatively significant positive loading in PC1 (0.36 – Table 3.4). The latter indicates that NH_4-N and the three DO probes in the Carrousel reactor (that have negative loadings in PC1 – Table 3.5) tend to decrease when NO_3-N concentration in the plug-flow increases. Given that all data-points in cluster 5 have positive scores in PC1, it can be concluded that are characterised 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 3.8, where the data-points of cluster 5 correspond to the daily low range of ammonium concentrations in both reactors.

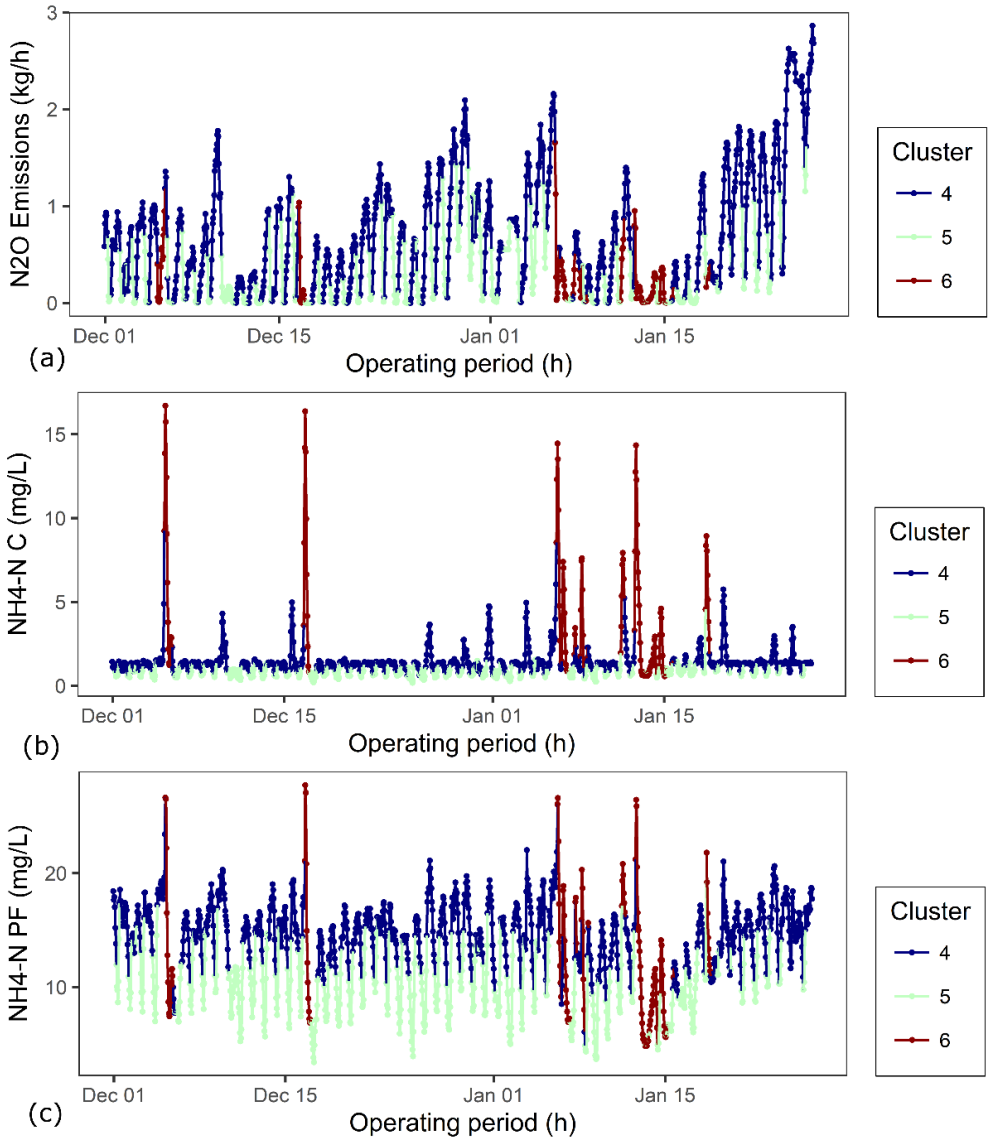


Figure 3.8: Profile of (a) N_2O emissions, (b) NH_4-N concentration in the Carrousel reactor and (c) NH_4-N concentration in the plug-flow reactor for sub-period 2; coloured points indicate the respective clusters

Figure 3.9 summarises scores of the PC2 and the respective clusters (coloured 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 occurs; the system is disturbed and cannot recover immediately. Spearman's rank correlation coefficient between PC2 and N_2O emissions is equal to 0.72 with the NO_3-N drops, directly resulting in reduced N_2O fluxes.

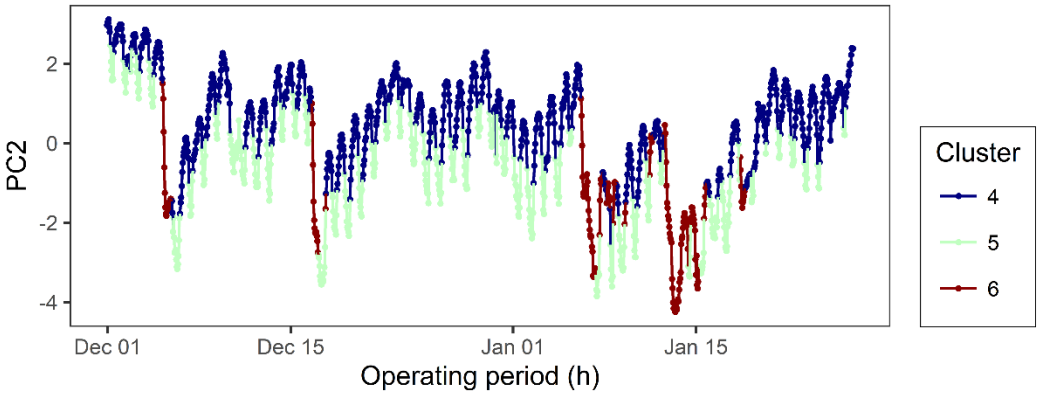


Figure 3.9: PC2 scores for sub-period 2

In sub-period 4, mechanisms triggering high N_2O emissions in the Carrousel reactor prevail (average = 5.6 kg/h). The PCA loadings are similar to sub-period 2, while the clustering results indicate 3 clusters; clusters 10 and 11 are affected by the diurnal variability and cluster 12 is affected by the precipitation events (Table 3.3). Again the 3 DO sensor data in the Carrousel reactor have significant negative loadings in PC1. However, ammonium concentration in the Carrousel reactor is 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 are observed in the effluent of the Carrousel reactor (17 data points belong to cluster 12). The correlation coefficient of PC1 with NH_4-N 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) is 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).

In PC2, the NO_3-N concentration and temperature have significant positive loadings (Table 3.5). The score plot of PC2 (Figure 3.10 (a)), has an increasing trend and therefore, shows that nitrate and temperature increase. The latter is validated in the profiles of NO_3-N concentrations in the Carrousel reactor (Figure 3.10 (b)) and NO_3-N concentration in the plug-flow reactor and temperature (Appendix B Figure B.29). In the beginning of the sub-

period very low concentrations of nitrate are observed in the system and they gradually increase especially after the 28th of March. The Spearman's correlation coefficient between N_2O emissions and PC2 scores relatively high and equal to 0.62. However, contrary to sub-period 2, the clustering analysis showed that there is no nitrate accumulation (Table 3.3). The average nitrate concentration in the plug-flow reactor is equal to 0.2 mg/l until the 28th of March and increases up to 1.6 mg/l until the end of the sub-period. Therefore, the observations in section 3.4.3 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 can be disturbed by the low temperature resulting in nitrite accumulation). PCA also showed that the metabolism of bacteria requires time to recover from the disruption and emissions start decreasing in sub-period 5.

Table 3.5: 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

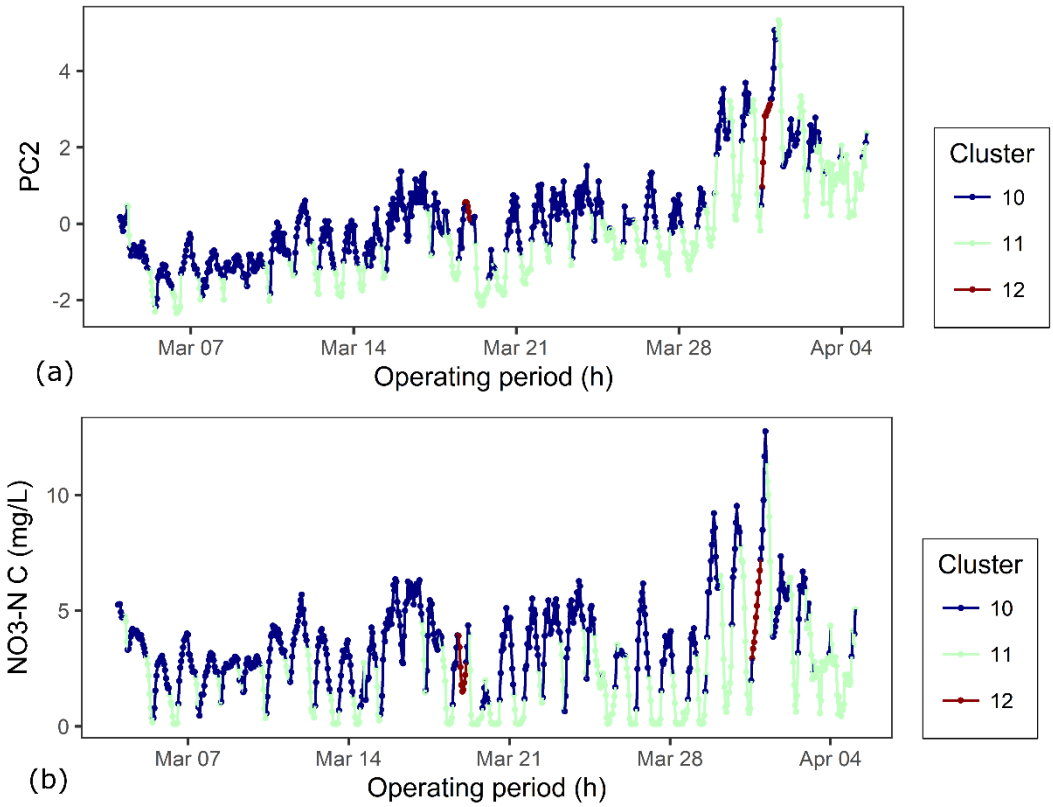


Figure 3.10 (a) PC2 scores for sub-period 4 and (b) NO_3-N concentration in the Carrousel reactor for sub-period 4.

3.4.5 *N_2O emissions generation pathways*

The results from the application of multivariate statistical techniques can be used for the identification and explanation of potential pathways for N_2O emissions generation. 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 the current study, the combination of hierarchical k-means clustering and PCA linked the different emissions ranges with all the monitored variables in the reactor (i.e. Figure 3.7) and enabled the identification of a set of variables that are connected with N_2O emissions in each sub-period (i.e. Figure 3.8). Therefore, the analysis facilitated the identification of the dominating pathways for N_2O emissions during the 15-month monitoring campaign.

In sub-periods with lower average N_2O emissions fluxes (1, 6, and 7) emission peaks coincide with ammonium peaks in the effluent of the plug-flow reactor and therefore in the Carrousel reactor. Wunderlin et al., (2012) demonstrated that N_2O can be partly produced by hydroxylamine oxidation, with excess ammonia, low nitrite concentration and high ammonia oxidation rate. In that case, average emission fluxes range between 0.05 kg/h (sub-period 1) to 2.54 kg/h (sub-period 6) and tend to be higher at higher temperatures and DO levels. The high DO concentrations coincide with peaks in nitrite and nitrate concentrations indicating insufficient denitrification zones in the reactor. AOB can use nitrite instead of oxygen as electron acceptor (Kampschreur et al., 2009) especially in oxygen limiting conditions (low DO zones exist even when all aerators are under operation); thus nitrifier denitrification by AOB can potentially contribute in N_2O emissions. In line with the results of the current work, 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 increase significantly and peak at low nitrate concentrations in both reactors (i.e., sub-periods 3 and 4) and high nitrite concentrations (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). In a pilot Carrousel reactor (Zheng et al., 2015) observed a synergistic N_2O generation between nitrifier denitrification and heterotrophic denitrification, in which 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. Low nitrate concentrations together with high TSS resulted in low N_2O emissions. The latter can be explained by the higher efficiency in NO_3-N removal. The combined results of PCA and hierarchical k-means clustering can guide through the most significant N_2O production pathways in different sub-periods (Appendix B).

3.5 Summary of main findings

- This study applies a data-driven approach to analyse long-term N_2O emission dynamics and generation mechanisms utilizing high temporal resolution data, i.e. hourly process variables' data, through the application of statistical-based methods. The data-set was acquired from a 15-month N_2O monitoring campaign and was divided into 10 sub-periods based on the N_2O emissions profile.
- Spearman's rank correlation showed significant univariate correlations of N_2O with ammonium, nitrate and nitrite. The correlations fluctuate between the 10 sub-periods. However, it was observed that low correlation coefficients can indicate non-monotonic interrelationships that Spearman's rank correlation cannot identify.
- Hierarchical k-means clustering was applied and provided information on the existence of reoccurring patterns and their effect on N_2O emissions. N_2O emission peaks were linked with the diurnal behaviour of the nutrients' concentrations and with rain weather events, whereas low nitrate concentrations in the preceding plug flow reactor (<1 mg/l) resulted in increased loadings in the subsequent Carrousel and high N_2O emissions.
- The PCA 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 provided 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. Therefore, multivariate analysis can be used to assist researchers and operators to understand and control the emissions using long term historical data.

4. Data-driven *versus* conventional N_2O EF quantification methods in wastewater; how can we quantify reliable annual EFs?

4.1 Introduction

Nitrous oxide (N_2O) emitted during biological nutrients removal, can significantly contribute to the total carbon footprint of Wastewater Treatment Plants (WWTPs). The recent roadmap to carbon neutrality in urban water published by Water and Wastewater Utilities for Climate Mitigation (WaCCliM) project and the International Water Association (IWA) (Ballard et al., 2018), states that direct N_2O should be considered for the carbon footprint assessment and reporting.

N_2O fluxes in wastewater processes are characterised by significant spatial and temporal variability due to the different interacting biological processes that consume or produce N_2O and the variation of operational and environmental conditions (Daelman et al., 2015; Gruber et al., 2019). A recent analysis of N_2O emission factors (EF) for over 70 full-scale wastewater treatment processes revealed that the sampling frequency and sampling techniques applied in N_2O monitoring campaigns, can significantly affect the quantified EFs (Vasilaki et al., 2019). For instance, most of the monitoring campaigns lasting less than one month have reported EFs less than 0.3 % of the N-load. On the other hand, studies lasting over a year result in a median EF equal to 1.7 % of the N-load. The IPCC guidelines for the estimation of N_2O in WWTPs were updated in 2019; the suggesting an EF of 1.6 % for of the total N-load (IPCC, 2019). However, uncertainties remain; the use of measured emissions data is suggested for the estimation of country-specific EF in large WWTPs (IPCC, 2019). The development of process-based reliable N_2O EFs requires long-term monitoring campaigns of over 1-year (Gruber et al., 2019; Vasilaki et al., 2019).

Long-term N_2O sampling (continuous or via grab-samples) is still rarely performed in WWTPs. High cost and complexities of long-term online monitoring are the main limiting

factor. There is still lack of a holistic low-cost, practical approach for the quantification of N_2O EFs. Therefore, new approaches are required for the quantification of EFs, minimizing sampling rate and advising on the duration and frequency of sampling campaigns.

A large amount of raw, heterogeneous operational data is available from WWTP operations (Olsson et al., 2014). Several studies have demonstrated that utilisation of historical data (i.e. DO, mixed liquor suspended solids (MLSS), NH_4^+ concentrations) from WWTPs can feed statistical methods and predict the profile of target process variables or key performance indicators that cannot be monitored online; an overview can be found in the study of Haimi et al., (2013). Additionally, data-driven techniques have been extensively used to capture the non-linearities and complex structures of wastewater treatment processes towards their optimisation, monitoring and better control (Haimi et al., 2013; Corominas et al., 2018; Newhart et al., 2019). Vasilaki et al. (2018) showed that variables monitored online can be utilised to provide insights on the long-term behaviour and abrupt changes of N_2O dynamics with the application of clustering and dimensionality reduction techniques. However, advanced information extraction methods have rarely been used to analyse data from N_2O monitoring campaigns. Recently, Sun et al., 2017 developed a back-propagation artificial neural network (ANN) to simulate N_2O emissions in an anaerobic-oxic (A/O) process. The authors demonstrated the feasibility and simplicity of predicting N_2O emissions with the use of data-driven models.

Univariate and multivariate changepoint detection techniques have been widely used to detect changes in underlying distribution of sequences and regime shifts in several applications including investigation of distributional changes in financial markets (Allen et al., 2018) and climate change investigation studies (Kotta et al., 2018). Li et al. (2015), recently, applied a non-parametric multivariate changepoint detection algorithm (e-divisive; (James and Matteson, 2013)) to detect changes in water quality variables in a shallow lake (total nitrogen, total phosphorus and Chlorophyll) and linked the changepoints (CPs) with changes in the dynamics and patterns of sediment release.

In this work, the most prevalent sampling approaches in wastewater industry are presented and compared. A data-driven sampling approach and two conventional monitoring

approaches have been compared and assessed. The result of the comparison proposes the most accurate and efficient amongst the three methods. Specifically, an approach that uses CPs to analyze the behaviour of online monitored variables linked with N_2O generation (i.e. DO, NH_4^+), is proposed, to detect i) periods (between the CP intervals) with steady N_2O emissions profile and ii) changes in the temporal range and dynamics of N_2O emissions. Multivariate changepoint detection was applied to identify structural changes in the variables monitored online (i.e. NH_4^+, DO, flow-rate, NO_3^-) in a full-scale Carrousel reactor. Subsequently, the CPs were linked with changes in the N_2O emissions behaviour and range during a 15-month monitoring campaign (Daelman et al., 2015). A classification model was developed to predict the range of N_2O emission loads (i.e. low, medium, high) based on the CP intervals. This approach can support operators to minimise GHG sampling requirements, without compromising long-term EF estimates. The accurate quantification of annual N_2O EF requires samples collection between all CP intervals and a few sampling days can be sufficient to estimate a representative EF for different CP intervals. The classification model can support the estimation of the N_2O emission range for new incoming data in the WWTP.

4.2 Process description and the source of data

N_2O measurements and the extensive data-set of the operational variables from the studies of Daelman et al. (2015) and Vasilaki et al. (2018) were used in the analysis. The dataset belongs to one of the longest N_2O monitoring campaigns undertaken in the wastewater sector (15 months). A plug-flow reactor linked with two subsequent parallel Carrousel reactors was monitored. A full description of the WWTP can be found in the study of Daelman et al. (2015).

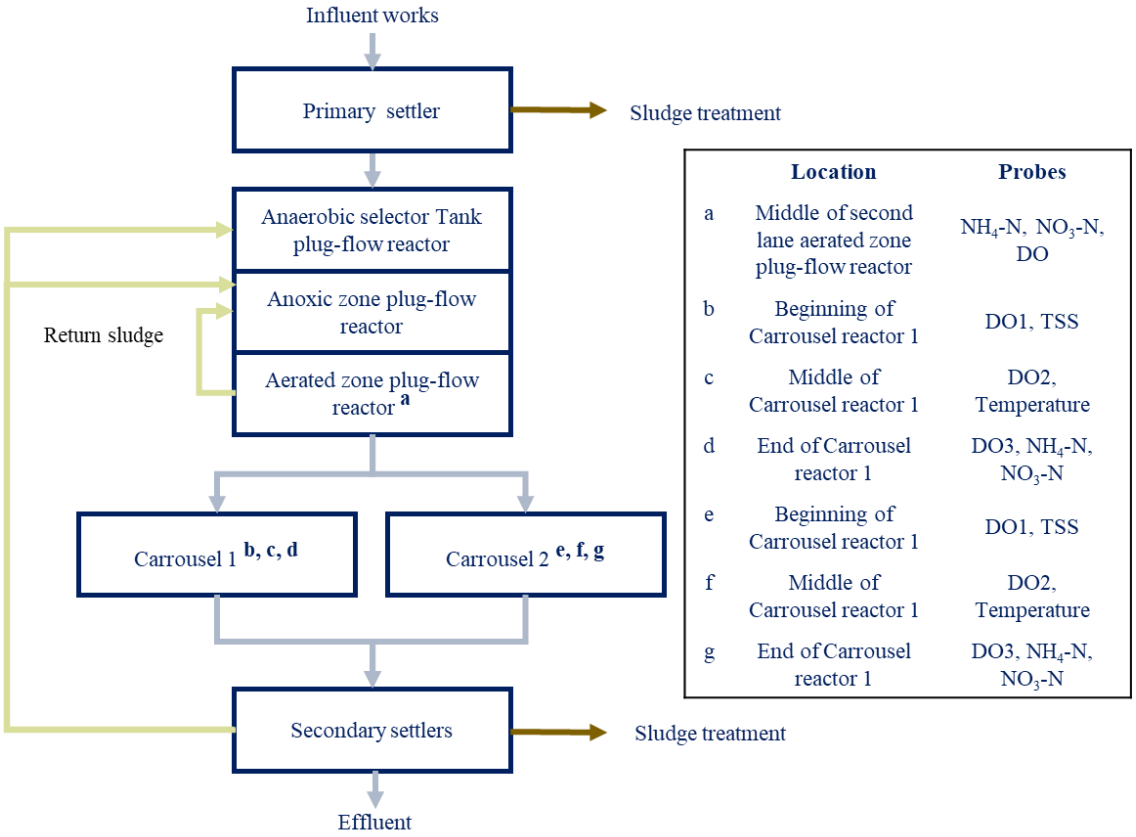


Figure 4.1: Flow-chart of the secondary treatment at Kralingseveer WWTP and sensors location

The analysis and the development of the methodological approach was based on data obtained from the Carrousel reactor 1 (Figure 4.1). The data matrix used in the analysis, the location of the sensors in the system and the details of the operational control are provided in the study of Vasilaki et al., (2018). The system includes the following probes: DO (DO1, DO2, DO3) in the beginning, middle and end of the Carrousel reactor, ammonium nitrogen (NH_4-N) and nitrate nitrogen (NO_3-N) in the effluent of the Carrousel reactor, NH_4-N from the middle of the second oxic zone in the plug flow reactor, temperature and influent flow-rate .

The behaviour of N_2O emissions at Carrousel reactor 1 showed a high level of volatility during the 15-month monitoring campaign and characterised by significant diurnal and seasonal variations (see supplementary material). The daily emission loads ranged from < 0.004 kg N_2O / day to >150 kg N_2O / day. Daelman et al. (2013) simulated different

sampling strategies, based on data collected from the same plug-flow – Carrousel reactor linking EFs with different sampling strategies. The authors simulated the sampling strategies using a long-term dataset. They reported that short-term campaigns (grab sample, 24 h and 7-day sampling), cannot accurately estimate annual EFs, while there is a high probability to underestimate actual emissions. The relative error of the estimated annual N_2O emissions ranged between -22 % and 35 % (95 % of the cases) by simulating a 50-days N_2O sampling campaign (random 24h periods on working days were selected). The authors found that long-term offline/online sampling capturing seasonality and temperature effects is needed for reliable EF assessment. Reliable estimation of N_2O emissions, can provide guidance on N_2O mitigation measures and support WWTPs towards their carbon neutrality goals. However, there is high cost and resources related to long-term, of N_2O online monitoring. Therefore, minimizing the sampling requirements can help water utilities to integrate N_2O monitoring in practice.

The behavior of N_2O emissions at Carrousel reactor 1 showed high level of volatility during the 15-month monitoring campaign and characterised by significant diurnal and seasonal variations. Figure 4.2, shows the daily emission loads ranging from < 0.004 kg N_2O / day to >150 kg N_2O / day. Vasilaki et al. (2018) applied a changepoint detection technique on the N_2O emissions hourly timeseries in Carrousel reactor 1, combined with hierarchical k-means clustering to investigate the N_2O emission patterns and identify links with the variables monitored online. The study concluded that i) the dependencies between N_2O and other operational variables (i.e. NH_4^+, NO_3^-, DO) varied in different sub-periods, ii) the system disturbances are mainly linked with events of elevated influent flow-rates, iii) specific ranges of operating variables have historically resulted in low or high ranges of N_2O in the different sub-periods. These findings have been used to develop the methodological framework of section 4.3.

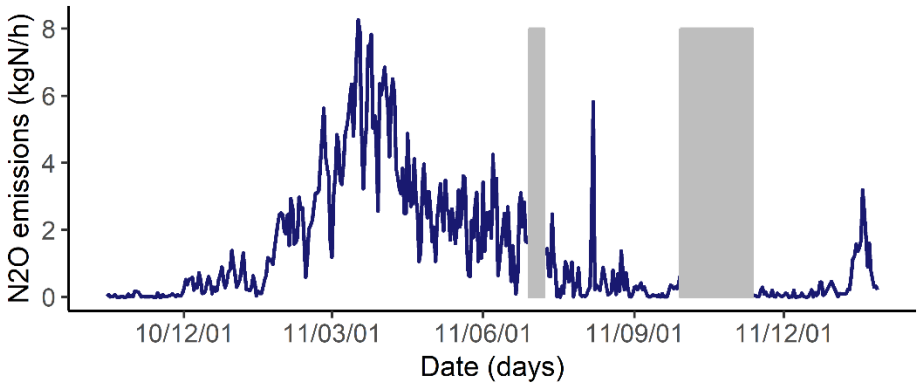


Figure 4.2: Daily N_2O emission loads; grey areas represent missing data

4.3 Methodological framework and data analytics

Figure 4.3 summarises the methodology applied in the current study. Pre-processed data obtained from the work of Vasilaki et. al. (2018) were used in the analysis. Previous examination showed that disturbances (i.e. precipitation events) significantly affect the NH_4-N effluent concentrations. Thus, the first step of the analysis was to isolate and categorize abnormal diurnal behavior of the influent flow-rate that affected system performance. Subsequently, sensor data were used to segment the behavior of the system into sub-periods with different behavior and operational variables ranges (i.e. NH_4-N, DO). The aim was to investigate whether changes in the N_2O emissions coincide with the changes in the range of operational variables. For this purpose, multivariate changepoint detection techniques were applied to categorize one-year historical sensor data into sub-periods exhibiting different behavior. The sequential segmentation of the operational variables enabled the quantification of N_2O EF over 1 year using a small number of random samples between segments Average estimated N_2O emissions were then compared with the respective EFs conventional monitoring techniques (equivalent sampling duration), following the methodology applied in the study of Daelman et. al. (2013a). Finally, features were extracted representing the diurnal behavior of the operational variables and classification models were trained to predict the range of N_2O emissions. The analysis was based on the N_2O emission ranges between the changepoint segments.

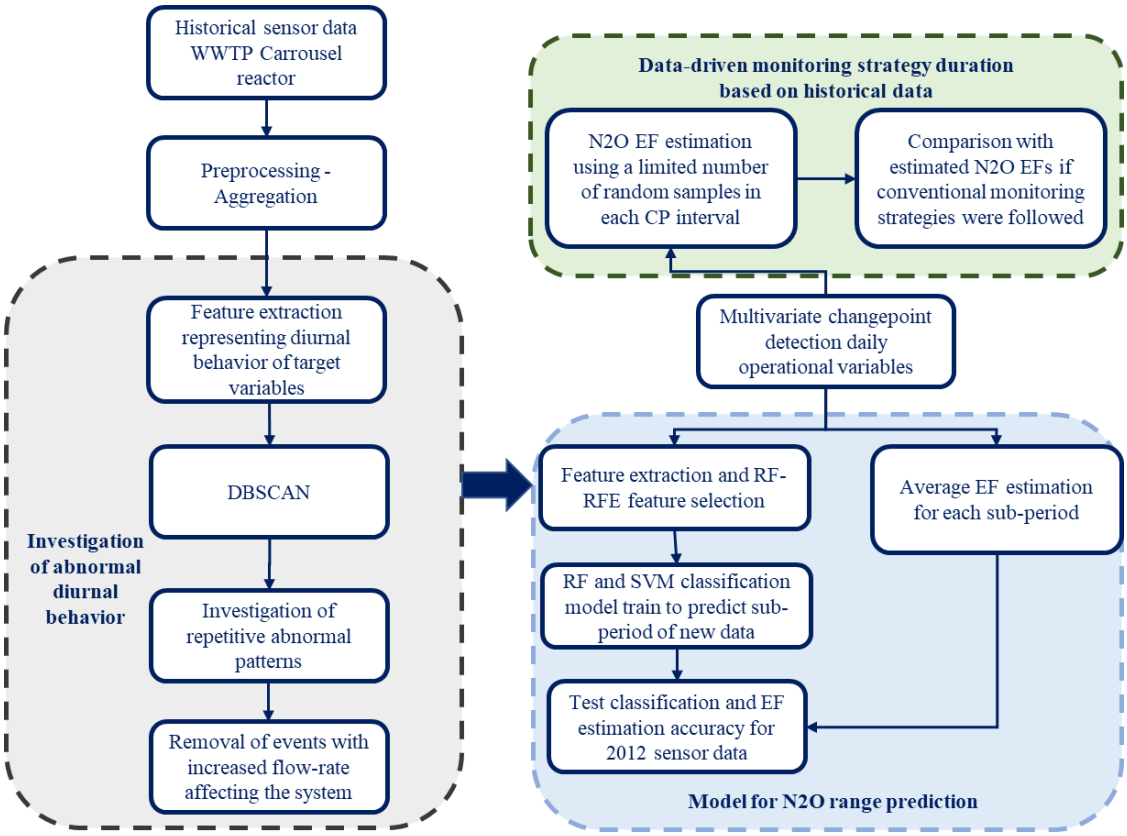


Figure 4.3: Methodology used for the optimisation of duration of monitoring campaign and sampling frequency. The analysis is based on historical data and the development of SVM and RF classifiers to predict N_2O emissions range

4.3.1 *Identification and isolation of influent-flow-rate increase*

A variation in the behavior of the online monitored variables was observed during the campaign. Previous analysis (Vasilaki et al., 2018) showed that abrupt and rapid increases in the influent flow-rate were linked with precipitation events and often resulted into peaks in ammonium concentration in the effluent of the Carrousel reactor. The following steps were performed in order to detect and isolate diurnal influent flow-rate patterns that affected the performance of the system (Figure 4.3): i) features were extracted representing the diurnal behavior of influent flow-rate and ammonium concentration, ii) the selected features were transformed into a lower dimension space using principal component

analysis (PCA) and iii) DBSCAN was applied to detect days that did not exhibit the expected dynamics and range of the target variables.

PCA (Jolliffe, 2002) was applied to reduce feature dimensionality by eliminating a small proportion of variance in the data. The principal components are linear combinations of the data variables. The loadings of the variables in each principal component can map their relationship with the respective principal component. 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. A detailed description of the methodology can be found in the study of Vasilaki et al., (2018)

Subsequently, density-Based Spatial Clustering of Applications with Noise (DBSCAN) (Ester et al., 1996) was applied to the features and clusters with regions of high and low density were detected. DBSCAN has been applied to identify outliers in various studies considering monthly temperature data (Çelik et al., 2011), building energy data (Jalori and Reddy, 2015) and multivariate sensor data (i.e. precipitation, humidity) (Saeedi Emadi and Mazinani, 2018). For instance, Jalori and Reddy, (2015) created 25-dimensional vectors of hourly building energy consumption (representing different daily energy consumption profiles) and used DBSCAN to isolate abnormal and group typical daily energy profiles.

In this study, DBSCAN was applied to the first three PCs extracted from the selected feature vector (explaining ~90% of the total variance) to isolate data at a distance greater than a pre-defined distance. The method relies on a density-based concept for separating data in high-density areas of the space (clusters) from data in low-density areas (events with high-influent flow-rates affecting the system performance). In DBSCAN two parameters are defined; the neighbourhood distance epsilon (*eps*) that defines the distance threshold in the data space and minimum number of points (*MinPts*). DBSCAN classifies the data-points into three groups: core points (data-points with at least *MinPts* neighbours within the *eps* distance), border points (data points that are neighbours of at least one core

point) and noise points (neither a core nor a border point). Data-points with distance lower than the *eps* distance from a core point are assigned to core point's cluster (direct density reachable points). Additionally, if one of the direct density reachable points is a core point, again its neighbouring data-points are assigned to the same cluster (density reachable points). Noise points are not density reachable the core points. In the analysis, *MinPts* was selected twice the number of variables as suggested by Jörg et al. (1998). Additionally, the *eps* distance was defined based on the 'knee' in the plot of the descending k-nearest-neighbour distances calculated for the data-points. DBSCAN was executed in the R package dbscan (Hahsler et al., 2017).

4.3.2 *Changepoint detection*

In this study, following a similar approach to Li et al., (2015), the aim was to investigate whether distributional changes and level shifts of variables conventionally monitored in wastewater systems can be used to detect changes in the range and formation N_2O emissions. Identifying changes in the online data collected from wastewater treatment processes is not straightforward; the time-series consist of a combination of seasonal, gradual and abrupt changes. For this purpose, the e-divisive algorithm was used from the R package ecp (James and Matteson, 2013). E-divisive changepoint detection algorithm is a non-parametric method that slices the timeseries by detecting changes in the characteristic functions of the underlying distributions (that define a probability distribution) between segments. The method assumes that the α absolute moment (for $\alpha \in (0, 2]$) exists and that observations are independent. E-divisive is an iterative procedure where in each iteration one single changepoint that divides the timeseries into two segments that maximise the difference between the characteristic functions of the segments is detected. Subsequently, the statistical significance of the changepoint is evaluated based on a permutation test (James and Matteson, 2013). The procedure is repeated until the statistically significant CPs have been identified. In the implementation, 21 days (3 weeks) were selected as the minimum distance between possible CPs, in order to account for

seasonal variability. The confidence was defined equal to 95% in order to control the false positive rate of CPs.

4.3.3 *N_2O EF estimation based on changepoint detection*

Different realistic N_2O sampling scenarios were tested following the methodology described in Daelman et al. (2013a), to evaluate whether changepoint detection applied to historical data can reduce the required number of samples for the determination of N_2O EFs. For this purpose, the daily emission load was calculated (kgN_2O/day), for the first year of the monitoring campaign in the Northern Carrousel reactor (N_2O emissions dataset - DatN_2O).

Three different scenarios were considered; i) random 3-day monitoring between the CP intervals (total samples equal to 36 days/year) (sampling strategy 1 -ST1), ii) monitoring N_2O emissions for 3 random days each month for 1 year to account for seasonal variability of N_2O emissions (36 days/year) (sampling strategy 2 -ST2), iii) random monitoring for 36 days/year (sampling strategy 3 -ST3).

In all sampling strategies, it is assumed that emissions were monitored continuously for 24 h (starting from 00:01 a.m. of the chosen day). Emissions averaged over the 24 h periods, represent the daily average N_2O emitted (kg N_2O-N / d). In ST1, DatN_2O dataset was used to extract randomly 3 days from each CP interval (total 36 days). In ST2, 3 days were extracted randomly from each month (total 36 days), whereas in ST3 36 days were randomly selected over the whole DatN_2O dataset (total 36 days). Subsequently, the average N_2O emissions over these 36 days in all scenarios were estimated and were considered to represent the annual EF estimates. This procedure was repeated 10,000 times. Therefore, for each scenario, 10000 average annual N_2O emission loads were simulated, and a frequency histogram with the potential annual N_2O emission estimates was developed and compared with the observed average N_2O emissions.

4.3.4 *Feature extraction*

The data included in the analysis are characterised by 24-hour cyclical behavior, therefore 24 hours were selected as time interval. A 24-dimension vector was developed for each variable monitored in the system (representing hourly average). Subsequently, features were extracted, and a feature vector was developed, representing the behavior of the system; in total >100 features were extracted that can be grouped into three main categories (Figure 4.4). The first category consists of first-order statistical features including measures of central tendency (i.e. mean), measures of variability (i.e. standard deviation), measures of shape (kurtosis, skewness) and basic statistical functions such as daily maximum, minimum and interquartile range (IQR). First-order features are calculated using the real values of the timeseries and provide information on the diurnal behaviour of the variables. For instance, skewness is a measure of the data asymmetry around the mean and kurtosis is an indicator of the sharpness of the probability distribution compared to a normal distribution. The second category consists of second-order features calculated based on the differences between neighbouring values. The 24-dimension vectors ($y(t)$) were transformed based on the Equation (4.1) (Nanopoulos et al., 2001):

$$y'(t) = y(t + D) - y(t);\ 1 \leq D \leq n - D, \qquad (4.1)$$

Where D is the temporal distance between neighbouring points that are compared ($D = 1$). Subsequently, the mean and standard deviation were calculated from $y'(t)$.

The third group of the features was developed based on specific diurnal sub-events. It captures the behavior of operational variables under specific conditions. The intensity and presence of these events and patterns can provide information on the temporal behavior of the system. In the system, aerator 1 operates under on/off pattern (when ammonium is higher than 1.2 mg/l), while aerators 2 and 3 operate always and peak when ammonium is higher than 0.6 and 0.9 mg/l, respectively. Therefore, one subset of this group of features aimed at capturing the behavior of DO and nitrate concentrations when ammonium concentrations in the Carrousel reactor effluent was higher than 1.2 mg/L and lower than 0.6 mg/L. Additionally, the concentration of ammonium and nitrate in the plug-flow reactor

and the flow-rates provide an indication of the loads entering the Carrousel reactor. Subsequently, the second subset of features belonging to this group, aimed at capturing the diurnal duration of low (<10 mg/L) or high (>18 mg/L) ammonium concentration in the plug-flow reactor and the respective behavior of nitrate concentration. Calculated y′(t) values were also used to calculate a subset of features belonging to group 3. For instance, there are periods with a strong relationship between N_2O concentration and nitrate concentration in the Carrousel effluent. Therefore, the diurnal duration and strength of increasing/decreasing nitrate concentrations were calculated and linked with and the response of other variables in the system. A detailed list of the features tested is provided in the Appendix C (Table C.1).

Finally, Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) (Torres et al., 2011) was applied, to deconstruct the Temperature and NH_4-N concentration in the plug-flow reactor, into Intrinsic Mode Functions (lMFs) representing different oscillatory components. These variables were selected to investigate seasonal and cyclical components. CEEMDAN is an extension to Empirical Mode Decomposition (EMD) (Huang et al., 1998) and to the ensemble empirical mode decomposition (EEMD) (Wu et al., 2009). Based on EMD, the IMFs are constructed following two conditions: i) the number of extrema and zero crossings is either equal or differs by one at maximum and ii) the average value of envelope from local maxima and minima is equal to zero everywhere. EMD has been criticized for the "mode mixing problem" that occurs when similar oscillations are observed in more than one modes or when oscillations with dissimilar amplitude in the same mode (Wu and Huang, 2009). EEMD (Wu et al., 2009), constructs the IMFs by averaging the IMFs (constructed via EMD) over an ensemble of iterations; in each iteration the signal is augmented by the addition of different realisations of Gaussian noise with limited variance. However, there are still some limits with the application of EEMD (i.e. the noise added to the constructed signal). In CEEMDAN the noise is included to the remaining of the previous trial (Colominas et al., 2012; Torres et al., 2011). In each trial, the modes are constructed (from the original signal and white noise) independently from other realisations; the residues

from each iteration are different. Results from CEEMDAN decomposition are provided in the Appendix C (Figures C.1-C.2).

In this study CEEMDAN was used to extract the trend of the variables monitored in the system. Trend and variability of the variables are defined based on the study of Wu et al. (2007) following the methodology applied in Antico et al., (2014). Specifically, the variability consists of the modes with oscillatory characteristics less than 3 months; the trend is extracted by subtracting all the variability modes with oscillatory periods less than 90 days from the original variable. The frequency matching the peak of the raw periodogram is used to define the oscillatory period. The behaviour of wastewater treatment plant processes is affected by environmental factors (i.e. temperature); the exact seasonal trend is strongly linked with the local environmental conditions and on many occasions cannot be extrinsically derived. CEEMDAN is an adaptive approach based on information extracted from the raw data (Antico et al., 2014) and it can be useful to extract trends when analyzing wastewater treatment processes data. In the algorithm, the noise level was set to 0.02, the realisations to 1000 and the maximum sifting iterations to 1000. Subsequently, the aggregated daily mean of ammonium concentration and temperature were added to the set of features.

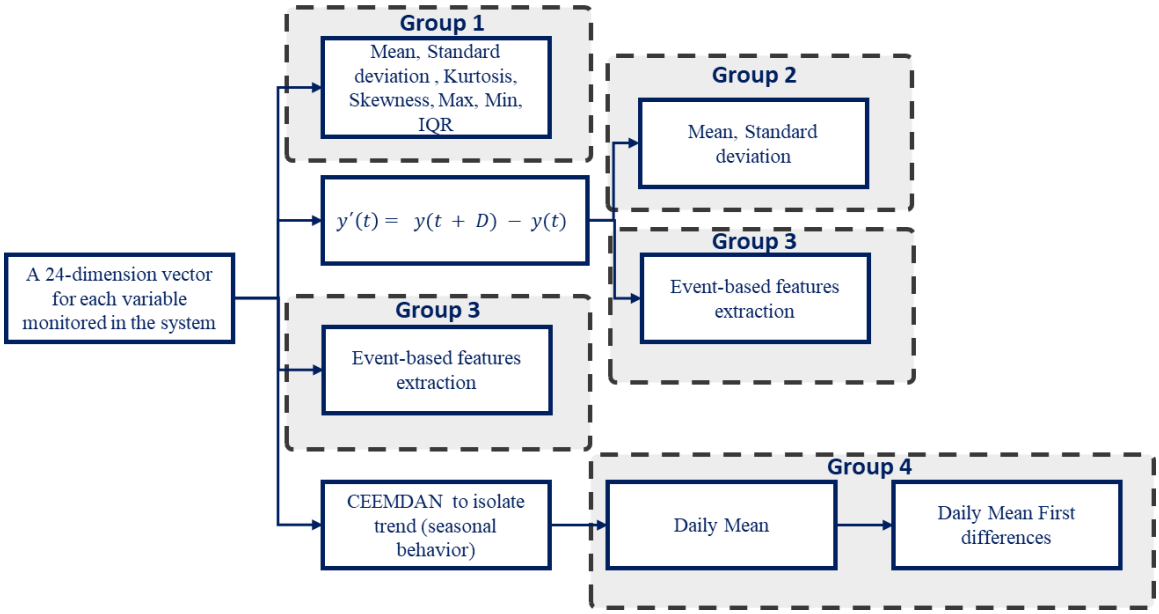


Figure 4.4: Feature extraction

4.3.5 *Feature Selection*

The objective of the feature selection step is to isolate feature-subsets, that can distinguish days belonging to different ranges of N_2O emissions. The supervised feature selection methodologies for classification tasks can be categorized into three groups: i) filter, ii) wrapper and iii) embedded (Kohavi and John, 1997; Saeys et al., 2007). Filtering models rank the features based on metrics of basic properties of the training dataset, such as distance and correlation (Tang et al., 2014) (i.e. Fisher score (Duda et al., 2001) or approaches based on mutual information (Peng et al., 2005)). Filtering approaches rank features separately from the classification algorithm, whereas wrapper approaches, use the predictive accuracy of a classification algorithm to evaluate feature importance (Liu and Motoda, 1998; Saeys et al., 2007). Classification algorithms with integrated feature exclusion during the training procedure, belong to the embedded feature selection methods (i.e. (Cawley et al., 2007).

Feature selection has been widely applied in environmental modelling, i.e. for groundwater quality monitoring (Rodriguez-Galiano et al., 2018; Tesoriero et al., 2017; Wheeler et al., 2015) and in renewable energy prediction problems (Salcedo-Sanz et al., 2018). In the wastewater sector, Tomperi et al., (2017) recently used five different features (i.e. based on forward selection, stepwise regression and genetic algorithms) together with a multivariable linear regression (MLR), to optimise the prediction of quality wastewater parameters from process measurements and high-resolution optical monitoring.

In this study, a recursive feature elimination (RFE) approach (Guyon et al., 2002) that implements backward elimination of features, wrapped with a standard random forest classification algorithm is applied to the feature vector for the selection of features.

Random forest classification (Breiman, 2001) is a nonparametric machine learning method where multitude of decision trees are constructed from a random subset of the features and trained in a bootstrap sample of the training set (consisting of around 2/3 of the data producing uncorrelated predictions); the final class prediction consists of the repeated outputs of these trees. It is one of the most powerful methods for feature selection and

classification. The prediction of the decision trees in the samples that were not included in the bootstrap sample (out-of-bag (OOB) samples) are compared with the prediction of the decision trees after permutation of the values of individual features. Significant features will alter the prediction results when shuffled, whereas non-significant features do not affect the tree outcome. The features' score in this study is derived based on the contribution of a specific feature in the misclassification rate of the forest when their values are permuted. Consider that t is the tree of each forest, OOB_{ti} is the sample and $errOOB_t$ is the misclassification rate for a specific tree t, $\widetilde{OOB_t}^j$ is the permuted OOB after random permutation of the values of a feature X^j , $err\widetilde{OOB_t}^j$ is the misclassification error of the tree in the perturbed sample $ntree$ is the total number of trees in the Random Forest, the raw feature X^j importance score over all the sum of all trees t, is given by Equation (4.2):

$$VI(X^j) = \frac{1}{ntree}\sum_t \left(errOOB_t - \ err\widetilde{OOB_t}^j\right)_t \tag{4.2}$$

The Z-score of the raw feature importance score (assuming normality), is calculated by dividing the raw importance score $VI(X^j)$ with the standard error of the raw feature importance score $\sigma^j \Big/ \sqrt{ntree}$ where σ^j is the standard error of the decrease in the number of correct class predictions due to permutation of feature X^j based on Equation (4.2):

$$C(X^j) = \frac{VI(X^j)}{\sigma^j \Big/ \sqrt{ntree}} \tag{4.3}$$

RFE algorithms belong to a cluster of methods where the least significant features are iteratively stripped towards minimizing a pre-defined stripping criterion (Kursa, 2014). It initiates with a random forest built utilizing all predictive features. During each iteration, features' importance is ranked based on Equation (4.3), the least important feature is eliminated, and the model is refitted and evaluated. Features with the strongest classification performance are used in the final model. The procedure is recursive because feature importance varies under different feature subsets over the stepwise feature elimination procedure (Granitto et al., 2006) The algorithm was implemented in caret package (Kuhn, 2008) in R software.

To evaluate the stability of the feature selection approach the following ensembled feature selection methodology was applied (shown in Figure 4.5), adapted from Abeel et al., (2010). Initially, the complete dataset was randomly split into two sets; the training and the test set (75/25) and RFE-RF was used for feature selection. In order to generate ensembles of feature rankings from the training set, a repeated 10-fold cross validation was performed with repetitions equal to 3. Feature rankings for the training set were estimated based on the complete linear aggregation method (Abeel et al., 2010). Considering an ensemble with s feature selectors $E = \{F_1, F_2, \dots, F_s\}$; each F_i gives a feature ranking equal to $f_i = (f_i^1, f_i^2, \dots, f_i^N)$ where N is the number of features and the different feature rankings acquired for the training set by the repeated k-fold cross validation are aggregated into one feature ranking, calculated based on Equation (4.4):

$$f = \left(\sum_{i=1}^{s} w_i f_i^1, \dots, \sum_{i=1}^{s} w_i f_i^N \right) \tag{4.4}$$

where, w_i is a weighting function. In the complete linear aggregation method, the aggregated feature ranking f was calculated as the sum of all feature ranks acquired over the repeated k-fold cross validation samples. Therefore, w_i was set equal to 1.

The procedure was repeated 50 times for different splits of training and test sets. In order to acquire unbiased performance estimates the following procedure was followed. During each train and test data set splits, the final feature selection was based on the feature subset that minimised the prediction error. The number of features selected ranges from 5 to 12 and a RF classification model was trained consisting of 1000 trees. The prediction accuracy was tested on the validation set. Additionally, the selected features were used to build and compare an SVM classification model using the same training and test sets. This procedure was repeated 50 times for the different training and test dataset splits.

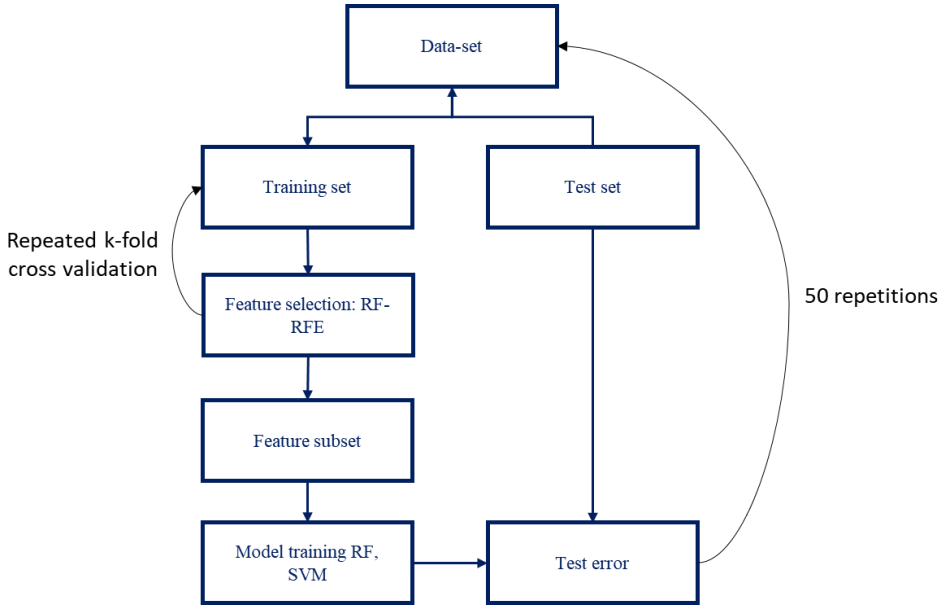


Figure 4.5: Procedure followed to evaluate the stability of the feature selection and performance of the SVM and random forest classifiers

4.3.6 *Support Vector Machine classification*

SVMs were initially developed by Cortes and Vapnik, (1995) for binary classification. Briefly, the SVM method is outlined first for the binary linearly separable classification case. In the binary classification model two classes exist and an object (i.e. set of features extracted from sensor measurements) belongs to one of these classes. Considering a d-dimensional N set of vectors x_i, with $x_i \in \mathbb{R}^d$ where $i = 1,\ldots,N$. Each set is identified to belong to class y_i with $y_i = 1$ for one class and $y_i = -1$ for the other class, $y_i \in \{-1,1\}$. If the two classes are linearly separable, then a family of linear separators exists, also called separating hyperplane. The hyperplane is defined as:

$$\mathrm{w}^T x + w_0 = 0 \tag{4.5}$$

where w is the weight vector with dimensions equal to the dimensions of x_i and $w_0 \in \mathbb{R}$ is the bias (a scalar). The weight vector w and the bias w_0, determine the separating hyperplane location. The equation of the hyperplane defines the function $f(x) = \mathrm{sign}(\mathrm{w}^T x + w_0)$, which represents the output of the algorithm for a new point x_i. Support vectors are defined as the new points x_i that satisfy the constraint $f(x_i) = +1$ for $y_i = 1$

and $f(x_i) = -1$ for $y_i = -1$. Other x_i points not defined as support vectors are equal to $f(x_i) > +1$ for $y_i = 1$ or $f(x_i) < -1$ for $y_i = -1$. Therefore, the constraints for all x_i points can be defined as (Ekici, 2009):

$$y_i f(x_i) = y_i(w^T x_i + w_0) \geq 1, \qquad i = 1, 2, \ldots, N \tag{4.6}$$

Assuming that a separating hyperplane exists, natural classifiers are constructed assigning the test observations to a class based on their location in the hyperplane. The margin of a separating hyperplane is the minimum distance of any data point x_i to the hyperplane and can be expressed as $2/\|w\|$. The support vectors are the observations with the minimum distance to the decision boundary. Support vector machines select the 'maximum margin hyperplane'; the separating hyperplane that has the farthest minimum distance to the observations x_i. Classes with large margins are clearly separable and provide a 'safety' for the generalisation of the algorithm when applied to new points. In practical applications, the overlapping of a number of data belonging to the two classes, is common. Therefore, soft margins are introduced (C regularization parameter, ξ_i the slack variable) to allow a number of misclassifications to identify feasible solutions when the training dataset is not strictly linearly separable. This is equivalent to:

$$Minimize \ \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N} \xi_i \tag{4.7}$$

$$\text{subject to } y_i(w^T x_i + w_0) \geq 1 - \xi_i, i = 1,2, \ldots, N$$

$$and \ \xi_i \geq 0, i = 1,2, \ldots, N$$

Where the regularization parameter C is a positive constant and ξ_i represents the distance of data x_i from the decision boundary.

The solution of this problem is the saddle point of the Lagrangian:

$$L_p = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\xi_i - \sum_{i=1}^{N}\alpha_i\,(y_i(w^T x_i + w_0) - 1 + \xi_i) - \sum_{i=1}^{N}\beta_i\xi_i \tag{4.8}$$

Where the coefficients α_i and β_i are the Langrage multipliers (constraint to be ≥0). The minimisation of the Lagrange function derivates based on w, w_0 and ξ will lead to an optimum solution of Equation (4.8). After substitution and simplification, the final form of the linear classifier is:

$$f(x) = \text{sign}\left(\sum_{i=1}^{N_{SV}}\alpha_i\, y_i(x_i^T \cdot x) + w_0\right) \tag{4.9}$$

Where: N_{SV} is the number of support vectors. This last equation allows classifying a new vector x unknown in the training database.

In cases where linear separation is not possible, the d-dimensional input vector x is mapped to the d_h −dimensional feature space, where the linear separation of the input data is feasible, via a function $\phi(x) : \mathbb{R}^d \rightarrow \mathbb{R}^{d_h}$. Subsequently, Equation (4.9) is defined as:

$$f(x) = \text{sign}\left(\sum_{i=1}^{N_{SV}}\alpha_i\, y_i(\varphi(x_i)^T \cdot \varphi(x)) + w_0\right) \tag{4.10}$$

An appropriate kernel function K(·,·) can be selected for the mapping. When the Mercer's condition is satisfied, then $K(x, x_i) = \varphi(x_i)^T \cdot \varphi(x)$ and an explicit construction of φ(x) mapping is not required (and thus to calculate the data coordinates) since solely the inner product of the data point mappings in the feature space K(·,·) is only required for the optimisation. This is commonly called the 'kernel trick' and enables SVMs to operate even in possibly infinite feature space (where data are mapped), without in practice executing calculations there (Luts et al., 2010). There are several kernel functions used (linear, polynomial, sigmoid and radial basis function (RBF)). One of the most widely used kernel functions is the RBF (Singh et al., 2011) that is expressed as:

$$\mathrm{K}(x, x_i) = exp(-\gamma \left\| x_i - x_j \right\|^2) \quad (4.11)$$

Where, γ, determines the kernel amplitude and is connected to the model's generalisation capabilities.

One versus one approach was applied to classify into more than two classes. K SVMs are fitted, each time comparing one of all the K classes to the remaining K − 1 classes. All SVMs were trained using RBF kernel. SVM models were trained using repeated 10-fold cross validation (3 repetitions) to select the cost and γ parameters. The dataset was randomly divided into test and train, with 75% of the available data used for training the SVM model and 35% used for testing.

4.3.1 *Model performance evaluation*

The performance of the classification SVM models were evaluated based on the accuracy, kappa shown below:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TP + FN} \quad (4.12)$$

$$\text{Kappa} = \frac{\left(Accuracy - \frac{\left(\frac{((TP + FP) * (TP + FN))}{(TP + FP + TN + FN)} + \frac{((TN + FN) * (TN + FP))}{(TP + FP + TN + FN)} \right)}{(TP + FP + TN + FN)} \right)}{\left(1 - \frac{\left(\frac{((TP + FP) * (TP + FN))}{(TP + FP + TN + FN)} + \frac{((TN + FN) * (TN + FP))}{(TP + FP + TN + FN)} \right)}{(TP + FP + TN + FN)} \right)} \quad (4.13)$$

Where: TP is the number of true positives, FP false positives, FN false negatives, and TN true negatives.

4.4 Results and discussion

4.4.1 *Detection of abnormal events*

DBSCAN was applied to detect events that affect influent flow-rate and NH_4-N concentration in the effluent of Carrousel reactor as described in section 4.3.1. Table 4.1 shows the features that were considered in DBSCAN; these were selected in order to represent the diurnal behaviour of the target variables. The eps parameter was determined based on the "knee" of k-nearest neighbours of the data plotted in increasing order (Appendix C, Figure C.3). Based on this procedure *MinPts* and *eps* were set equal to 6 and 0.4 respectively.

Table 4.1: Features selection

Features selected
Average influent flow-rate
Average NH_4-N concentration in the Carrousel effluent
Maximum daily influent flow-rate
Maximum NH_4-N concentration in the Carrousel effluent
Average daily standard deviation of the first derivative of influent flow-rate
Average daily standard deviation of the first derivative of NH_4-N concentration in the Carrousel effluent
Kurtosis influent flow-rate
Kurtosis NH_4-N concentration in the Carrousel effluent

The behavior of the influent flow-rate and effluent NH_4-N concentration, in the days isolated by DBSCAN are shown in Figure 4.6. In total 155 days were isolated and are mainly characterised by events with elevated influent flow-rate or/and peaks of the NH_4-N concentration in the Carrousel effluent. Subsequent inspection showed that these events varied in intensity and duration; therefore, they were categorized into three major groups. Group 1 consists of system disturbances with duration equal or less than 24 h. Days belonging to group 1 are characterised by elevated influent flow-rate (days with precipitation) and peaks in the effluent NH_4-N concentration during the same day and thus, low removal efficiency of NH_4-N. In group 1, the system resumes to normal NH_4-N removal efficiency after 24h. Group 2 consists of system disturbances lasting more than

24h. Multiple days with precipitation, at close temporal proximity, that affect the system performance for several days were assigned to this group. Finally, elevated influent flow-rate events lasting less than 24h, affected significantly the behavior of NH_4-N concentration in the Carrousel effluent for several days. These occasions were assigned to group 3. In total, 54 different events were detected (155 days) that belong to one of the three groups.

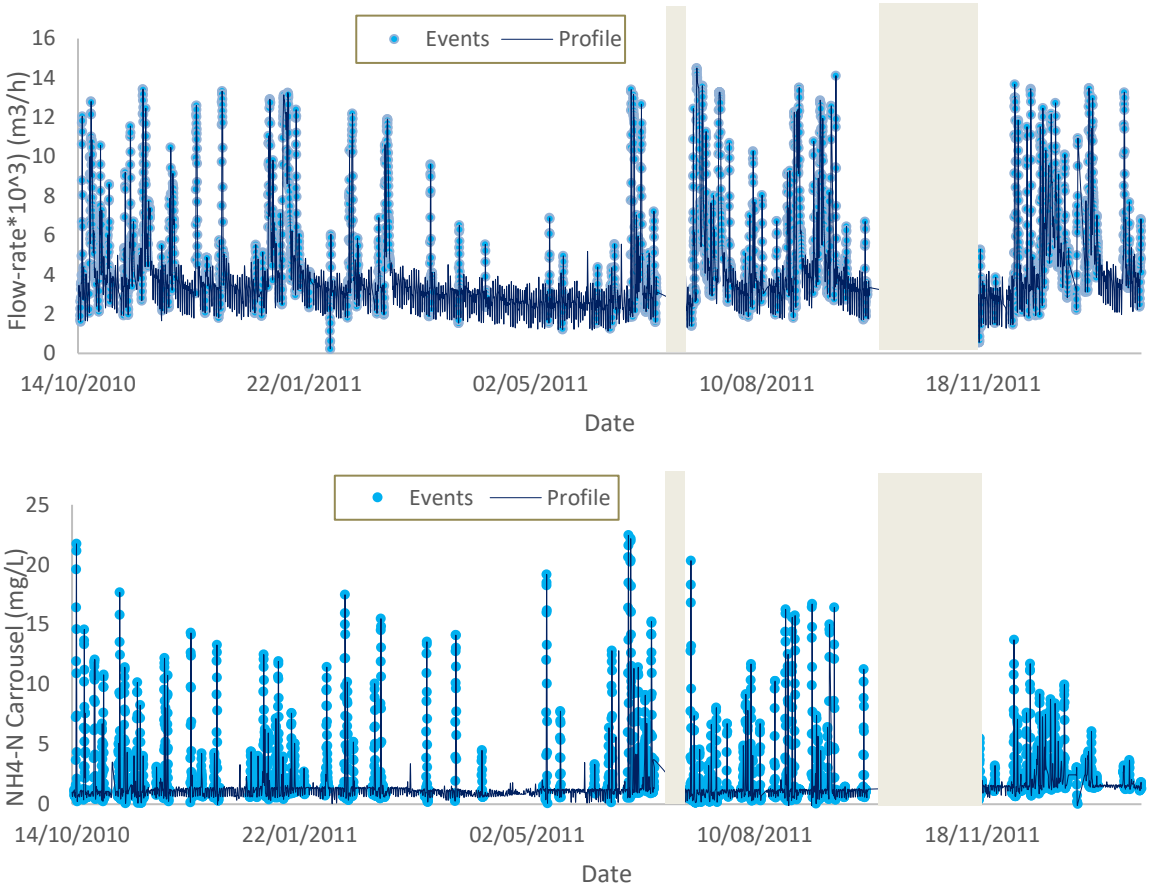


Figure 4.6: Influent flow-rate (top) and NH_4-N concentration in the Carrousel effluent (bottom) profiles (blue lines) and events (light blue points) isolated by DBSCAN

Figure 4.7 shows examples of the events belonging in groups 1, 2 and 3 and the behavior of N_2O emissions. Blue lines represent the events detected by DBSCAN and red lines represent the normal operational conditions. Overall, ~30% of the events belong to group 1. The average daily influent flow-rate is ~4000 m^3/h; therefore, days in group 1 have moderate increase of the influent flow-rate (flow-rate peaks <7000 m3/h and NH_4-N concentration in the Carrousel peaks< 6 mg/L). Figure 4.7 (a-d) shows an event with the

highest influent flow-rate peak; N_2O emissions are not significantly affected. Overall, the behavior of N_2O emissions of 1-day events at temperatures between 12-16 °C is not significantly affected compared to the average behavior of emissions the day prior to the event (Figure 4.7 (d)). However, significant peaks, of N_2O emissions, coinciding with group 1 events, are observed at higher temperatures. Days belonging to group 2 are characterised by influent flow-rate peaks above 8000 m^3/h, whereas NH_4-N concentration peaks are higher than 9 mg/L in the effluent of the Carrousel reactor (Figure 4.7 (e-h)). Again, the behavior of N_2O emissions varies. However, emissions tend to drop after the peak of the influent flow-rate (NH_4-N concentrations in the plug-flow<8 mg/L). Finally, most peaks of N_2O emissions between June and September 2019 belong to group 3 (Figure 4.7 (e-h)).

The categorization of the system operational behavior based on historical reoccurring disturbances can help operators to identify events that can significantly impact the performance of the system and apply mitigation strategies (i.e. regulate the anaerobic supernatant stream to the mainstream line to reduce the system loads).

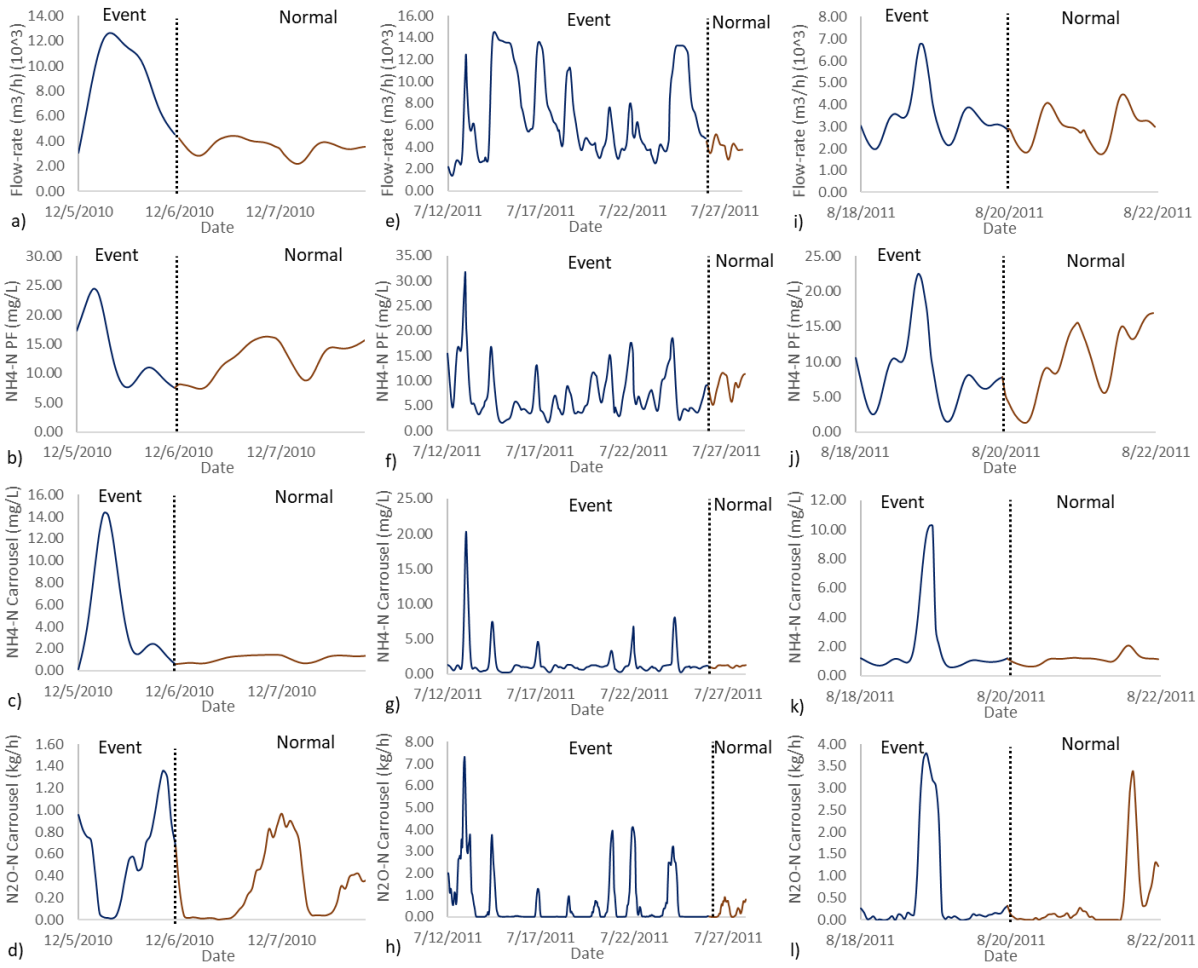


Figure 4.7: Flow-rate, NH_4-N concentrations in the plug-flow and Carrousel effluent and N_2O emissions' profiles for days detected by DBSCAN belonging to events group 1 (a-d), group 2 (e-h) and group 3 (i-l)

4.4.2 *Changepoint detection*

Changepoint detection was applied in order to identify the changes in the profiles of the operational variables that affect the performance of the Carrousel reactor. Daily averages of i) NH_4-N load (kg/h) in the plug-flow reactor (as an indication of the influent ammonium in the Carrousel reactor), ii) NH_4-N load (kg/h) in the effluent of the Carrousel reactor, iii) DO1 and DO2 concentrations were used. Daily averages of these variables were considered in order to avoid the diurnal cyclic characteristics of the variables. The volumetric flow-

rate of the plug-flow and Carrousel reactors were calculated based on Daelman et al., (2015).

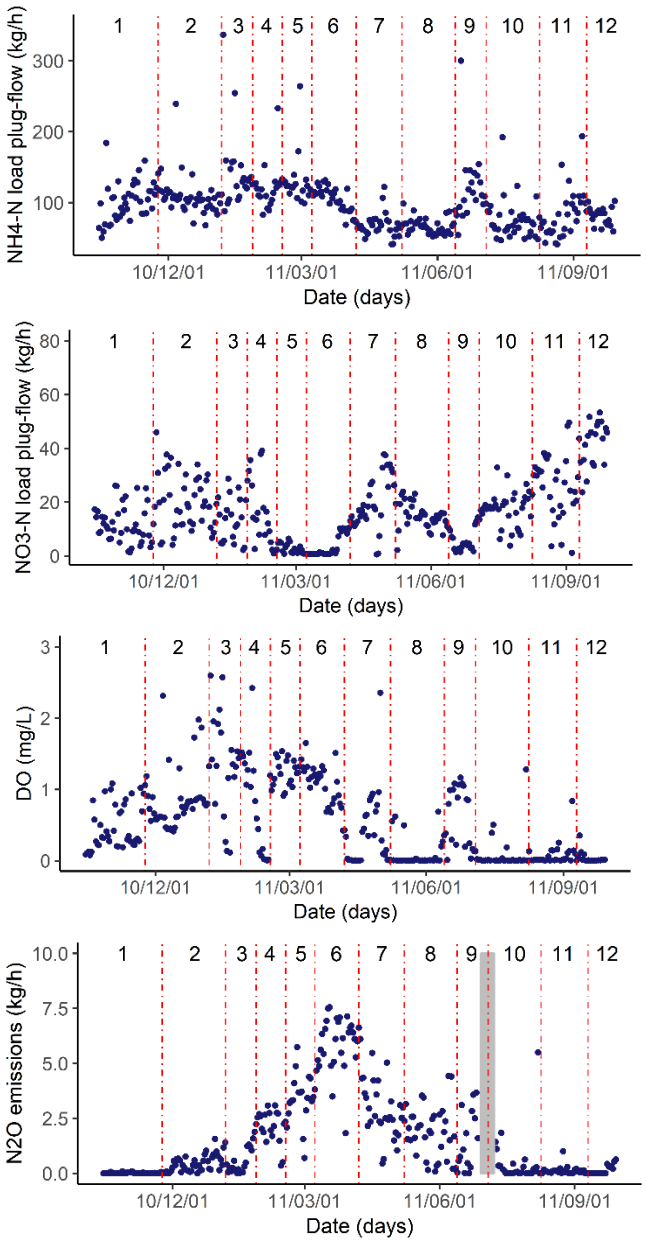


Figure 4.8: CPs intervals for a) NH_4-N load (kg/h), b) NO_3-N load (kg/h), c) DO1 and DO2 average concentration, d)N_2O-N emissions behavior between CP intervals

The minimum transition interval was equal to 21 days (~3 weeks) on the assumption that biological processes can be affected by seasonality. The multivariate changepoint detection

analysis identified 12 statistically significant CPs (with significance level <<0.05) for the first year of the monitoring campaign. Figure 4.8 shows the identified CPs and the respective profile of N_2O emissions for each period interval between CPs. On many occasions, CPs coincide with the changes of the N_2O emissions profile during the monitoring period. For instance, the highest drop in the average N_2O emissions between adjoining periods (CPs 6 and 7) coincide with a drop in the ammonium load in the plug-flow reactor and an increase in the average nitrate-nitrogen load in the plug-flow reactor. Similarly, the drop of N_2O emissions between CPs 2 and 3, coincides with an increase in the DO1 concentration in the Carrousel reactor.

4.4.3 *Accuracy of the monitoring strategy based on system CPs*

As shown in section 4.4.2, the behavior and range of N_2O emissions changes between the detected CP intervals. Quantification of reliable N_2O EFs in wastewater treatment processes is still not straightforward; monitoring campaign duration and strategy can significantly impact the reliability of the results. Seasonal effects have also significant impact on N_2O emissions (Vasilaki et al., 2019). Daelman et al., (2013a) simulated different sampling strategies, based on the data collected from the same plug-flow – Carrousel system linking EFs estimates with different sampling strategies. The authors found that long-term offline/online sampling capturing seasonality and temperature effects, is significant for reliable EF assessment. The aim of this analysis is to simulate a knowledge-based N_2O sampling campaign between CPs and evaluate EF estimates following a similar approach to the study of Daelman et al. (2013a). Additionally, the knowledge-based sampling campaign is compared with two alternative monitoring strategies: i) random 24-h monitoring and ii) random 24-h monitoring for specific days at each month capturing the seasonal variability.

Figure 4.9 (a) shows the relative frequency histogram of the estimated annual N_2O load (kg/day) when 24h sampling for 36 random days is applied, during the first year of the monitoring campaign (n=10000 repetitions). The red vertical line represents the measured average annual N_2O load (equal to 39~ kg/day). In total 43% of the simulations resulted

to an EF ranging between 35 and 43 kg/day (less than 10% error from the actual annual N_2O load quantified by the monitoring campaign). Additionally, the probability to underestimate the N_2O load by more than 10% is equal to ~30%. Figure 4.9 (b), shows the histogram of the estimated annual N_2O load (kg/day) when 24h sampling for three random days between the CPs (12 CP intervals) was assumed (n=10000 repetitions). Overall, the likelihood to estimate an average N_2O load between 35 and 43 kg N_2O/day was equal to ~80%, with >99% of the simulated N_2O estimates ranging between 32 and 46 kg/day. In this case, the probability to underestimate the emissions by more than 10% was approximately 5%. Finally, when random sampling for 3 days per month was tested (Figure 4.9 (c)), the probability to estimate an N_2O load ranging between 35 and 43 kg/day, was equal to ~70%, whereas the probability to underestimate the emissions by more than 10%, was approximately 25%.

The behaviour of the operational variables needs to be considered together with seasonal effects when sampling campaigns are planned. In the investigated system, limited sampling days between the CPs could give a realistic quantification of the actual EF during a whole year. The proposed approach can be applied to identify N_2O emissions "hotspot" periods and guide towards the identification of the operational periods that require intensive investigation of N_2O pathways and mitigation measures.

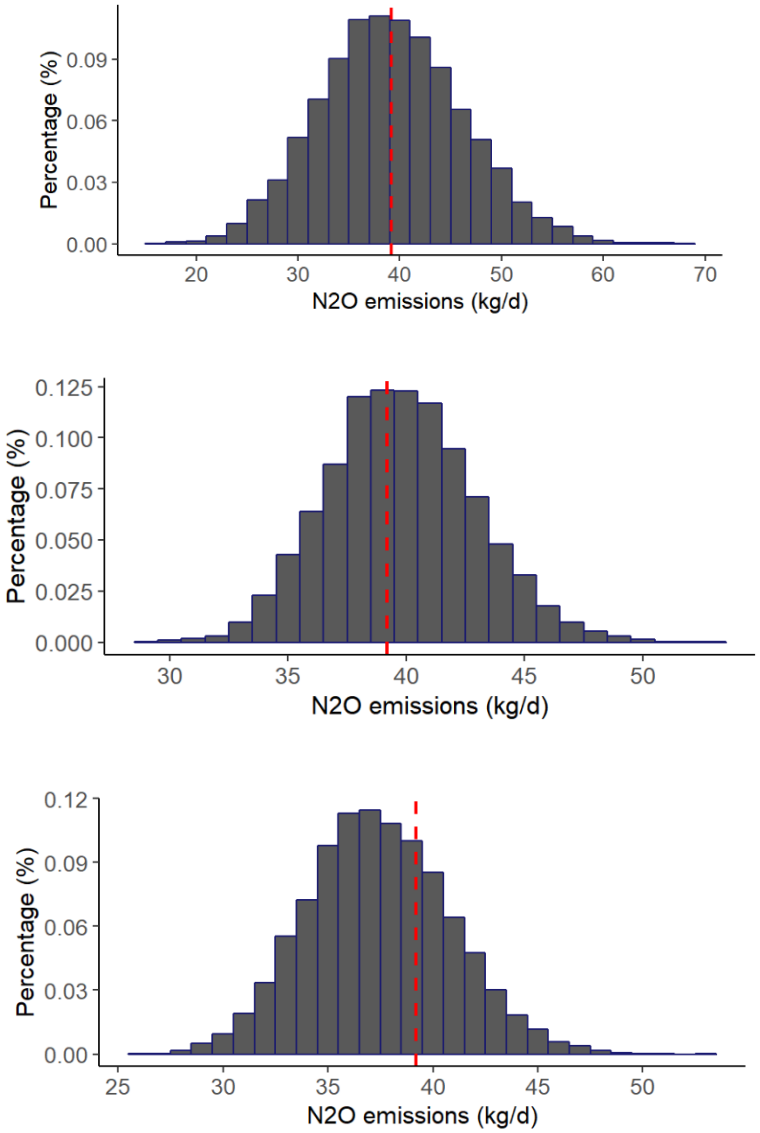


Figure 4.9: Relative frequency histogram of the estimated average annual N_2O load for (a) a simulated monitoring campaign with duration equal to 36 random days, (b) a simulated monitoring campaign with 3 random days between CP intervals and (c) a simulated monitoring campaign with 3 random days between different months for 1 year

4.4.4 *Feature selection and classification results*

A classification algorithm was constructed to predict low, medium or high N_2O emissions based on the operational behaviour of the system. The categorization of the different

classes was based i) on the CP analysis, ii) on the seasonal effects. Therefore, two periods (from the CP intervals) characterised by similar N_2O ranges but not sequential, were assigned in two different classes. Table 4.2 shows the average N_2O emissions in each class and the changepoint intervals for each class (the changepoint intervals are shown in Figure 4.8). However, five only classes were considered in the feature selection and construction of the classification algorithms.

Table 4.2: Classes considered in the classification based on the CP intervals

Class	N_2O emissions (kg/d)	Changepoint intervals	N_2O emissions level
1	8.3 (±8.3)	1, 2, 3	Low
2	68.3 (±23.2)	4, 5	Medium
3	141.7 (±31.6)	6	High
4	58.9 (±25.0)	7, 8, 9	Medium
5	7.2 (±7.9)	10, 11, 12	Low

Feature selection is a significant step of several high-dimensional classification applications (Tang et al., 2014). However, many studies have shown that selected features depend significantly on the training sample, and thus, a feature selection algorithm can be unstable (He and Yu, 2010; Kalousis et al., 2007). Therefore, in many cases feature selection stability needs to be considered together with model prediction accuracy in the evaluation of the classification/regression performance (Pes et al., 2017; Saeys et al., 2008).

Figure 4.10, shows the most common features for feature subset sizes equal to 6 and 10 for the resampling perturbations. Overall, 4 features coincided in all subsets with feature size equal to 6 and 8 features coincided in almost all subsets with feature size equal to 10. The selected features are divided between feature group 1 (i.e. first-order statistical features - maximum DO2, minimum influent flow-rate), feature group 3 (that capture the behavior of operational variables under specific conditions) and feature group 4 (i.e. trend extracted by CEEMDAN). For instance, the features describing the behavior DO2 concentration for NH_4-N higher than 1.2 mg/L in the Carrousel effluent were selected in all feature subsets. The trends of NH_4-N concentration in the plug-flow reactor and temperature extracted by

CEEMDAN, were also included in all feature subsets. Finally, the complete description of the features is provided in the Appendix C (Table C.1).

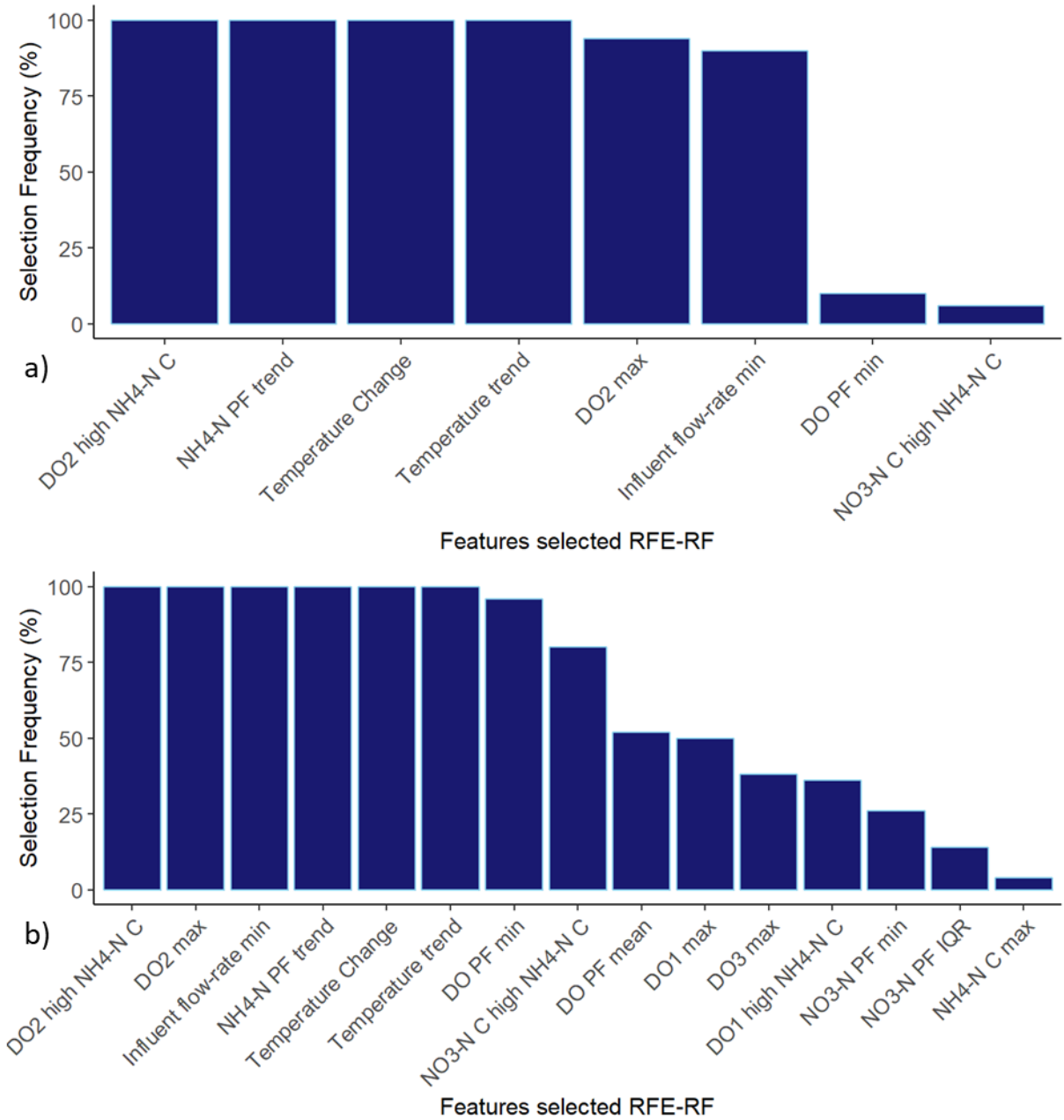


Figure 4.10: Features selected for subset sizes equal to (a) 6 and (b) 10. Details for the selected features are provided in the supplementary material Appendix C (Table C.1). (max: maximum, min:minimum, IQR: interquantile range, DO2/DO1/NO_3-N high NH_4-

N C: average DO2/DO1/NO_3-N concentration in the Carrousel for the hours NH_4-N concentration in the Carrousel effluent>1.2 mg/L)

The results of the SVM and RF classification models from the different resampling perturbations for both the train and test data sets are shown in Table 4.3. Feature subsets that minimise the classification error were selected in each resampling. The results show similar behavior both for the RF and SVM classifiers, whereas classification accuracy in the test dataset is high even for small feature subsets (> 97%). In 58% of the resampling perturbations the size of the best feature subset was equal to 6 (with the variables selected shown in Figure 4.10). In total, 8 variables were selected for 14% of the resampling perturbations, 10 for 20% of the resampling perturbations whereas all the features were selected for 4% of the resamples.

Table 4.3: Evaluation of SVM and RF classifiers for different feature subset sizes

Feature subset size	Model	Dataset	Accuracy (%)	Kappa (%)
6	RF	Train	97.4 (± 1)	96.6 (± 1)
		Test	97.4 (± 2)	96.7 (± 4)
	SVM	Train	98.9 (± 1)	98.6 (± 1)
		Test	95.1 (± 3)	93.7 (± 4)
8	RF	Train	97.3 (± 2)	96.6 (± 2)
		Test	98.2 (± 2)	97.7 (± 2)
	SVM	Train	98.7 (± 1)	98.4 (± 2)
		Test	95.9 (± 2)	94.8 (± 2)
10	RF	Train	97.4 (± 3)	96.6 (± 3)
		Test	98.3 (± 1)	97.7 (± 2)
	SVM	Train	99.0 (± 1)	98.8 (± 2)
		Test	96.3 (± 3)	95.2 (± 4)

Subsequently, data from the last period of the monitoring campaign were tested to assess the predictive capabilities of the models in previously unseen operational periods. In total ~30 days were tested (precipitation events were not considered). Figure 4.11 displays the predicted classes (SVM classification) for different best feature subset sizes (majority rule for the different resampling perturbations) and the daily N_2O emissions. In the first 20 days all models have predicted low emission-risk classes. Class 1 represents operational conditions from the beginning of the monitoring campaign, whereas class 5 represents

conditions from the end of the monitoring campaign (after ~1 year). The investigated period is ~ 1 year after the start of the monitoring campaign and the observed N_2O emissions are again low. Therefore, the alternation of the predicted classes between class 1 and 5 can be expected, mainly due to the impact of seasonal effects on the influent concentration and on N_2O generation. Finally, the model with best feature subset size equal to 10 was the only one able to detect the change in the N_2O range after the first 20 days.

Additional data are required to investigate the generalisation capabilities of the SVM classifier. The methodological approach followed is able to predict the range of N_2O emissions, as long as the system operates within the predefined and investigated range.

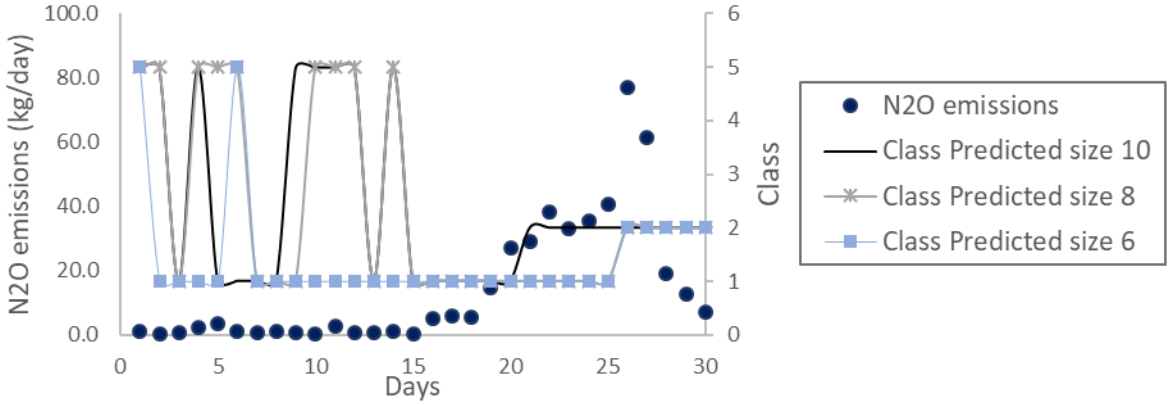


Figure 4.11: Daily N_2O emissions and predicted classes (SVM classification) for the last period of the monitoring campaign (data not used during changepoint detection)

Under the investigated conditions, the accuracy of the classifier with feature subset size equal to 10, was satisfactory, even when data, from the second year of the campaign that were not used during training (30 days) were tested. Therefore, the SVM classifier can be used (with caution) to detect periods with operational behaviour that has been historically linked with elevated emissions.

The development of mitigation measures in the predicted high-risk N_2O emission periods, can be supported with the integration of mechanistic models or practical, simplified theoretical approaches. The latter will facilitate the identification potential triggering mechanisms linked with the period-specific operational conditions. For example, a simplified N_2O risk-based model, was developed by Porro et al. (2014) considering

thresholds of ASM state variables linked with the generation of N_2O emissions (i.e. DO, nitrite, COD:N) based on the treatment step (nitrification, nitrification, transition zones).

Additional, long-term monitoring campaigns, in continuous wastewater systems are required, to validate the proposed strategy and standardise the selection of operational variables that need to be considered during changepoint detection and classification. Finally, the development of a novelty detection approach needs also to be integrated in the procedure, that will detect new, unobserved operational states and provide feedback to the algorithms on re-calibration requirements.

4.5 Summary of main findings

This study shows that information hidden in conventional variables monitored in wastewater can be mined to reduce N2O sampling frequency without compromising the quantification of annual N2O EFs and ultimately predict the risk of elevated emissions.

The main findings of this chapter are summarized below:

- The isolation and categorization of re-occurring system disturbances (i.e. precipitation events) showed that environmental conditions and events with increased flow-rate affect the system response in terms of N_2O emissions and NH_4-N removal efficiencies.

- The application of changepoint detection in the process operational variables provided insights on structural changepoints of the N_2O emissions profile. In total, 12 statistically significant CPs were detected for the first year of the monitoring campaign. Abrupt decreases of the N_2O emission profile were linked with drops in the ammonium load, increase in the nitrate-nitrogen load of the plug-flow reactor and increase in the DO1 concentration of the Carrousel reactor

- Limited 24-h N_2O samples between the CP intervals are sufficient to estimate the average N_2O EF for the whole year, while conventional strategies resulted in lower accuracy of the N_2O EF.

- An SVM classification model was constructed to predict operational periods linked with specific N_2O emission ranges. The results indicate that analysis of historical data and investigation of seasonal effects can be of paramount importance in the planning of monitoring campaigns. The proposed approach can be applied when long-term online sampling is not technically and economically feasible. The proposed solution is capable of pinpointing the N_2O emissions "hotspot" periods and guiding towards the identification of operational periods that require extensive investigation of N_2O pathways and mitigation measures.

5. A knowledge discovery framework to predict the N_2O emissions in the wastewater sector

5.1 Introduction

In recent years the sustainability and operational efficiency of wastewater treatment plants (WWTPs) have come to the fore (Liu et al., 2018). Several biological technologies such as partial-nitritation – anammox (anaerobic ammonium oxidation) have emerged, towards the efficient, low-cost treatment of high-strength municipal wastewater streams (Lackner et al., 2014; Zhou et al., 2018). The anaerobic supernatant is a by-product of dewatering of the anaerobic digestion effluent and represents less than 1-2% of the total influent flow in the WWTP. It contains 10–30% of the N load and 20–30% of the P load (Janus and van der Roest, 1997; van Loosdrecht and Salem, 2006). Sidestream treatment of the anaerobic supernatant can contribute to the reduction of energy consumption for N-removal, decrease of nitrogen loads in the secondary treatment, and the minimisation of risks related to exceeding effluent regulatory requirements of nitrogen concentrations in the water line of WWTPs (Eskicioglu et al., 2018). However, the performance and environmental evaluation of different sidestream technologies is still under investigation (Eskicioglu et al., 2018; Rodriguez-Garcia et al., 2014).

SCENA (Short-Cut Enhanced Nutrient Abatement) is a new sidestream process, that combines the conversion of NH_4^+ to NO_2^- under aerobic conditions (nitritation) with the subsequent reduction of NO_2^- to nitrogen gas and enhanced biological phosphorus uptake by polyphosphate-accumulating organisms (DPAOs) in a sequencing batch reactor (SBR) (Frison et al., 2015). External volatile fatty acids (VFAs), are produced via acidogenic fermentation of the primary and secondary sludge on-site and dosed into the SBR. In a recent study, Longo et al. (2016), quantified the environmental and cost benefits and impacts of the integration of the SCENA process in a full-scale WWTP. They reported major energy savings for aeration after the integration of sidestream SCENA process. The direct N_2O emissions were equal to 1.42% of the influent N-load. Short-term monitoring

campaigns were implemented, while the effect of operational conditions on N_2O generation was not investigated.

N_2O is a potent cause of global warming, its global warming potential is 265 - 298 times more than that of CO_2 (IPCC, 2013). The emission of N_2O in full-scale sidestream partial-nitritation/partial-nitritation–anammox or nitrification-denitrification systems range from 0.17% to 5.1% of the influent N-load (average equal to ~2.1% of the N-load is emitted (Vasilaki et al., 2019). Schaubroeck et al. (2015) showed that N_2O emissions from a full-scale sidestream DEMON process in Austria were significantly higher than the direct N_2O emissions from the mainstream treatment in a full-scale WWTP. On average, 0.256 g N_2O were emitted compared to 0.005 g emitted in the secondary treatment per m^3 treated wastewater. The increased direct N_2O emissions can be mainly attributed to low DO concentrations, higher ammonia oxidation rates (AOR) and NO_2^- build-up (Desloover et al., 2011; Kampschreur et al., 2008b); conditions that also prevail in the SCENA process. The variability of EF reported in sidestream technologies can be partially attributed to both complex relationships between emitted N_2O and operational conditions and different configurations (i.e. SBR, continuous systems), loads (i.e. NH_4^+ concentrations), feeding strategies and operational control (i.e. DO set-points). Additionally, different interactions between operational variables trigger a different response of N_2O generation. For instance, in a recent modelling study of a granular one-stage partial-nitritation-anammox reactor, Wan et al. (2019) showed that higher temperatures resulted in increased N_2O emissions in the presence of COD (chemical oxygen demand) and in decreased N_2O emissions in the absence of COD (due to increased anammox activity and reduction of NO_2^- accumulation in higher temperature). Additionally, the long-term temporal variations of direct N_2O emissions were not adequately assessed in sidestream technologies; the majority of the monitoring campaigns in sidestream reactors lasted less than 5 days (Vasilaki et al., 2019).

The digitalisation of water services and the data-driven knowledge discovery from wastewater treatment plant may increase the resilience of water utilities under climate change and other water-related challenges (Sarni et al., 2019). Recent studies have provided extensive overviews of the use of data-driven techniques in the wastewater sector

for different applications including the development of soft-sensors, fault prediction and multi-objective optimisation of control of water utilities (Corominas et al., 2018; Haimi et al., 2013; Newhart et al., 2019). Data-mining and extraction of the information hidden in the raw sensor signals can facilitate the identification of patterns and hidden structures and reveal significant information on the behaviour of N_2O emissions in continuous wastewater treatment processes (Vasilaki et al., 2018). The SBR in the SCENA process is multiphase (i.e. anaerobic, aerobic, anoxic conditions) applying different operational variables (unsynchronised data), non-linear and subject to different disturbances, such as influent compositions and fermentation liquid characteristics. Moreover, SBR process data are based on a 3d-structure that consists of the number of i) cycles, ii) variables and iii) sampling points within each cycle. Therefore, the identification of process abnormalities and patterns can be complicated. N_2O emissions could be affected by both within-cycle and between-cycle batch dynamics.

In this study, sensor and laboratory analyses data from a full-scale SCENA SBR were analysed to provide insights on the N_2O emissions behaviour and generation. A structured approach was followed for knowledge discovery from the available dataset using a combination of abnormal events detection, classification and regression techniques. The objectives of the study were to i) investigate whether the sensors integrated in the system (i.e. conductivity, pH) can provide actionable information on the dynamics of N_2O emissions, ii) detect hotspots for the accumulation and emission of N_2O and iii) develop data-driven regression and classification models to predict the dissolved N_2O behaviour and concentration for the different phases (anaerobic, aerobic, anoxic) of the SBR.

5.2 Process description and data origin

The Carbonera plant is designed to treat domestic wastewater of a population equivalent of 40,000 (dry weather flow equal to 10,000 m^3/d). After screening and degritting and primary sedimentation, the effluent from the primary clarifier is sent to a Schreiber reactor (single basin – working volume 4671 m^3). The Schreiber process consists of a continuously sequencing reactor (CSR) favouring simultaneous nitrification-denitrification via time-

based intermittent aeration. Schreiber reactor effluent is pumped to two secondary clarifiers (2260 m^3 each) and subsequently to the tertiary treatment unit for disinfection by peracetic acid and filtration before final discharge in the Melma River (sensitive water body).

Waste activated sludge (WAS) generated by the biological treatment is recycled to the primary sedimentation unit and mixed with primary sludge. Daily, ~70-100 m^3/d of mixed sludge are pumped from the primary settler to a dynamic thickener with a flowrate of 20 m^3/h after polyelectrolyte dosage (0.8% of active compound solution). The final concentration of the thickened mixed sludge is around 5% total solids (TS). About 75% of the mixed thickened sludge is fed to an anaerobic digestion unit (1800 m^3 working volume). Digestate is dewatered by a centrifuge with the addition of polyelectrolyte (Hidrofloc CL 1908, Hidrodepur, Italy); the solid fraction is mechanically composted (mixing and aeration for 20 days) and used as agricultural fertilizer. The anaerobic supernatant is sent to the equalization tank (of 90 m^3) in the SCENA system for the biological N and P removal.

The remaining portion of mixed sludge (25%) is fed to a sequencing batch fermentation reactor (SBFR) with hydraulic retention time (HRT) equal to 5 days (working volume equal to 50 m^3). The SBFR is operated under mesophilic condition (37°C) for the fermentation of thickened sewage sludge and the on-site production of carbon source enriched of VFAs (mainly acetic and propionic acids). Daily, around 10 m^3 of fermentation sludge are extracted and replaced with fresh thickened sludge. The solid/liquid separation of the fermented sewage sludge is carried out by a screw-press (SCAE) (after polyelectrolyte (Hidrofloc CL 1908, Hidrodepur, Italy) dosage), generating ~2-4 m^3/h of fermentation liquid rich of volatile fatty acids (in total, ~ 11.5 m^3/d). The latter is collected in a storage tank of 20 m^3 and automatically dosed during the anaerobic and anoxic phases of a short-cut sequencing batch reactor (SBR) based on pH and conductivity sensors. The solid fermented fraction (13-15% total solids based) is mixed with the thickened mixed sludge and fed to the anaerobic digestor.

The anaerobic supernatant is treated in an SBR with a maximal working volume of 70 m^3 (3-4 cycles daily). The SBR is fed with ~7-9 m^3 of anaerobic supernatant in each cycle that

is treated via nitrite enhanced phosphorus removal associated with nitritation-denitritation (SCENA process). Information about the SBR configuration and data used in the analysis are shown in Figure 5.1. The typical SBR cycle consists of feeding (6-8 min), anaerobic conditions (30 min), aerobic conditions (200-240 min), anoxic (~60-140 min), settling (30 min) and discharge (8 min). The sensors integrated in the SBR include: pH, Dissolved Oxygen (DO), conductivity, Oxidation Reduction Potential (ORP), mixed liquor suspended Solids (MLSS) and temperature. Conductivity and pH are used to control the length of the aerobic and anoxic phases and the carbon source dosage. Additionally, variable frequency driver is used to control the air flow-rate of the blowers, maintaining the dissolved oxygen during aerobic phase in the range of 1.0 to 1.5 mg/L. The aeration system consists of volumetric blowers (nominal power 11 kW) and n 80 diffusers (INVENT), providing ~500 m^3/h of compressed air at 400 mbar of pressure (design oxygen transfer efficiency up to 15%). The treated supernatant is recirculated back to the WWTP headworks.

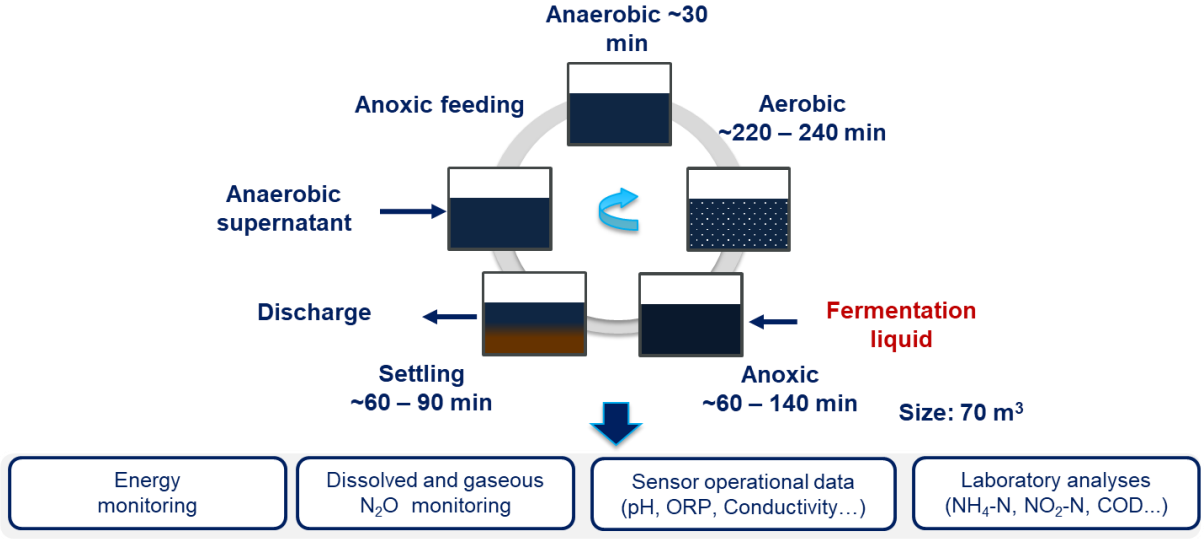


Figure 5.1: Schematic representation of a complete cycle in the SCENA process and datasets used in the analysis

5.2.1 *N_2O monitoring*

A monitoring campaign was conducted in the sidestream line at Carbonera WWTP treatment plant for approximately 4 months (January 2019 – April 2019). Dissolved N_2O concentrations were measured using a polarographic Clark-type electrode (Unisense, Aarhus, Denmark). The N_2O data were recorded and processed 24 h per day by the modular multichannel measuring device for liquid analysis (Unisense Environment, Denmark).

The calibration was performed bimonthly or when the average wastewater temperature changed by more than 3 degrees. After installing the N_2O wastewater sensor or replacing the sensor head, the sensor was stabilised overnight as recommended by Unisense for zero drift. For the calibration of the probes a stable bucket was filled with 4 L of secondary effluent. The temperature was measured, and the sensor was placed in the bucket until it stabilised. Two-point calibration mode for the N_2O sensor was applied. The zero-level calibration was done with secondary effluent gas-sparged with argon, in order to remove the dissolved nitrogenous gas. For the second point calibration 8 mL of N_2O calibration solution were used. The solution was injected to the secondary effluent.

To supplement the long-term monitoring campaign with Unisense probes, N_2O emissions in the headspace of the SBR reactor, were also continuously monitored with MIR9000CLD analyser (Environment Italia S.p.A.) during March – April 2019. The analyser measured N_2O through infrared spectroscopy (IRS). Weekly calibration using standard gas cylinders was performed. The gas flow was pumped, transported by a heating tube at 120 °C, filtered for dust removal and cooled at 4 °C. During aerobic phases, the outgoing gas flow-rate was assumed the same as the incoming gas flow-rate, neglecting the consumption and production of gasses in the reactor. The complete methodology is described in Spinelli et al., (2018)

5.2.2 *Complementary monitoring/analyses*

Liquid composite samples of the influent and effluent of the SCENA reactor were collected 2-4 days per week. The influent and effluent samples were analysed in terms of pH, COD,

ammonia-nitrogen (NH_4-N), soluble COD (sCOD) and nitrite-nitrogen (NO_2-N), nitrate-nitrogen (NO_3-N) and orthophosphate as phosphorus (PO_4-P) according to standard methods (APHA, 2005). Additionally, pH, conductivity, ORP, MLSS, temperature and energy consumption were monitored online in the SBR.

5.2.3 *N_2O emissions calculation*

- **Gas analyser**

The gases emitted during the aerated phases of the SBR reactor were calculated based on Equation (5.1):

$$Gas\ emitted_{(aerated)} = \left[\sum (c_{gas} \times Q_{aer} \times \Delta t\right] \quad (5.1)$$

Where:

- Gas emitted$_{(aerated)}$: GHG gas (N_2O) emitted during aerated operational times (N mg)
- c_{gas}: c_{gas} N_2O= N_2O (ppmv N_2O) × 1/0.08205 atm L mol^{-1} K^{-1} × (28/T(K)) × 10^{-6}
- Q_{aer}: gas flow coming out of the reactor during aerated zones (L m^{-1})
- Δt – time interval by which the off-gas concentration was recorded (m)

- **Conversion of N_2O and into CO2 equivalents**

CO_2 equivalents for the N_2O ($CO_{2eq_{(N_2O)}}$) emissions assessed for the system were calculated based on Equation (5.2) in respect to the NH_4-N removed:

$$CO_{2eq_{(N_2O)}} = \frac{N_2O_{emitted}\,(kg) \times 265}{NH_4 - N_{removed}(kg)} \quad (5.2)$$

CO_2 equivalents for the electricity consumption ($CO_{2eq_{(elec)}}$) emissions assessed for the system were calculated based on Equation (5.3) in respect to the NH_4-N removed. The emission conversion factor for the electricity grid mix in Italy from Ecoinvent 3 database has been used:

$$CO_{2eq_{(N_2O)}} = \frac{Electricity_{consumped}\,(kWh) \times 0.503}{NH_4 - N_{removed}(kg)} \quad (5.3)$$

5.3 Data analysis

5.3.1 *Methodological Framework*

Figure 5.2 summarises the methodological framework followed in this study. Phase one includes preliminary analysis of the collected data. Features extraction and density-based clustering was applied (Ester et al., 1996), to isolate abnormal cycles. The methodology and results of abnormal cycles' isolation are given in the supplementary material (section S4). In phase two, the behaviour of N_2O emissions and dissolved N_2O concentration during normal operation was investigated; efforts were focused to identify dependencies with the operational dataset and laboratory analyses. Finally, in phase three, classification and regression models were trained to predict the behaviour of aerobic dissolved N_2O concentration in the different cycles. Support vector machine classification (SVM) and regression (SVR) models were constructed (Cortes and Vapnik, 1995).

The first step for the prediction of the average aerobic dissolved N_2O concentration included the training of an SVM classifier (ANOXSVM) to predict whether dissolved N_2O will be consumed during the anoxic phase. This was significant, given that accumulated dissolved N_2O in the beginning of the aerobic phase, will be stripped during aeration. All cycles were divided in two classes: class anoxA (dissolved N_2O < 0.6 mg/L) and class anoxB (dissolved N_2O > 0.6 mg/L). The dissolved N_2O concentration threshold was set equal to 0.6 mg/L, since in ~88% of these cases, N_2O was consumed by the end of subsequent anaerobic phase. In cycles belonging to class anoxA, no N_2O carryover was assumed. It is important to note that the term anaerobic phase, is used to describe the first operational phase of the SBR (Figure 5.1) within each cycle and is not necessarily representative of the actual conditions in the reactor.

Subsequently, an SVM classifier (ANSVM) was trained to predict if dissolved N_2O will be consumed in the subsequent anaerobic phase. The threshold of N_2O at the end of the anaerobic phase was set equal to 2.6 mg/L (sensor calibration limit). Therefore, anaerobic phases with accumulated N_2O were classified in two groups: class anaerA (N_2O concentration < 2.6 mg/L) and anaerB (N_2O concentration > 2.6 mg/L) . Cycles belonging to anaerA class, were used to train an SVR model (ANSVR) to predict the dissolved N_2O concentration at the end of the anaerobic phase.

Finally, an SVR model was trained to predict the average N_2O concentration during the aerobic phase (AERSVR), utilizing the ANSVR model predictions for cycles with initial aerobic N_2O less than 2.6 mg/L. Finally, the aerobic SVR model was also tested to cycles belonging in class anaerB (N_2O concentration > 2.6 mg/L). In anaerB cycles, initial aerobic N_2O accumulation exceeds the calibration limit of the sensor. Additionally, aerobic N_2O accumulation starts before completion of the stripping of pre-existing dissolved N_2O. In these cases, the average dissolved N_2O concentration of the cycle, was calculated considering the period from the first minimum of dissolved N_2O concentration until the end of aeration (or after 30 min if a local minimum did not exist). Additionally, initial N_2O accumulation was assumed to be equal to 0.6 mg/L (average minimum after initial N_2O stripping observed in these cycles).

In practice, the methodology followed was not linear as it is illustrated in Figure 5.2; it involves several backward and forward loops between the different steps. The feedback loops were necessary to leverage the knowledge discovered and adjust the data-preparation (i.e. new features extraction, different pre-processing) and mining phases.

Details of the specific methods applied are provided in the relevant sections. In practice, the methodology followed was not linear as it is illustrated in Figure 5.2; it involved several backward and forward loops between the different steps. The feedback loops were necessary to leverage the knowledge discovered and adjust the data-preparation (i.e. new features extraction, different pre-processing) and mining phases.

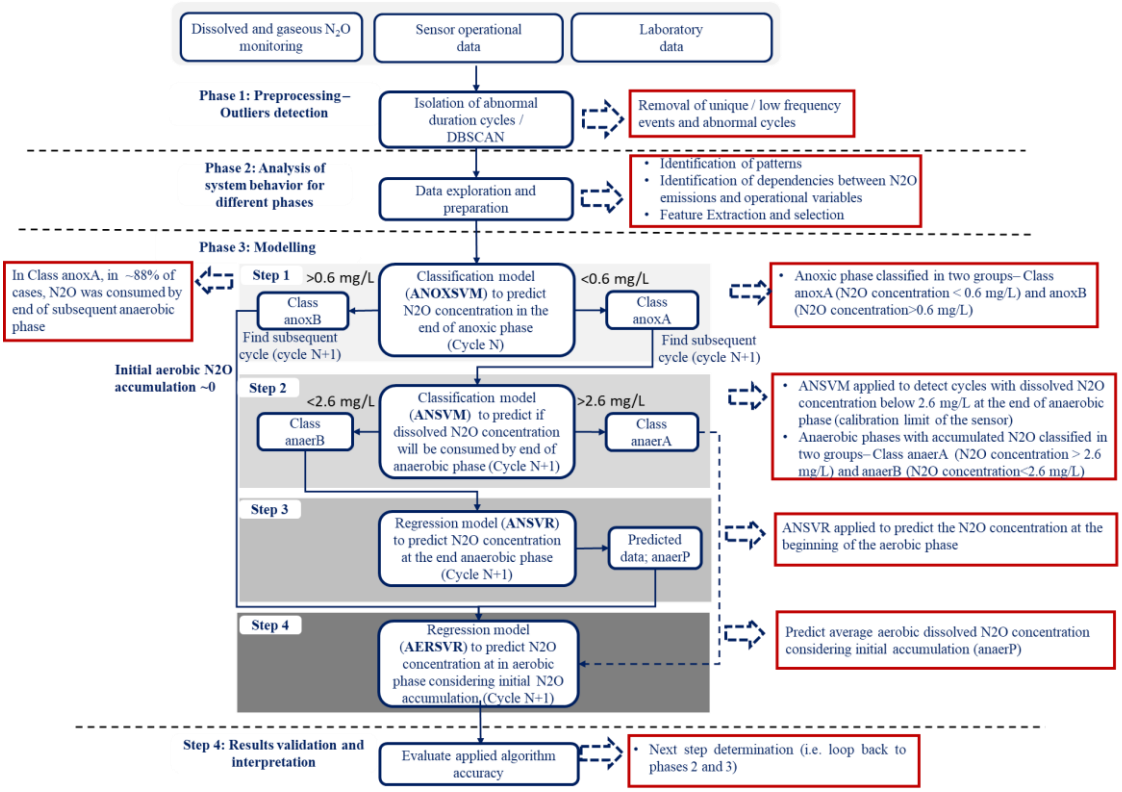


Figure 5.2: Methodological Framework followed in the study

5.3.2 *Abnormal cycles detection*

Sensor data collected from the SCENA process consisted of a three-dimensional data matrix, X ($I \times J_x \times K$), where I is total number of cycles, J_x represent the total process variables and K is the number of sampling time intervals in a cycle (equal to 1 min). Variable-wise unfolding was applied keeping the variables' dimension and merging the other two dimensions (Wold et al., 1998). Subsequently the data were z-normalised (zero mean and unit variance) preserving the variable trajectory information in the data. Each phase (i.e. anaerobic vs aerobic vs anoxic) is characterised by its own underlying dynamics of the variables monitored (i.e. ORP, pH, conductivity) and the system can exhibit different performance between different phases. Therefore, appropriate features were extracted from the variables monitored for each phase and cycle, capturing information on the performance of the cycle/phase. Feature extraction, compressed data dimension to a finite number of descriptive features able to describe whether the variables

monitored, exhibited the expected dynamics and range during each phase of the cycles analysed, with respect to pre-determined patterns and normal ranges. The features selected for each phase are shown in Table 5.1. The selection of features was based on the profiles of the variables monitored in cycles with abnormal effluent PO_4-P, NO_2-N and NH_4-N concentrations based on the available laboratory measurements.

Table 5.1: Features used for outliers detection

Anaerobic	Aerobic	Anoxic
Minimum pH	Maximum pH	Minimum conductivity rate
ORP change	Conductivity kurtosis	Maximum pH
Conductivity change	Minimum rate conductivity	Mean conductivity
Time ORP min/Total phase time	Mean ORP	Mean pH
	Time minimum rate ORP/Total phase time	pH rate at end of phase
		ORP kurtosis

Subsequently density-based spatial clustering of applications with noise (DBSCAN), (Ester et al., 1996) was applied to the phase-based datasets as described in section 4.3.1. The aim was to detect cycles with behaviour deviating from the expected observed under normal operational conditions. Laboratory analyses were used to provide insights on possible data patterns indicating poor performance.

5.3.3 *Support Vector Machines classification and Support vector regression*

Support vector machines (SVMs) are a range of supervised non-parametric classification and regression algorithms that have various applications in several fields including remote-sensing (Mountrakis et al., 2011), hydrology (Raghavendra and Deka, 2014), bioinformatics (Byvatov and Schneider, 2003) and wastewater (Corominas et al., 2018). For instance, in wastewater, support vector regression (SVR) has been successfully applied to data generated from mechanistic modelling of biological processes (Fang et al., 2011; Xie et al., 2011) or to experimental data (Seshan et al., 2014) to predict reactors' performance.

As shown in Figure 5.2, SVM classification was applied in the analysis to predict the behavior of dissolved concentration in different phases of the SBR operation. SVM aims to define an optimum separating hyperplane in the feature space that maximises the margin between two different classes; a complete description of the method is provided in section 4.3.6. Binary classification was performed in this study. The algorithm was implemented with the kernlab package (Karatzoglou et al., 2004) in R software. Repeated 10-fold cross validation (3 repetitions) was applied to select the cost and γ parameters over a grid-search in the caret package (Kuhn, 2008) in R software. The dataset was randomly divided into test and train, with 70% of the available data used for training the SVM model and 30% used for testing. Over-sampling was applied for the minority classes within the 10-fold cross validation loop (before training)

SVMs have been also used to regression tasks with continuous output variables (Mountrakis et al., 2011). In the SVR case, the aim of the method is to identify the hyperplane that has the minimum distance to all data points.

Considering a train dataset (x_i, y_i), x_i is a n-dimensional M set of vectors, with $x_i \in \mathbb{R}^n$ where $i = 1, \ldots, M$ and y_i the target property of each x_i case, with $y_i \in \{-1,1\}$ During the training phase, the aim of the algorithm is to define the function $f(x) = (\mathrm{w}x + w_0)$, with $w, x \in \mathbb{R}^n$ and $w_0 \in \mathbb{R}$ that deviates from the target variable y_i by ε at maximum, which can be expressed as $[y_i - f(x_i)] < \varepsilon$, with ε being the insensitive loss function. Therefore, the following quadratic programming problem needs to be solved (Singh et al., 2011):

$$Minimize\ \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{M}(\xi_i + \xi_i^*) \tag{5.4}$$

$$\text{Subject to } y_i - w^T x_i - w_0 \geq \varepsilon + \xi_i$$

$$\text{and } w.x + w_0 - y_i \leq \varepsilon + \xi_i^*, i = 1,2,\ldots,N \text{ and } \xi_i \xi_i^* \geq 0$$

The optimum regression function after the introduction of kernel function and transformation of Equation (5.4) in a similar way as shown in section 4.3.6, can be expressed as:

$$\mathrm{f(x)} = \left(\sum_{i=1}^{M_{SV}} (\alpha_i - a_i^*)\mathrm{K}(x, x_i) + w_0 \right) \tag{5.5}$$

Where: α_i a_i^* the Lagrange multipliers with $\mathrm{C} \geq \alpha_i, a_i^* \geq 0, i = 1,2,\ldots,M$ and $\mathrm{K}(x, x_i)$ the kernel function. In this case, data points with non-zero α_i a_i^* represent the support vectors.

Local models were developed based on observations from each phase of the SBR reactor instead of the dataset from the duration of the whole cycle. The underlying characteristics and dependencies of the operational variables vary between anoxic, aerobic and anaerobic conditions. Additionally, the performance of the system under the different phases within the cycle can also vary. There are significant benefits in the development of local phase-based models. The behaviour of dissolved N_2O and triggering operational conditions vary between the different phases; local models enable to investigate the phase-based dependency structures that would not be possible using the whole cycle dataset.

5.3.4 *Model performance evaluation*

The performance of the classification SVM models were evaluated based on the accuracy, and kappa as described in section 4.3.1 and from the sensitivity and specificity as expressed below.

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{5.6}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{5.7}$$

Where: TP is the number of true positives, FP false positives, FN false negatives, and TN true negatives.

Similarly, the regression models were evaluated considering the root mean squared error (RMSE) and R-squared (R2) that are expressed as:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{M}(y_{pred,i} - y_i)^2}{N}} \tag{5.8}$$

$$\text{R} = \frac{\sum_{i=1}^{M} y_{pred,i} y_i - \frac{1}{M}\sum_{i=1}^{M} y_i \sum_{i=1}^{M} y_{pred,i}}{\sqrt{\left[\sum_{i=1}^{M} y^2{}_i - \frac{1}{M}(\sum_{i-1}^{M} y_i)^2\right]}\sqrt{\left[\sum_{i=1}^{M} y^2{}_{pred,i} - \frac{1}{M}\left(\sum_{i-1}^{M} y_{pred,i}\right)^2\right]}} \tag{5.9}$$

Where: y_i the measured target variable, $y_{pred,i}$ the predicted value, M the number of observations and $R - squared = R^2$.

5.4 Results and discussion

5.4.1 *SCENA performance*

The SBR treats up to 43 kg of N/day of anaerobic supernatant, which results in a volumetric nitrogen loading rate up to 0.62 kgN/m^3 day. The performance of the SBR reactor in terms of NH_4-N removal, was stable during the monitoring campaign. Deviations in the profile of the N_2O emissions, electricity consumption, operational variables and pollutant removal efficiencies, were mainly due to i) limitation of anaerobic supernatant during weekends and extension of aerobic phase length in the SBR and ii) low availability of carbon source. During system's normal operation (January 2019 - April 2019), the average removal efficiency of NH_4-N, TN and PO_4-P was 78%, ~77% and 78% respectively. Influent and effluent concentrations of the SCENA system for the duration of the monitoring campaign are provided in

Table 5.2.

Table 5.2: Influent and effluent concentrations of the SCENA system

	Parameter	unit	mean	Sd
SBR Influent	NH_4-N	mg/L	992.5	90
	PO_4-P	mg/L	30.8	6.9
	pH		8.2	0.2
	sCOD	mg/L	1111.7	562
	Flow-rate	m^3/d	30 (8.4 per cycle)	2.2
	Air flow-rate	m^3/h	450 (170 - 520)	78
	Dimensions	mxmxm	8 x 3.5 x 2.5	
SBR Effluent	NH_4-N	mg/L	214.7	80.93
	NO_2-N	mg/L	3.23	9.7
	NO_3-N	mg/L	0.28	0.34
	PO_4-P	mg/L	6.78	2.22
	pH		8.04	0.3
SBR Reactor	MLSS	g/L	5.05	0.87
	HRT	d	2.39	
	SRT	d	13-15	
	pH		7.7	0.5
	T	°C	30.02	1.56
Fermentation Liquid	NH_4-N	mg/L	715	72.6
	PO_4-P	mg/L	86	12
	pH		5.6	0.6
	T	°C	36	5.1
	sCOD	mg/L	13082	2228
	ferm_Hac	mg/L	3250	546
	ferm_HPr	mg/L	2281	588
	ferm_Hbut	mg/L	1347	196
	Flow-rate to SBR	m^3/cycle	7.45 (~2.41 per cycle)	3.0

5.4.2 *N_2O Emission factor*

Over the period monitored with the gas analyser (March – April 2019), on average ~0.8 kg of N_2O-N was emitted in each cycle, equivalent to 7.6% of the NH_4-N load in the SBR. In terms of the NH_4-N oxidized the N_2O EF was equal to 11% (±4). Given that the N2O concentration during the anoxic phases exceeded the calibration limit of the sensor, only the emissions during the aerobic phase were considered. N_2O emissions exhibited significant variability ranging from 0.14 kg N_2O-N/cycle (1.3% of NH_4-N load) to ~2 kg

N_2O-N/cycle (19% of NH_4-N load) as shown in Figure 5.3 (a). Emission peaks higher than 1.5 kg N_2O-N/cycle and the increasing trend observed close to the end of the monitoring campaign coincide with peaks in the conductivity change in the aerobic phase of the cycles (Figure 5.3 (b)). Laboratory analyses performed approximately four times per week, did not demonstrate any significant changes in the influent COD, NH_4-N loads and removal efficiencies linked with the increasing trend of the emissions observed in Figure 5.3 (a). Given the wide range of the N_2O emissions observed in the system, in the following sections, efforts were focused to identify triggering operational conditions.

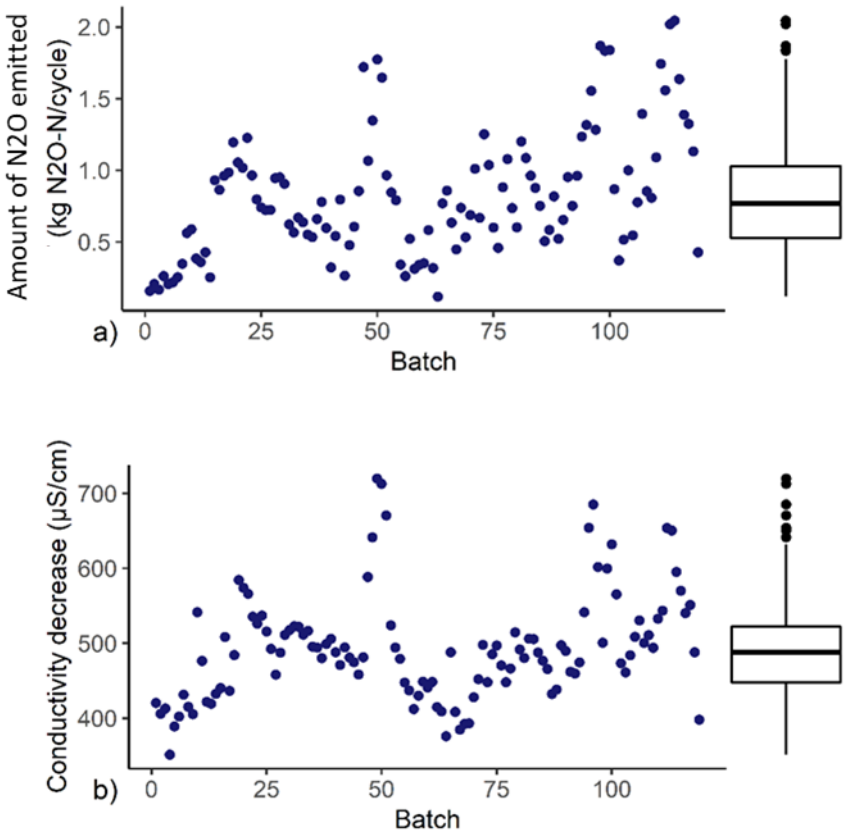


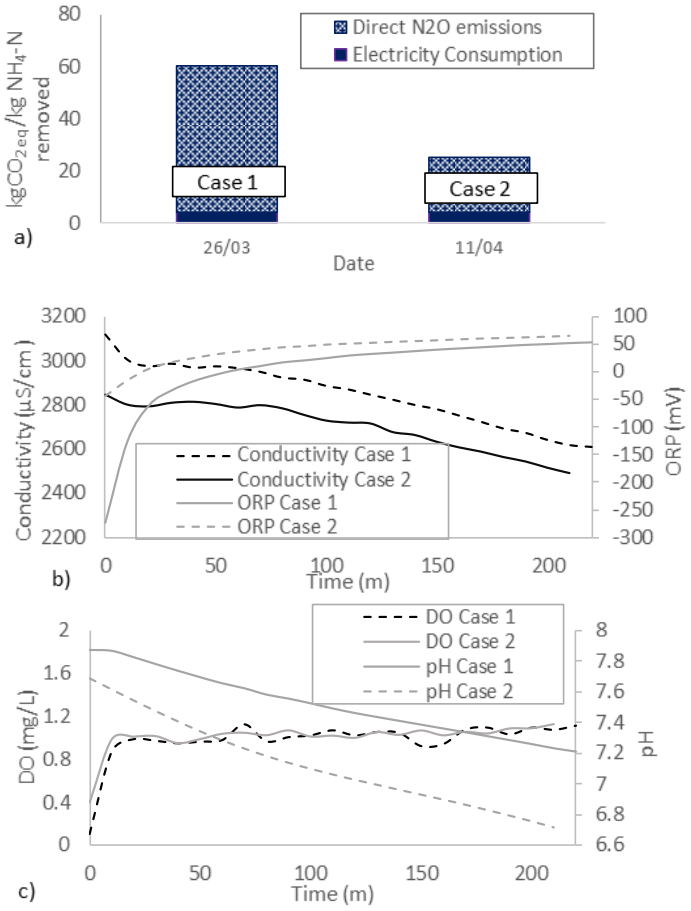
Figure 5.3: (a) N2O emissions and (b) aerobic phase conductivity decrease, during monitoring campaign (gas analyser, March-April)

5.4.3 *Energy consumption vs N_2O emissions*

The operational carbon footprint of the sidestream line was estimated considering the direct GHG emissions (from N_2O) and the electricity consumption. The electricity consumption was relatively steady over the monitoring period; on average ~5.4 kWh was consumed in the SBR for the removal of 1 kg of NH_4-N from the anaerobic supernatant. The average energy consumption of the SBR was equal to 49 (± 3.0) kWh per cycle and represented ~77% of the total electricity consumption of the SCENA system. On average ~48.7 kg of CO_{2eq} are generated for the removal of 1 kg of NH_4-N due to the direct N_2O emissions and electricity consumption in the system. The contribution of the total N_2O emissions to the operational carbon footprint of the SCENA process ranged from 66.7% to 96.8% when all the equipment electricity consumption (i.e. fermenter, dynamic thickener) were considered (average equal to ~88%). Given the variability of the N_2O emissions observed in the system (Figure 5.3) the kg of CO_{2eq} emitted per kg of NH_4-N removed ranged between 9.5 kg CO_{2eq} to 117.7 kg CO_{2eq}.

Figure 5.4 (a), shows the average operational carbon footprint (considering direct N_2O emissions and electricity consumption) of the SCENA system for two cases with different ranges of N_2O emissions. In the first case (26/03), a considerable amount of N_2O was emitted, equal to ~10.5% of the influent NH_4-N load. In the second case the emissions were significantly lower, equal to ~4% of the influent NH_4-N load. Both cases are characterised by similar influent NH_4-N concentrations, phase duration, temperature and ammonia removal efficiencies (~79%). The DO concentration is equal to ~1 mg/L. In case 1, the operational carbon footprint of the process is ~136% higher compared to case 2. This example shows that under similar conditions (considering laboratory analyses, average pH and DO), dissolved N_2O concentrations can vary significantly in the studied system. Investigation of the behaviour of conductivity during the two aerobic phases, showed higher conductivity and pH decrease in case one (~ 510 μS/cm and ~1 respectively) compared to case two (~350 μS/cm and 0.7 respectively) (Figure 5.4 (b) and (c)).

Additionally, the initial aerobic ORP in case 2, was higher (-43 mV) compared to case 1 (-274 mV) (Figure 5.4 (b)).



Operating Conditions	Unit	26/03	11/04
Influent NH_4-N	mg/L	1186.3	1154.13
DO	mg/L	1.0	1.0
Anaerobic/Aerobic/ Anoxic duration	h	0.5/3.6/1.5	0.5/3.7/1.5
Effluent NH_4-N	mg/L	211.1	185.9
Effluent NO_2-N	mg/L	~0	1.6
Temperature	℃	31.7	32.1

Figure 5.4: (a) Example of the effect of N_2O emissions in the operational carbon footprint for two cases, (b) aerobic profiles of conductivity, ORP and (c) DO and pH for the two cases shown in (a)

Therefore, efforts to understand the N_2O triggering operational conditions and mitigate GHG emissions, should consider the dynamic in-cycle behaviour of the variables monitored in the system. The relationship between the operational variables (i.e. DO, NH_4-N concentration, ORP, conductivity) will be discussed in the following sections.

5.4.4 *Variability of N_2O emissions during normal operation*

N_2O was emitted during aeration phase in all cycles and correlated significantly with the dissolved N_2O accumulation. One representative cycle profile for the dissolved N_2O concentration and N_2O emissions in cycles starting without dissolved N_2O accumulation from the previous cycle is shown in Figure 5.5, together with the DO, NH_4-N, conductivity, ORP and pH.

ORP at the beginning of the aerobic phase shows a correlation with the DO, whereas N_2O accumulation is minimum. Dissolved N_2O increases in the first 60-70 min of aeration (a small change in the pH slope can be seen coinciding with the peak of accumulated N_2O) indicating that the generated N_2O is higher than the stripped N_2O. N_2O accumulation shows a decreasing trend after ~90 minutes of aeration. Subsequently dissolved N_2O concentration increases when aeration stops, and the anoxic phase starts. This shows that production of N_2O continues under decreasing DO and until DO depletion. The calibration range of the dissolved N_2O probe is between 0 - 2.6 mg/L. Therefore, the accumulation of dissolved N_2O can be higher than the peak shown in Figure 5.5. During the anoxic phase, pH increases rapidly during the dosage of fermentation liquid, followed by a slow decrease upon the end of carbon dosage phase. A sudden change in the ORP signal slope ("nitrite knee") indicates the depletion of nitrite whereas TN still exists in the form of N_2O. Accumulated N_2O is subsequently depleted rapidly after NO_2^-N depletion.

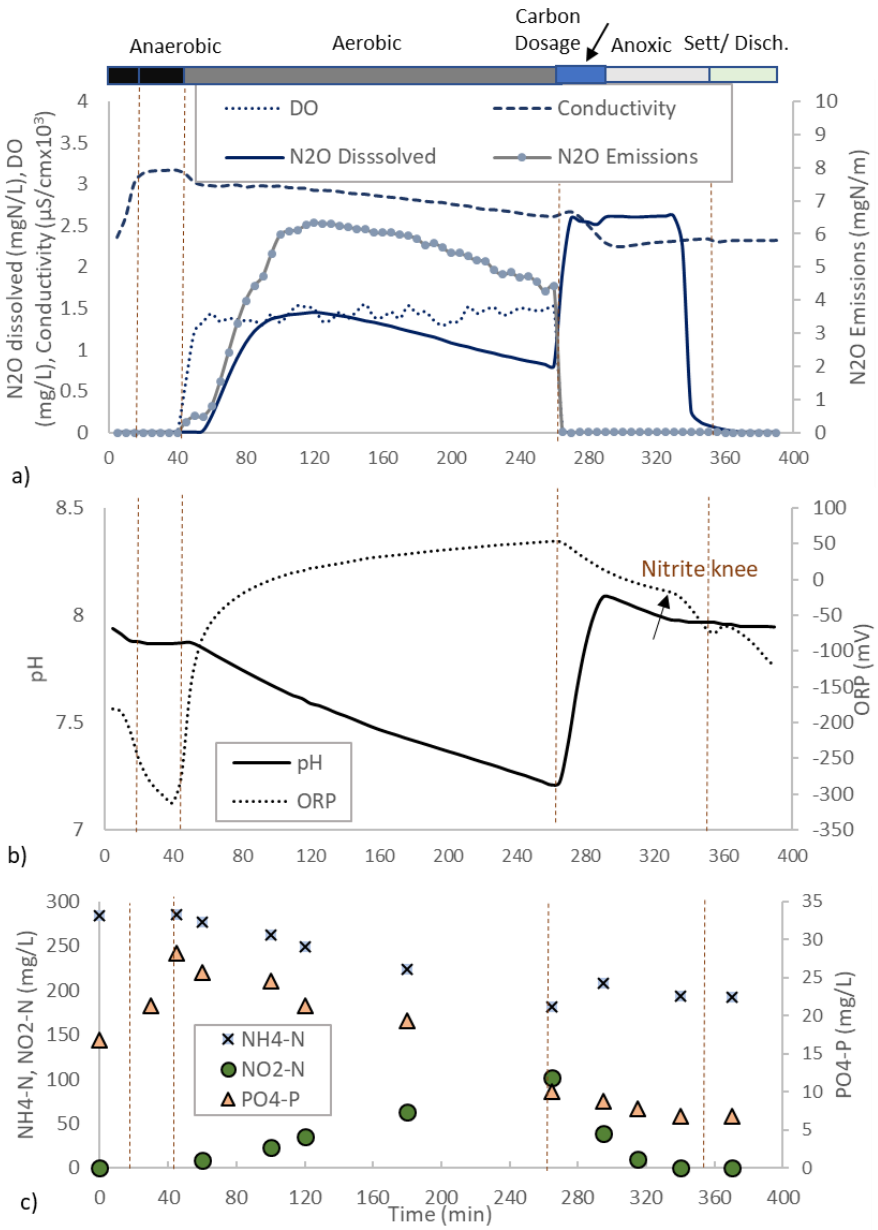


Figure 5.5: Representative cycle profile for the (a) dissolved N_2O concentration, N_2O emissions, conductivity, DO, (b) ORP and pH, and (c) NH_4-N, NO_2-N and PO_4-P concentrations

5.4.5 *Outliers Analysis*

In total 96 cycles were characterised by abnormal duration of the anaerobic, aerobic or anoxic phases mainly attributed to limitations in the anaerobic supernatant or fermentation

liquid. Phase duration abnormalities include: i) absence of one or more phases (total: 21 cycles), ii) duration of anaerobic phase less than 15 min and anoxic phase less than 45 min (total: 33 cycles), iii) duration of aerobic phase more than 360 min (total 42 cycles). The normal operation ranged mainly between 30-45 min, 60-150 min and 200-240 for the anaerobic, anoxic and aerobic phases

The laboratory analyses were performed 2-4 times per week; therefore, the isolation of cycles with poor performance was not feasible based only on laboratory measurements. Features representing the behaviour of variables monitored in the system for each phase were extracted and DBSCAN was applied to detect cycles deviating from normal operational conditions. In total, 66 cycles with abnormal behaviour in one or more phases of the cycle were detected (representing ~20% of the cycles analysed). Figure 5.6 - Figure 5.8, show the most frequent profiles of ORP, pH and conductivity identified as abnormal by DBSCAN, in each phase of the cycle and the representative behaviour of the variables during normal operation. NH_4-N, NO_2-N or PO_4-N effluent concentrations of the system are also reported for the examples presented.

The behaviour of the operational variables in the anaerobic phase events detected by DBSCAN can be categorized in two groups; representative examples of these groups are shown in Figure 5.6 (a) and (b). Figure 5.6 (c) shows typical profiles of the variables monitored during normal SBR operation. In total, 16 cycles were identified with the behaviour of anaerobic phase shown in Figure 5.6 (a and b). Specifically, 7 cycles were detected with behaviour of the operational variables similar to Figure 5.6 (a) and 9 cycles with behaviour of the operational variables similar to Figure 5.6 (b).

The example shown in Figure 5.6 (a) shows an anaerobic phase where the effluent PO_4-P concentration was equal to 19.72 mg/L (influent: 18.06 mg/L, effluent concentration is higher than the influent due to the PO_4-P content of the fermentation liquid that was dosed during the anoxic phase). After 15 min of anaerobic conditions the ORP stabilises and then slightly increases. Similarly, the ORP behaviour in Figure 5.6 (b), deviates significantly from behaviour during normal operation; the ORP increases during the anaerobic phase.

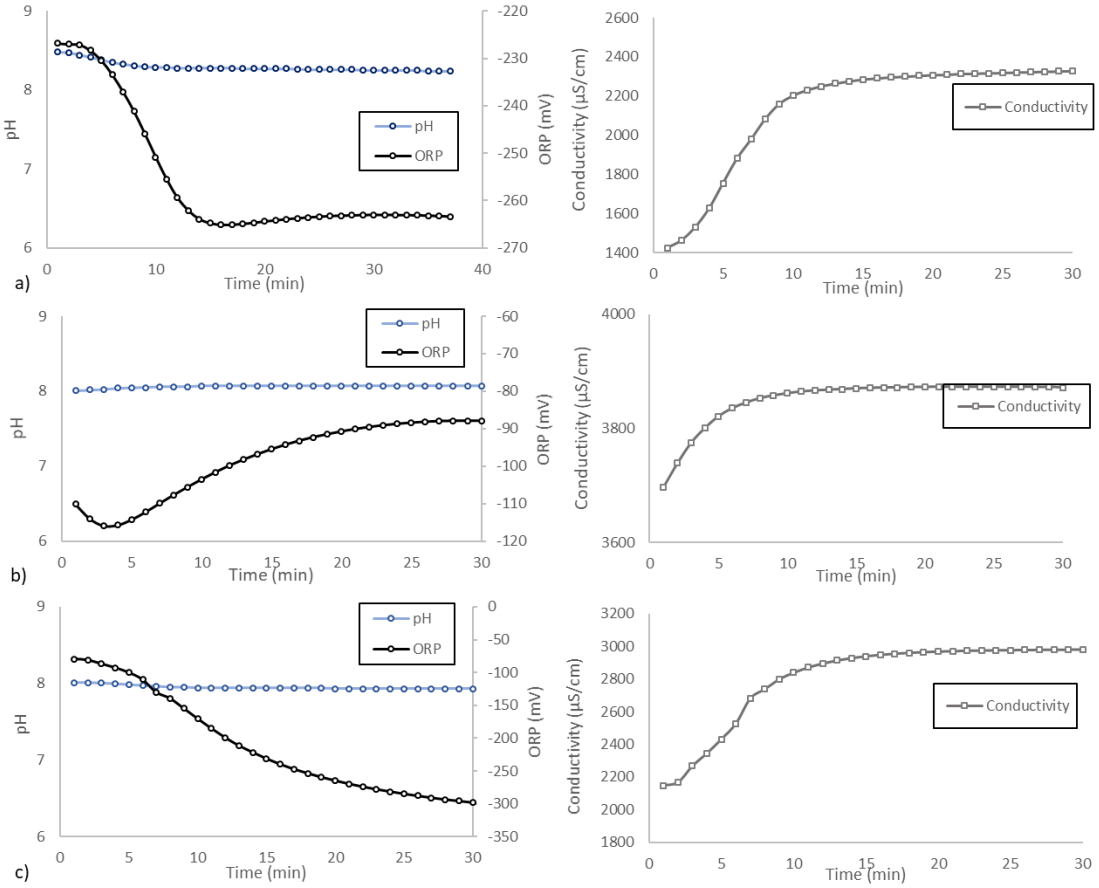


Figure 5.6: Profiles of ORP, pH and conductivity during anaerobic phase in (a-b) cycles with poor performance detected by DBSCAN, (c) normal operation

Examples of ORP, pH and conductivity profiles in abnormal anoxic phases are shown in Figure 5.7 (a and b) and normal profiles are shown in Figure 5.7 (c) . In total, 29 cycles detected by DBSCAN were characterised by the profiles shown in Figure 5.7 (a). Laboratory analyses were available for ~30% of these cycles and showed an average effluent NO_2-N concentration at the end of the cycle equal to 176.2 mg/L indicating incomplete denitritation. pH increases during the anoxic phase whereas conductivity increases for ~30 min (fermentation liquid feeding) and then decreases. During normal operation, pH increases rapidly during fermentation liquid feeding and stabilises after the nitrite "knee" (Figure 5.7 c). Figure 5.7 (b) is typical for cycles without fermentation liquid dosage in the anoxic phase. In total, 5 cycles showed this profile and were isolated.

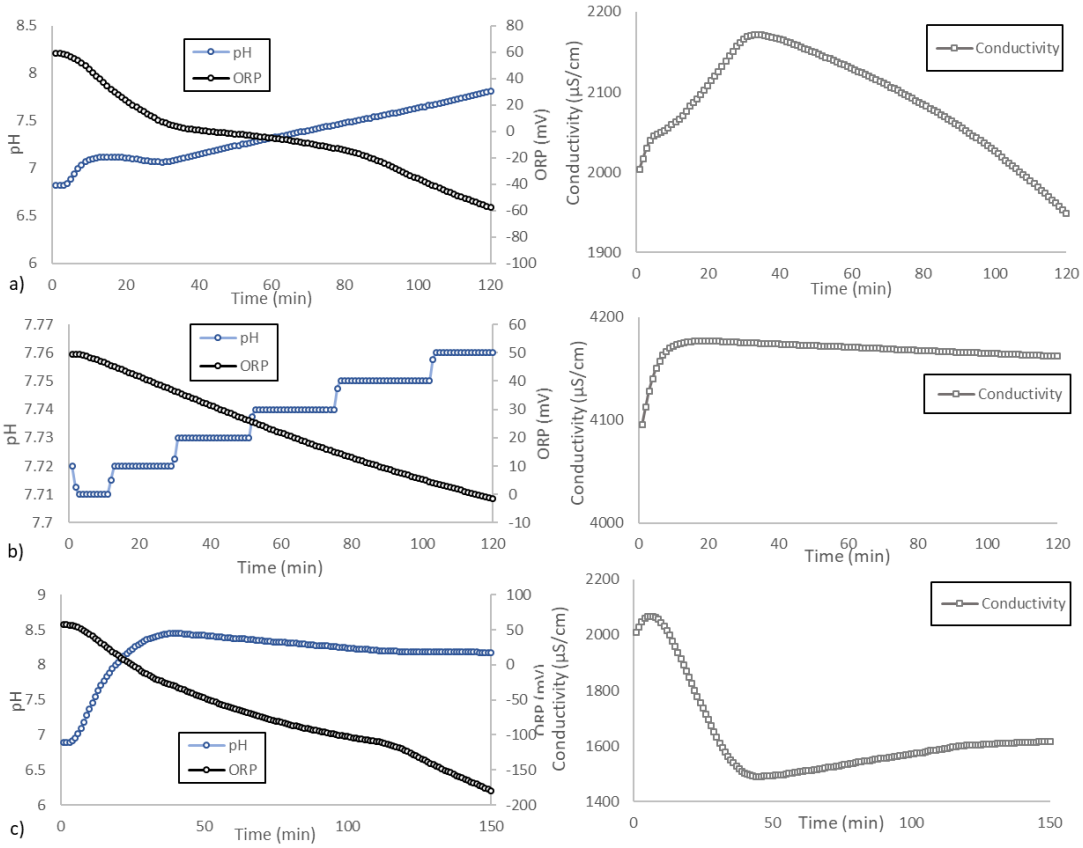


Figure 5.7: Profiles of ORP, pH and conductivity during anoxic phase in (a-b) cycles with poor performance detected by DBSCAN, (c) normal operation

Typical behaviour of the ORP, conductivity and pH during normal aerobic operation is shown in Figure 5.8 (c); the ORP increases rapidly during the first minutes of aeration and then at a reduced rate, pH decreases at almost a constant rate and conductivity decreases. The majority of the cycles isolated by DBSCAN can be categorized in two groups; Figure 5.8 (a) and (b) show the typical aerobic behaviour of operational variables for these two groups. The profiles shown in Figure 5.8 (a) are mainly characterised by very slow decrease rate of the conductivity after the first 30 min of the aerobic phase and relatively steady pH. Lower than average (<60%) NH_4-N removal efficiencies were linked with the conductivity behaviour shown in Figure 5.9 (a). The average dissolved N_2O concentration is equal to 0.85 mg/L for aerobic phases belonging to this group. Conductivity behaviour similar to Figure 5.8 (a) was detected for 25 cycles.

Similarly, the OPR profile in Figure 5.8 (b) (8 cycles total), deviates significantly from the ORP profile during normal aerobic operation (Figure 5.8 c). However, there were not available laboratory measurements to link the observed behaviour with operational efficiency. All aerobic cycles belonging to this group, are characterised by elevated dissolved N_2O concentration (>2.5 mg/L) at the beginning of the aerobic phase whereas the average aerobic dissolved N_2O concentration is ~1.45 mg/L.

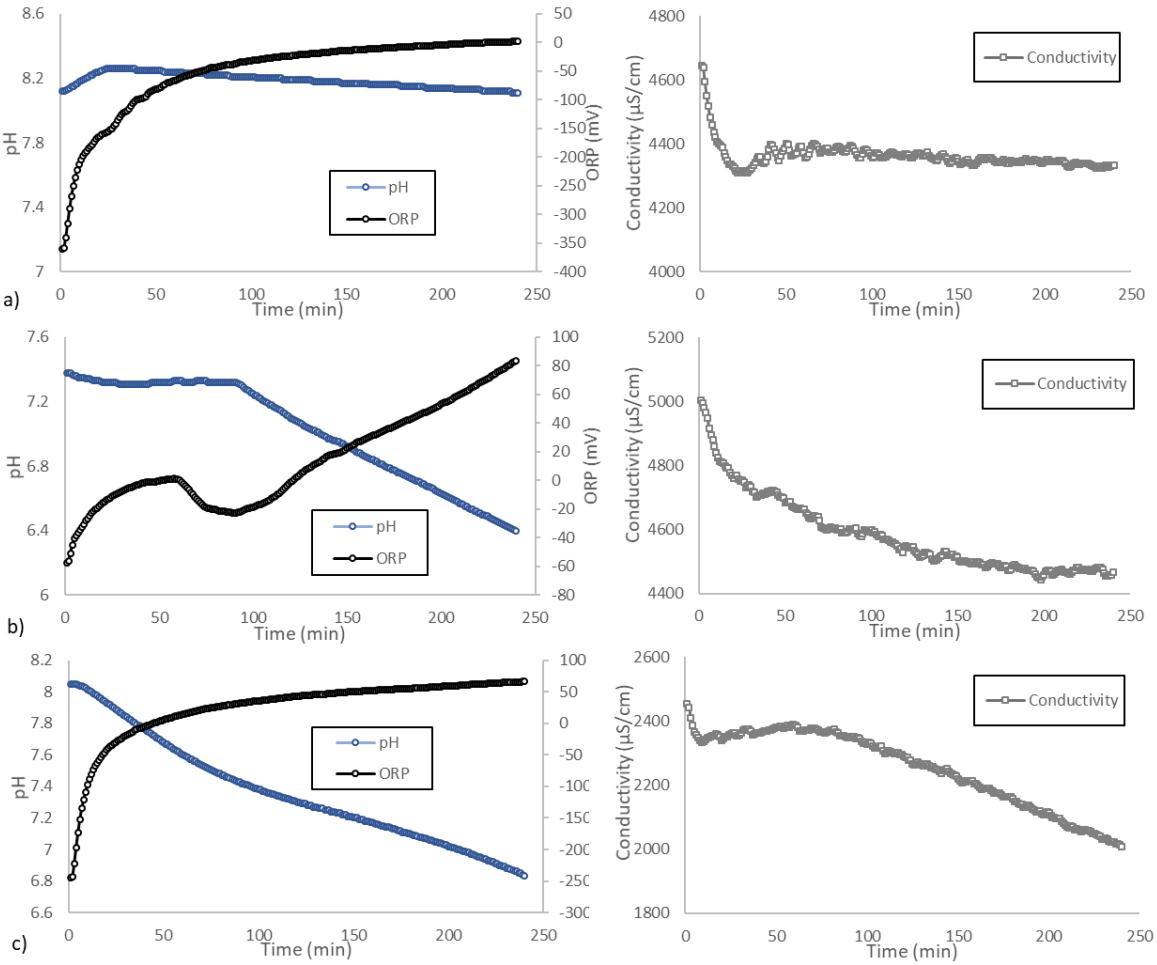


Figure 5.8: Profiles of ORP, pH and conductivity during aerobic phase in (a-b) cycles with poor performance detected by DBSCAN, (c) normal operation

The analysis showed that feature extraction combined with density-based clustering can be applied to the historical sensor data and detect cycles with abnormal SBR operation. In this study ~80% of the abnormal events detected by DBSCAN were linked with outliers in the

NH_4-N, NO_2-N and PO_4-P based on the laboratory measurements. Moreover, more than 95% of the events were characterised by significant deviations in the range and behaviour of the variables monitored in the system compared to normal reactor operation.

5.4.6 *The pattern of N_2O emissions*

Offline data from laboratory studies and the ranges of the operational variables were analysed in order to investigate significant changes that contribute to high accumulation of dissolved N_2O concentration and high N_2O emissions.

Figure 5.9 (a) shows the daily average dissolved N_2O concentration (coloured points) during aerobic phase versus conductivity at the end of aerobic phase and the effluent NH_4-N concentration. Conductivity is significantly related and can be linked with the NH_4-N concentration in the reactor (spearman correlation coefficient equal to 0.97). High average aerobic dissolved N_2O concentration (>1.5 mg/L) was mainly observed with NH_4-N concentrations lower than 150 mg/L and higher than 300 mg/L in the effluent of the SBR. Additionally, the spearman correlation coefficient between dissolved N_2O and average aerobic conductivity decrease rate (μS/cm/min) was equal to -0.7 and N_2O concentration peaks were observed for conductivity decrease rate > 1.8 μS/cm/min. The latter indicates that higher emissions occur under high ammonia removal efficiency that can be linked with higher ammonia oxidation rates (AOR) (i.e. due to pH values observed ~ 8) triggering the NH_2OH oxidation pathway or higher than average NO_2^-N accumulation (triggering nitrifier denitrification pathway). Domingo-Félez et al., (2014) found that N_2O production rates were positively correlated with the extant nitrification rate in a single-stage nitritation/Anammox reactor. Similarly, Law et al. (2011) identified a linear relationship between AOR and N_2O emissions in a partial nitritation SBR reactor treating the reject water from anaerobic digestion. Law et al. (2011) suggested that is attributed to higher accumulation of the ammonium oxidation intermediates (hydroxylamine (NH_2OH) and nitrosyl radical (NOH)) leading to faster N_2O formation or to the increased use of electrons reducing nitrite to nitric oxide (nitrifier denitrification pathway) under low DO concentrations. High nitrite accumulation has been also linked with elevated N_2O

emissions and the nitrifier denitrification pathway, especially under low DO concentrations (Tallec et al., 2006; Kampschreur et al., 2008b; Desloover et al., 2011; Peng et al., 2015; Massara et al., 2017; Law et al., 2012). For instance, Peng et al. (2017) and Kampschreur et al. (2009), in a nitritation-denitritation SBR and a full-scale single stage nitritation-Anammox reactor respectively, identified linear relationship between nitrite accumulation and N_2O emissions at DO levels below 1.5 mg/L. Similarly, Tallec et al (2006) in a nitrifying activated sludge observed eightfold increase of N_2O emissions with the addition of nitrite pulses (10 mg/L) at DO equal to 1 mg/L. Therefore, both hydroxylamine oxidation and the nitrifier denitrification are possible during aeration in the investigated SBR.

The average dissolved N_2O concentration during the aerobic phase of different cycles varied significantly in relation to the average DO concentration. Figure 5.9 (b), shows that the dissolved N_2O concentration peaks coincided with average DO concentrations less than 0.9 to 1 mg/L. The spearman correlation coefficient between dissolved N_2O and DO concentrations was equal to -0.7. The coloured points in the Figure, represent the ORP at the end of the aerobic phase; ORP is higher than 40 mV in the majority of the cycles with average aerobic dissolved N_2O concentration less than 1 mg/L. Only cycles without dissolved N_2O accumulation from the previous anoxic phase are shown in the graph. Stenström et al. (2014) showed decreasing DO concentrations lower than 1–1.5 mg/L are linked with higher nitrite accumulation and are positively correlated with N_2O emissions during nitrification in a full-scale predenitrification-nitrification SBR treating anaerobic supernatant. Similarly, Pijuan et al., (2014) reported an increase of N_2O emissions in a nitritation reactor with the reduction of DO from 4 to <1 mg/L. During the monitoring period, blowers operated at maximum flow-rate. Therefore, the presence of residual biodegradable COD concentration in the aerobic, is expected to decrease DO concentration. Similarly, higher influent NH_4^+ loads or higher ammonia oxidation rates (that can also result in increased NO_2^- accumulation) can impact the DO concentration in the system. The dissolved N_2O concentration can be affected by a combination of variables;

therefore, it cannot be deduced that the decreased DO is the sole contributing factor triggering the increased N_2O generation observed.

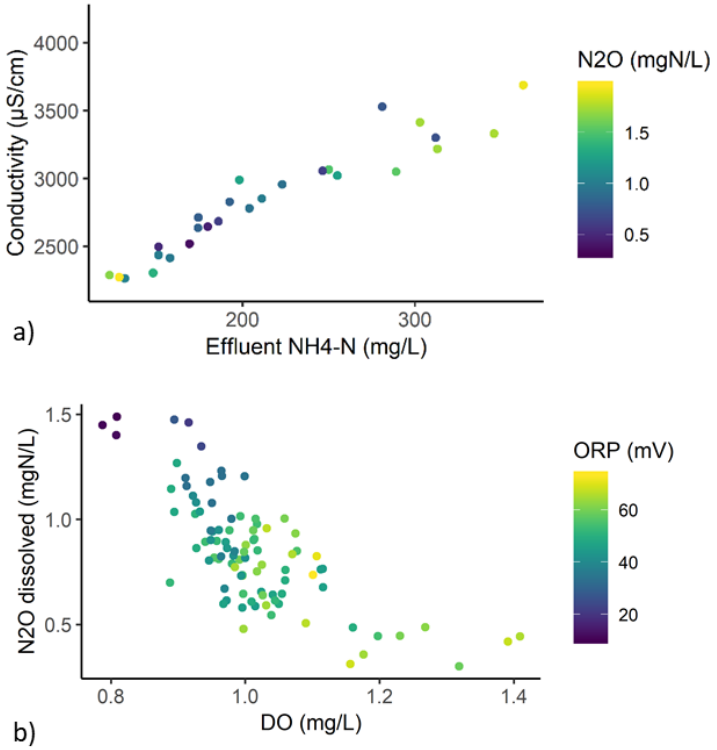


Figure 5.9: (a) Daily average conductivity at the end of the aerobic phase versis effluent NH_4-N concentration (coloured points: average dissolved N_2O accumulated in the aerobic phase), (b) Aerobic average accumulated dissolved N_2O in respect to DO concentration; only cycles without initial N_2O accumulation from the previous anoxic cycle are shown (coloured points: ORP at the end of the aerobic phase)

5.4.7 *Impact of accumulated N_2O in the end of anoxic and anaerobic phase*

Several parameters have been reported to affect the N_2O accumulation under anoxic conditions, such as the inhibition of the nitrous oxide reductase (Nos) by free nitrous acid (FNA) or high accumulation of NO_2^-, the electron competition between electron acceptors and the type of carbon source (Itokawa et al., 2001; Pan et al., 2013; Zhou et al., 2008; Zhu and Chen, 2011). Additionally, low values of COD/N can result in incomplete denitritation

and therefore, N_2O accumulation via the heterotrophic denitrification pathway during the anoxic phase of the SBR. Accumulated N_2O in the end of the anoxic phase is stripped in the subsequent cycle, increasing the N_2O emissions. Caranto et al. (2016) have recently showed that N_2O can be the main product of anaerobic NH_2OH oxidation catalysed by the cytochrome P460 in N. europaea. The latter can be an evidence of the biological N_2O generation under limited DO and high NH3 concentrations, both conditions occurring in the target system in the during the transition from aerobic to anoxic phases when N_2O accumulation rapidly increases.

In this study, the average soluble COD concentration in the fermentation liquid was equal to 13082 mg COD/L over the monitoring period (

Table 5.2). Overall, in >27% of the examined cycles the N_2O was completely consumed by the end of the anoxic phase. Zhu and Chen, (2011), showed that the use of sludge fermentation liquid as carbon source in an anaerobic-aerobic system treating high-strength stream, can reduce the N_2O production by up to 68.7% compared to alternative carbon sources (i.e. acetic acid). On the other hand, Li et al., (2013a) in a process utilizing PHA as internal carbon source, observed higher N_2O production and reduction rates at higher influent COD concentrations linked with higher anaerobic PHA synthesis (ranging from 100 to 500 mg/L). The higher N_2O production rates were attributed to the accumulated NO_2^- inhibiting the N_2O reduction.

The dissolved N_2O concentration in the anoxic phase exceeded the calibration limit of the sensors; only cycles in which "nitrite knee" was observed and N_2O reduced to values lower than 2.6 mg/L could be investigated. Therefore, the effect on NO_2^- in anoxic N_2O generation could not be studied. However, studies have shown that elevated NO_2^- concentrations during denitrification can reduce the denitrification rate and increase the N_2O accumulation (Schulthess et al., 1995). The electron competition between nitrite reductase NIR, nitric oxide reductase (NOR) and nitrous oxide reductace (NOS) is intensified under high NO_2^- concentrations; NOS is less competitive under limitation of electron donor and this will result in N_2O accumulation (Pan et al., 2013; Ren et al., 2019).

Based on the profiles shown in Figure 5.5, N_2O was always consumed after the depletion of NO_2^- during denitritation; specifically, dissolved N_2O concentration decreased after the "nitrite knee". Gabarró et al. (2014), studied a partial-nitritation reactor treating landfill leachate, and operated under alternating aerobic/anoxic conditions to allow heterotrophic denitritation. The authors demonstrated that significant N_2O accumulation was observed during anoxic periods. NO_2^- denitrification rate was higher under both biodegradable COD limiting conditions and after acetate addition compared to N_2O reduction; N_2O reduction rate was maximum after NO_2^- removal (similar to what was observed in this study). In denitrifying phosphorus removal processes, Li et al. (2013) showed that the N_2O accumulation can be higher compared to conventional denitrification; the authors suggested that in the electron competition between denitrifying enzymes and PHA, N_2O

reductase is less competitive. On the other hand, Ribera-Guardia et al. (2016) investigated the electron competition during denitrification (PHA as the sole carbon source) of enriched dPAO and dGAO biomass and found that higher N_2O accumulation in the latter culture. Additionally, the last step of denitrification was inhibited in dGAO cultures (N_2O accumulation up to ~84% of the N-reduced), under high levels of NO_2^- (~ 15 mgN/gVSS) whereas N_2O consumption in dPAO biomass was not affected. Wang et al., (2015) demonstrated that during denitrifying phosphorus removal, mitigation of NO_2^- accumulation is possible via continuous dosage of phosphate and nitrate. Wang et al., (2011), showed that optimisation of the synthesis of PHA during the anaerobic phase can mitigate the N_2O production during the anoxic phase leading to complete denitrification.

In the system, N_2O emissions and dissolved N_2O concentration in the aerobic phase is strongly related with incomplete denitritation in the previous cycle. In ~26% of the cycles with incomplete denitritation, the N_2O concentration did not decrease below ~2 mg/L in the anaerobic phase and therefore the stripping of accumulated N_2O in the subsequent aerobic phase was substantial. Figure 5.10 (a) shows representative profiles of the dissolved N_2O concentration and the N_2O emissions based on different initial concentrations of N_2O in the beginning of the aerobic phase. The profiles of the ORP, DO and pH are comparable in the preseted cycles (Figure 5.10 (b)). In cycle B ~0.56 kgN of N_2O were emitted during the aerobic phase, wheareas in cycle A N_2O emissions are equal to 0.33 kgN (given the duration of these cycles is not equal only 220 min were considered). The initial dissolved N_2O concentration in cycles A and B is equal to 0.27 and >2.6 mg/L respectively. The N_2O emissions increased significantly due to the accumulated N_2O at the beginning of the previous anoxic phase that was stripped at the beginning of aeration.

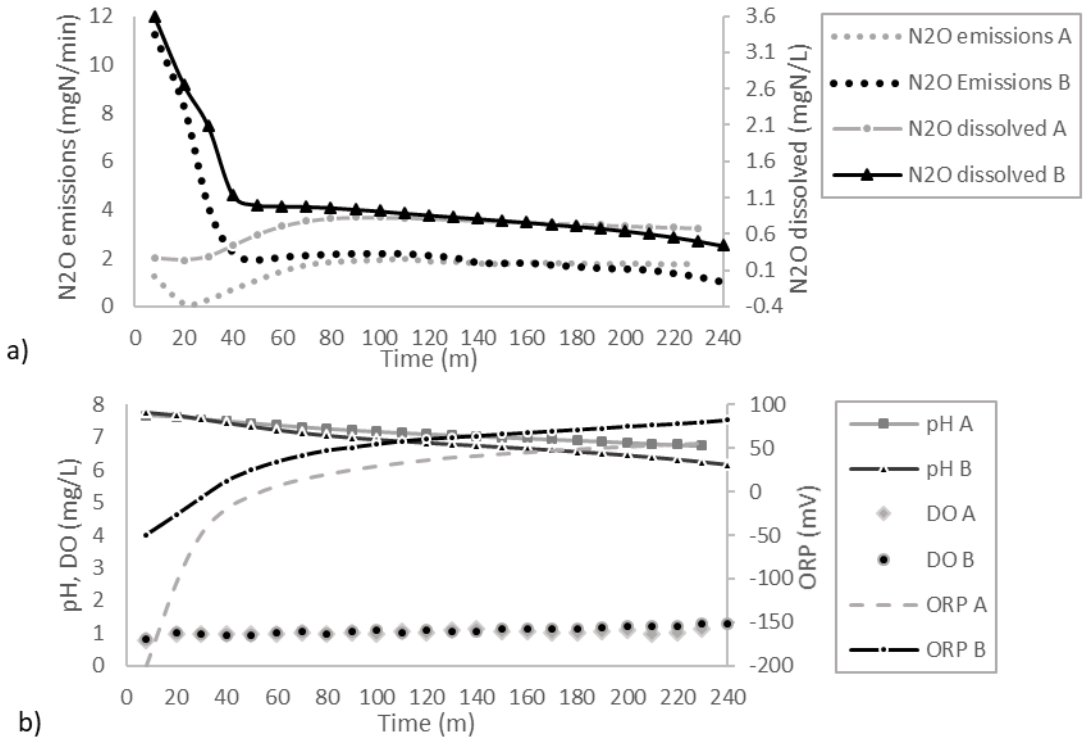


Figure 5.10: Representative profiles of ORP and DO (top) and dissolved N_2O concentration (bottom) based on different initial concentrations of N_2O in the beginning of the aerobic phase

Overall, in ~72% of the cycles, the dissolved N_2O concentration at the beginning of the anaerobic phase was higher than 0.3 mg/L. In cycles with dissolved N_2O concentration higher than 0.3 mg/L at the beginning of the anaerobic phase, the change in dissolved N_2O concentration during the anaerobic phase was highly correlated with the ORP at the beginning of the anaerobic phase. Additionally, the spearman correlation coefficient between the magnitude of the ORP reduction and magnitude of the dissolved N_2O reduction was equal to 0.7. Figure 5.11 shows the boxplots of dissolved N_2O reduction in relation to initial anaerobic ORP and ORP change for two cases: i) negligible dissolved N_2O change mainly due to influent dilution or anaerobic dissolved N2O concentration > 2.6 mg/L, and ii) occasions with N_2O reduction during the anaerobic phase. In Figure 5.11 (a) only occassions with ORP decrease higher than -50 mV are shown. The presence of nitrites in the bulk liquid during the (anaerobic) phase affected the ORP. NO_2-N depletion in the bulk liquid resulted in a sharp "nitrite knee" in the ORP profile (similar to the one

observed during the anoxic phase. Therefore, higher ORP change was expected in cycles with NO_2-N depletion and N_2O consumption during the anaerobic phase.

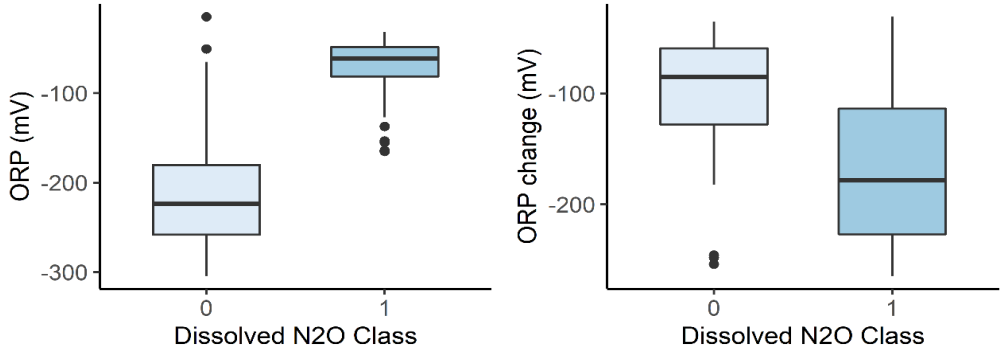


Figure 5.11: : Box-plots of the (a) initial anaerobic ORP and (b) the ORP change during the anaerobic phase for cycles with and without N_2O consumption (Class 0: no significant N_2O consumption or anaerobic N_2O concentration > 2.6 mg/L; Class 1: significant N_2O consumption)

Anaerobic phase term, is used to describe the first operational phase of the SBR (Figure 5.1) within each cycle and might not represent the actual conditions in the reactor. For instance, ORP ~ -80 mV in the anaerobic phase of the SBR indicates anoxic conditions, due to residual NO_2^-N concentration from the previous anoxic phase of the reactor.

5.4.8 *Prediction and control of N_2O accumulation in the anoxic and anaerobic phases*

As shown in section 5.4.7, the behaviour of ORP was significantly related with the behaviour of NO_2^- and consequentially of the dissolved N_2O concentration during the anaerobic phase. Therefore, in the ANSVM model, features related with the ORP profile were mainly used (Table 5.3). The classification results are shown in in the following section. Similarly, there was a strong link with the ORP behaviour and the "nitrite knee" with the N_2O accumulation during the anoxic phase. The features considered are shown in Table 5.3. Similarly, there was a strong link with the ORP behaviour and the nitrite "knee"

with the N_2O accumulation during the anoxic phase. The features considered in ANOXSVM model are shown in Table 5.3.

Table 5.3: Features used in the classification algorithm to predict the accumulation of dissolved N_2O at the end of the anoxic and anaerobic phases

Anaerobic	**Anoxic**	**Anaerobic regression**
ORP phase initial	Last ORP value	ORP phase initial
ORP change	ORP change	ORP change
First local maximum ORP first derivative	Mean pH	
Local minimum of ORP first derivative after first local maximum ORP first derivative	Difference between first local maximum (after carbon dosage) and subsequent local minimum of the ORP first derivative	pH phase initial
Duration between first local maximum and subsequent local minimum of the ORP first derivative	Duration of carbon dosage	Time of ORP first derivative minimum/duration of phase
pH phase initial	Duration between first local maximum (after carbon dosage) and subsequent local minimum of the ORP first derivative	Difference between first local maximum and subsequent local minimum of the ORP first derivative
Time local minimum ORP first derivative/Phase duration	Last ORP first derivate	

The classification matrices for train and test datasets of the ANSVM and ANOXSVM models are presented in Table 5.4. The average classification accuracy for the ANOXSVM model, was equal to 99% and 97% for the test and validation datasets. Similar results were obtained for the anaerobic phase with 95% and 98% accuracy in the train and test datasets respectively.

Jaramillo et al. (2018) developed an SVM classifier to estimate online the end of partial nitrification in a laboratory aerobic-anoxic SBR based on features extracted from pH and DO sensors over time-windows, resulting in 7.52% reduction in the operational time. In this study, the main focus was to estimate offline the behaviour of N_2O emissions based on historical batch data. The results from this study indicate that ORP and pH sensor data can be used to detect the consumption of N_2O during the nitritation/nitrification in SBR

reactors. The results show that knowledge-based feature-extraction and SVM classification could help in explaining the behaviour of the system and potentially optimise the control to consider the consumption of accumulated N_2O (i.e. in this system the denitritation can be stopped after the local maximum of the ORP rate after the "nitrite knee" in all the cycles investigated.)

Table 5.4: SVM classification results anaerobic phase

Phase	Dataset	Misclassified	Sensitivity	Specificity	Accuracy (%)	Kappa	Class
Anoxic phase cycle N	Train	anoxA: 1 anoxB: 0	1	0.99	99	0.97	anoxA: Final dissolved N_2O concentration end of anoxic < 0.6 mg/L anoxB: Final dissolved N_2O concentration end of anoxic > 0.6 mg/L
	Test	anoxA: 1 anoxB: 0	1	0.98	98	0.92	
Anaerobic phase cycle N+1	Train	anaerA:2 anaerB: 1	0.98	0.97	97	0.94	anaerA: N_2O end of anaerobic > 2.6 mg/L anaerB: N_2O end of anaerobic < 2.6 mg/L
	Test	anaerA: 1 anaerB: 0	1	0.97	98	0.95	

Figure 5.12 (a) and (b) illustrates the predicted and measured N_2O concentration at the end of the anaerobic phase (ANSVR model). The SVR parameters were optimised based on the root mean square error using the train dataset. RMSE of the SVR model was equal to 0.11 and 0.1 mg N_2O-N/L for the train and test datasets respectively (R-squared equal to 0.85 and 0.75 respectively). As shown in Figure 5.12 (b) the simulation results follow the behaviour of the actual dissolved N_2O concentrations observed. One of the major factors affecting the performance is the limited number of data points, but the prediction is still accurate.

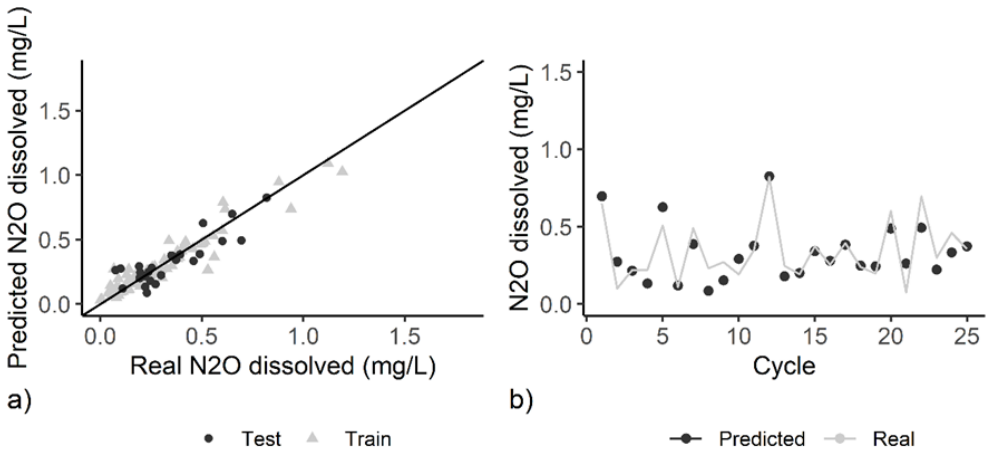


Figure 5.12: a) Predicted vs measured dissolved N_2O concentration in the end of the anaerobic phase (ANSVR) for the test and train datasets and (b) comparison of predicted and measured dissolved N_2O concentration for the test dataset

5.4.9 *Prediction of the N_2O concentration in aerobic phase*

The input features are shown in Table 5.5 and were selected based on the identified influential variables. The N_2O predicted values of the ANSVR model were used (anaerP). The procedure followed for the selection of model parameters was similar to the respective one followed for the anaerobic phase. Additionally, ANSVR test dataset cycles, were identified and used in AERSVR test dataset A. The model was also applied in anaerB cycles (test dataset B).

Table 5.5: Features selected in the SVR model for the aerobic phase

Aerobic Features
Average conductivity rate
ORP end of aeration
ORP increase during aeration
Conductivity at the beginning of aeration
Average DO
pH at the beginning of aeration
Conductivity increase (based on the conductivity at the end of the aerobic phase of the previous cycle)
Initial aerobic N_2O concentration (based on ANSVR predictions)

Figure 5.13 (a), shows the predicted and measured average aerobic N_2O concentration for the trained and test datasets. RMSE of the SVR model was equal to 0.06 and 0.11 mg N_2O-N /L for the train dataset and test dataset A respectively, whereas the R-squared was equal to 0.94 and 0.82 (Figure 10 (a) and (b)).

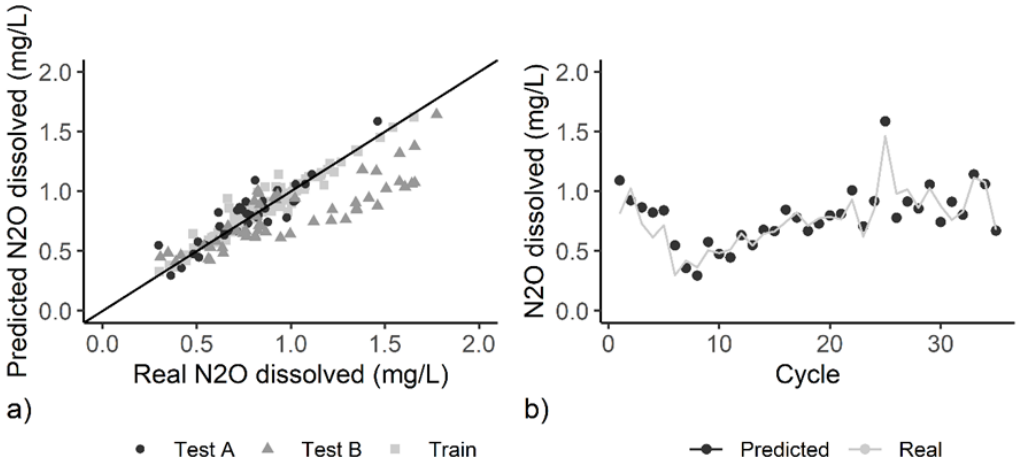


Figure 5.13: (a) Predicted vs measured dissolved N2O concentration (AERSVR) in the aerobic phase for the train dataset, the test dataset A and the test dataset B and (b) comparison of predicted and measured dissolved N2O concentration for the test dataset B

The RMSE of the predicted values for the test dataset B, was equal to 0.29 mg N_2O-N/L and the R-squared was equal to 0.72 (Figure 10 (a)). The AERSVR model underpredicted the average dissolved N_2O concentration of test B dataset. This is expected given that in test B dataset cycles, the initial aerobic N_2O accumulation exceeds the sensor calibration

limit. Therefore, on many occasions the initial aerobic N2O accumulation was also underestimated (section 5.3.1 - anaerB cycles). An example is shown if Figure 5.14. In cycle A, the average dissolved N_2O concentration (calculated as discussed in section 2.2.1 for anaerB cycles) is equal to 1.33 mg/L. The AERSVR model predicted 0.87 mg/L underestimating the actual concentration (considering initial accumulation equal to 0.6 mg/L). In cycle B, the AERSVR model predicted N_2O concentration equal to 0.61 mg/L (considering initial accumulation equal to 0.6); the observed average dissolved N_2O concentration (after the local minimum), was equal to 0.6 mg/L.

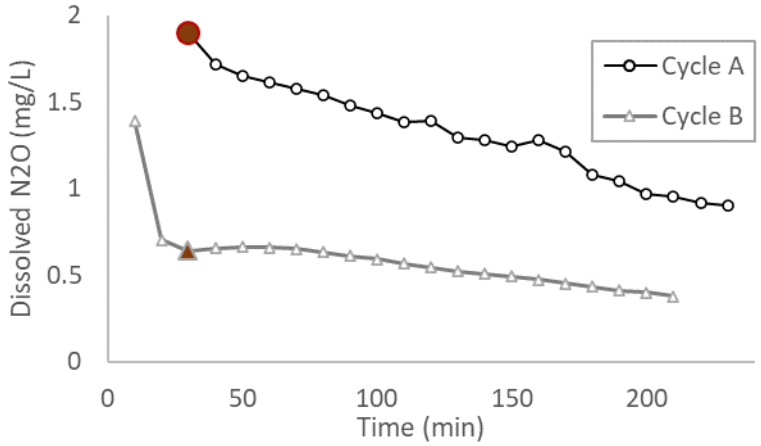


Figure 5.14: An example of dissolved N2O profiles for cycles belonging to anaerB cycles (test dataset B). The red points represent the first point considered for the calculation of the average aerobic N_2O accumulation (as described in section 5.3.1). Data points in the beginning of aeration exceeding sensor calibration limits are not shown.

The results show that under the investigated operational conditions, the framework shown in Figure 5.2**Σφάλμα! Το αρχείο προέλευσης της αναφοράς δεν βρέθηκε.** can provide a good estimation of the real dissolved N_2O behaviour and concentration observed during the different phases of SBR operation. Instabilities in the performance of machine learning models due to changes in the operational conditions in wastewater bioreactors have been reported in the literature (Shi and Xu, 2018). Therefore, long-term datasets and investigation of different patterns and dependencies should be investigated before model construction.

5.4.10 *Mitigation measures*

During aerobic phases, elevated average dissolved N_2O concentration was linked with DO less than 1 mg/L and increased conductivity decrease rates (conductivity values represent NH_4-N concentration values in the reactor). Therefore, cycles with increased conductivity decrease rate indicate higher NH_4-N removal efficiency and NO_2^-N accumulation. Dissolved N_2O concentrations lower than 0.6 mg/L were identified in cycles with average DO concentration equal to ~1.36 mg/L, and conductivity decrease rate > 1.8 μS/cm/min. Increasing the reactor DO concentration to values higher than 1.3 mg/L can result in decreased aerobic N_2O generation (Law et al., 2012). However, with the current anaerobic supernatant feeding strategy, blowers operate at maximum flowrate, so it is not possible to increase the aeration in the system.

On the other hand, the implementation of a step-feeding strategy could foster the reduction of N_2O emissions thanks to the lower NH_4-N and free ammonia (FA) concentration at the beginning of the cycle, which has been recognized as a triggering factor for N_2O production (Desloover et al., 2012). Conductivity at the end of the cycle can act as surrogate to estimate the effluent NH_4-N concentration of the reactor and optimize the anaerobic supernatant feeding load. Consequently, the aerobic initial NH_4-N concentration could be controlled to avoid either FA accumulation or high AOR with subsequent N_2O generation.

Additionally, frequent alternation of aerobic/anoxic phases can be introduced in order to avoid high nitrite accumulation. The impact of nitrite concentration on N_2O production can be also minimized by ensuring adequate DO levels within the reactor to inhibit the nitrifiers denitrification pathway (Blum et al., 2018; Law et al., 2013). Rodriguez-Caballero et al. (2015) reported that in a full-scale SBR treating municipal wastewater, intermittent aeration (alternation between 20–30 min oxic and anoxic) led to a minimization of N_2O compared to long oxic periods that enhanced N_2O emission. The authors related this behaviour to the presence of shorter aeration times with subsequently lower nitrite accumulation and N_2O production.

In addition, Su et al. (2019) reported that slightly acidic or neutral pH in nitritation reactors (at values that do not inhibit microbial activity) can decrease N_2O generation by up to seven times. Based on the pH profiles observed in this study, regulation of aerobic (alkalinity consumption) phase duration can be also considered to control the pH at lower levels.

The developed models can be used to estimate rapidly and precisely the hard-to-measure N_2O concentrations during aeration and detect N_2O accumulation in non-aerated phases. Additionally, it can alert operators about cycles with anoxic and anaerobic N_2O accumulation and elevated aerobic N_2O concentrations, that require modifications to the system operation. The ANOXSVM model can predict if N_2O is consumed in anoxic phases or if anoxic duration should be extended. Thus, additional provision of fermentation liquid can be performed to promote N_2O consumption through denitritation, when after 70-90 minutes the anoxic SVM model still indicates incomplete denitritation.

This study provides evidence on the relationship of DO, ORP and conductivity and pH with the dissolved N_2O concentration (in terms of correlation coefficients, behaviour and thresholds that indicate specific ranges of N_2O accumulation). These findings together with the models developed in this study, can be the basis for the development of intelligent control algorithms to integrate emissions control in sidestream SBR reactors performing nitritation/partial nitritation or other systems similar to SCENA. Moreover, features based on ORP, pH, DO and conductivity measurements in wastewater SBR processes, that can be used to predict dissolved N_2O concentrations have been identified. The developed framework can be also tested in continuous processes for the data-driven prediction of N_2O emissions.

5.5 Summary of main findings

In this study the behavior of N_2O emissions in a full-scale SBR reactor treating the anaerobic supernatant was investigated. Knowledge discovery and data-mining techniques

were employed to extract useful information on N_2O dynamics in the system and predict the behaviour of dissolved N_2O concentration.

- The N_2O emissions in SCENA process varies from 1.3% to 19% of NH_4-N load, therefore they can contribute considerably to the operational carbon footprint of the process (~90% on average).
- Average aerobic dissolved N_2O concentration could significantly under similar influent loads, DO, pH and removal efficiencies. Extracting information from the dynamic in-cycle behaviour of the variables monitored in the system is a significant step towards understanding N_2O behaviour.
- Aerobic dissolved N_2O concentration peaks (>1 mg/L), were observed in cycles with average DO concentrations less than 0.9-1 mg/L and ORP concentration at the end of the aerobic phase less than 40 mV. Conductivity was correlated with the reactor NH_4-N concentration (0.97). N_2O peaks were also observed in cycles with elevated decrease of conductivity during aeration. Step-feeding, control of initial NH_4-N concentrations and control of pH via the regulation of aerobic phase duration can mitigate the N_2O peaks observed in this study.
- The accumulation of N_2O at the end of the SBR anoxic phase was stripped in the subsequent aerobic phase and had a significant impact on the amount of N_2O emitted. The accumulated N_2O was consumed rapidly after nitrite 'knee' that was linked with the nitrite depletion. The ANOXSVM model can be used to detect if anoxic duration should be extended or additional fermentation liquid provided to enhance N_2O consumption in anoxic phases.
- This study shows that low-cost sensors, conventionally used to monitor SBR systems (i.e. pH, DO, ORP), have good capabilities to predict the dissolved N_2O behaviour and concentrations when couple with knowledge discovery techniques. The AERSVR model, showed reliable estimations of the aerobic N_2O concentration and can provide guidance to WWTPs operators, on whether N_2O levels are acceptable or mitigation actions are required.

6. Conclusions and Recommendations for future research

The data generated from sensors and laboratory measurements can provide valuable information for understanding, control and mitigation of direct N_2O emissions generated at wastewater treatment plants (WWTPs). This thesis demonstrated that the application of standardised data-mining frameworks to data obtained from wastewater treatment plants coupled with expert knowledge, can extract this information and translate it to actionable insights, supporting strategic plans for sustainable WWTP operation.

The examined techniques for outliers detection, clustering, classification and regression were able to i) detect and isolate re-occurring process disturbances ii) provide insights in the long-term dynamic behaviour of N_2O emissions, iii) link the behaviour of operational variables with specific ranges of N_2O emissions, iv) predict N_2O emission fluxes and risks for elevated N_2O emissions and v) provide feedback to the design of N_2O monitoring campaigns for the reliable estimation of N_2O emission factors.

Process-based N_2O EF benchmarking is challenging due to i) differences in the N_2O generation triggered by the site-specific operational characteristics, environmental conditions and control parameters and ii) the sensitivity of the quantified EF to differences in monitoring strategies and duration of monitoring campaigns. The analysis of the N_2O emission factors (EF), quantified in over 70 full-scale wastewater treatment processes revealed that the frequency and sampling techniques applied in N_2O monitoring campaigns, can impact significantly the quantified EFs (Chapter 2). EF boxplots were developed for different groups of process that can be used to compare new processes in terms of their EF magnitude. On average 0.87% of the N-load is emitted in the form of N_2O in the investigated mainstream processes. However, when considering the duration of the monitoring campaign, studies lasting over a year result in a median EF equal to 1.7 % of the N-load. On the other hand, most of the monitoring campaigns lasting less than one month have reported EFs less than 0.3 % of the N-load. Therefore, short-term monitoring periods may fail to capture underlying seasonal variations in the N_2O formation (or be affected by short-term process perturbations), and, consequently, result into unreliable EFs.

Similarly, the analysis showed that studies monitoring N_2O emissions in mainstream wastewater processes continuously (i.e. online via gas analysers), have quantified higher N_2O EFs than studies monitoring N_2O emissions discontinuously (i.e. offline via grab-samples). The average EF of mainstream wastewater processes monitored continuously and discontinuously is 1.2% and 0.44% of the N-load respectively. Low-frequency sampling campaigns have a high risk to not capture sufficiently short-term changes in pollutant concentrations, operational conditions and system disturbances impacting N_2O generation.

An N_2O EF database has been developed in this work, including information on configurations, control strategies, operational conditions, monitoring strategies, generation mechanisms and mitigation measures (subject to the information reported in the studies) for different processes and process-groups.

Two major drawbacks were identified in the conventional techniques used for N_2O data management and analysis. Firstly, the application of simple descriptive statistics, simple univariate and bi-variate graphical representations of operational variables, cannot adequately assess the combined effect of operational variables on N_2O generation. Secondly, visualizations of multivariate long-term timeseries can become overcrowded and cluttered, hindering significant events and short-term dependencies and limiting the knowledge that can be extracted from the wastewater sensor signals. The development of structured approaches that utilise readily available wastewater data (i.e. from sensors and actuators) to extract information concerning the N_2O emission patterns, combined with advanced visualization and dimensionality reduction techniques, can facilitate the interpretation of the long-term N_2O emissions behaviour.

Therefore, efforts were focused towards the development of a methodological framework for knowledge discovery from wastewater treatment datasets that can be applied and supplement current data management and analysis practices. Data-driven pattern recognition and modelling techniques were used to extract insights on the long-term N_2O triggering operational conditions and dependencies with environmental and operational conditions from different full-scale biological processes.

Multivariate statistical techniques were used to extract information from the long-term N_2O monitoring campaign of a full-scale Carrousel reactor (Chapter 3). The analysis showed that data-mining techniques can be used to assist researchers and operators to detect, understand and visualize the temporal behaviour and characteristics of the operational and environmental variables monitored online and their impact on N_2O formation. Additionally, the segmentation of the system based on differences in the behaviour of N_2O emissions enabled the detection of strong and varying local dependencies with the operational variables that were not visible when the complete timeseries were considered. The investigation of dependencies between N_2O emissions and operational variables in biological processes needs to account for i) the temporal variability of operational and environmental conditions that result into changes of the N_2O triggering mechanisms, ii) system disturbances that can influence short-term (i.e. 1 day) or even for longer periods (i.e. one week) both the system performance and the N_2O generation and ii) the combined effect of the operational variables on N_2O emissions.

Specifically, in the investigated system, fluctuating dependencies of the N_2O emissions with the operational variables were quantified, for temporal segments characterised by different range and pattern of N_2O emissions. Therefore, abrupt structural changes in the profile of N_2O emissions can be a sign of changes in the underlying mechanisms or environmental conditions triggering the N_2O generation. Hierarchical k-means clustering and principal component analysis (PCA) applied in the different segments, provided insights on the combined operational conditions linked with specific ranges of N_2O emissions and isolated sets of variables that relate to N_2O emissions. Additionally, the behaviour and ranges of the operational variables in the clusters can guide towards the identification of possible N_2O triggering mechanisms. For instance, clusters characterised with high dissolved oxygen (DO) and peaks in nitrite and nitrate concentrations indicated insufficient denitrification zones in the reactor.

Additionally, findings of this research showed that discretisation of the operational behavior of biological processes (based on the patterns of the variables monitored online and seasonal effects), can be used to predict the range of N_2O emissions. Additionally, the

frequency and duration of the sampling requirements, for a reliable estimation of N_2O EF, can be reduced (Chapter 4).

Specifically, the investigated Carrousel reactor, structural changes in the profiles of operational variables were linked with changes in the N_2O emissions range. Simulation results showed that even three-day sampling campaigns between the detected changepoints have a high probability (~80%) to result to an EF quantification with less than 10% error. Conventionally applied monitoring strategies with equivalent duration (i.e. random sampling, monthly sampling) have significantly higher probabilities to result in high errors in the EF quantification and are more likely to underestimate the EF in the investigated system.

Events with abnormal diurnal behaviour of the influent flow-rate and effluent NH_4-N concentrations (linked with precipitation events and poor system performance) were isolated and categorized based on duration/intensity using a combination of feature extraction and density-based clustering. This can facilitate the investigation of N_2O emission behaviour under different system disturbances. In the investigated reactor, the N_2O response for the observed system disturbances varied mainly based on the temperature, the magnitude of influent flow-rate peaks and duration of the events, the NH_4-N concentration and the local N_2O emission patterns.

Finally, the analysis showed that support vector machine (SVM) classification models can be trained to detect operational behaviour of the system. The classification model, constructed in this work, predicted the state of the system (based on the segments identified by the changepoint detection method). Since the different segments were characterised by relatively stationary N_2O fluxes, the SVM predicted classes provided a good approximation of the expected range of N_2O emission loads.

The results indicate that analysis of historical data and investigation of seasonal effects can be of paramount importance in the planning of monitoring campaigns sampling frequency and duration. The proposed approach can be applied when long-term online sampling is not feasible (due to budget or equipment limitations) to identify N_2O emissions “hotspot”

periods and guide towards the identification of the operational periods requiring extensive investigation of N_2O pathways and mitigation measures.

The developed framework for knowledge discovery was applied to a full-scale sidestream SBR, treating the anaerobic supernatant (SCENA) (Chapter 5). Overall, considerable direct N2O emissions, were quantified in the SCENA process (7.6% of the NH_4-N load). Therefore, understanding and control of N_2O emissions is significant in order to improve the sustainability of the process. Conductivity, oxidation reduction potential, pH and DO can provide useful insights on the N_2O emissions behaviour, as long as i) the different cycle phases are analysed separately, ii) cycles with poor performance are detected and isolated and iii) different operational conditions especially in the aerobic phase are identified and clustered together. SVM classification and regression models were trained to predict whether dissolved N_2O will be consumed during the anoxic and anaerobic phases and the average dissolved N_2O concentration during aeration. The applied methodology can provide a good estimation of the real dissolved N_2O behaviour and concentration during the different phases of SBR operation.

The analysis showed that, investigation of different patterns and dependencies is a significant step before the development of machine learning regression models. Introduction of periods characterised by significant changes in the operational conditions can deteriorate model performance.

6.1 Recommendations for future work

In this thesis, efforts were focused to investigate how data mining techniques for detection of abnormal events, pattern recognition, classification and regression can be used to extract information hidden in the raw sensor signals and provide insights on the long-term N_2O emissions' dynamics in different wastewater treatment processes with different control strategies. Main aspect of the analysis was to investigate the nature of actionable information that can be extracted when these techniques are used. The results showed that a structured framework for knowledge discovery from wastewater databases supported by a combination of domain knowledge and data-mining techniques can be used to support WWTP operation and facilitate the integration of sustainability metrics in the decision making.

The current thesis showed that information extracted from several conventional wastewater sensors can provide valuable insights on the long-term N_2O behaviour. Further research is required investigating the optimal sensor location and the optimal combination of variables monitored, for N_2O emissions' control for different wastewater configurations. Wastewater processes are characterised by non-stationarity, high dynamics and variations at different scales in time. Here, the development of novel methods and standardised frameworks that inherently consider these features while still being practicable is due.

Future research can also explore the possibility of coupling sophisticated statistical tools (e.g. multivariate statistics, machine learning algorithms) with multiple-pathway mechanistic models for full-scale applications, to facilitate the fast and adaptable online implementation of model predictive control and forecasting decision support tools. For instance, machine learning models trained at the residuals of mechanistic models can enhance the generalisation capabilities of these models. On the other hand, computationally universal mechanistic models can be used to simulate variables not conventionally monitored in WWTPs (i.e. NO_2^-, NO). Simulations of key variables coupled with the raw sensor signals can be used in the knowledge discovery process to enhance the reliability of the findings and improve the generalisation capabilities of the data-driven models. The

development of layer based on machine learning techniques can be integrated in a universal complex N_2O model to facilitate individual process parameters calibration. Multivariate statistics and pattern recognition algorithms can be applied to the variables monitored online in WWTPs to differentiate operational conditions and guide towards different calibration requirements within the same process. Finally, multivariate statistics can be applied to identify and isolate complex relationships between system variables and guide towards process-specific simplified modelling approaches. Such integrated, practical tools can help plant operators to design effective mitigation strategies.

Additionally, several aspects need to be considered before the integration of data-mining techniques into the data management practices of water utilities.

Water utilities have been dominated by traditional operations focusing on long-term investments and continuity. Historically, water utilities have separate departments doing separate jobs. Data analysis and algorithmic calculations on all data of all departments are not performed; standardised approaches are missing. The techniques developed in the current thesis need to be advanced to practical tools and interfaces that can provide the desired information at a simple and intuitive way. For instance, segmentation, clustering, classification and regression techniques integrated into SCADA systems can be used to benchmark and predict key performance indicators (KPIs) (performance, cost, environmental aspects) under different operational conditions (i.e. based on seasonal influent composition variations, different process rates affected by environmental conditions, system shocks etc). User-friendly dashboards, communicating the results in a simple and informative will help operators to detect operational modes in which the system is underperforming, analyse risks and prioritise optimisation needs providing a platform for continuous internal multivariate benchmarking of WWTP performance.

References

Abeel, T., Helleputte, T., Van de Peer, Y., Dupont, P., Saeys, Y., 2010. Robust biomarker identification for cancer diagnosis with ensemble feature selection methods. Bioinforma. Oxf. Engl. 26, 392–398. https://doi.org/10.1093/bioinformatics/btp630

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. N2O 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., 2010a. 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

Ahn, J.H., Kim, S., Park, H., Rahm, B., Pagilla, K., Chandran, K., 2010b. N2O Emissions from Activated Sludge Processes, 2008−2009: Results of a National Monitoring Survey in the United States. Environ. Sci. Technol. 44, 4505–4511. https://doi.org/10.1021/es903845y

Aigner, W., Miksch, S., Müller, W., Schumann, H., Tominski, C., 2008. Visual Methods for Analyzing Time-Oriented Data. IEEE Trans. Vis. Comput. Graph. 14, 47–60. https://doi.org/10.1109/TVCG.2007.70415

Allen, D.E., McALEER, M., Powell, R.J., Singh, A.K., 2018. Non-parametric multiple change point analysis of the global financial crisis. Ann. Financ. Econ. 13, 1850008. https://doi.org/10.1142/S2010495218500082

Antico, A., Schlotthauer, G., Torres, M.E., 2014. Analysis of hydroclimatic variability and trends using a novel empirical mode decomposition: Application to the Paraná River Basin [WWW Document]. J. Geophys. Res. Atmospheres. https://doi.org/10.1002/2013JD020420

APHA, 2005. Standard methods for the examination of water and wastewater. WEF 21, 258–259.

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

Arnell, M., Rahmberg, M., Oliveira, F., Jeppsson, U., 2017. Multi-objective performance assessment of wastewater treatment plants combining plant-wide process models and life cycle assessment. J. Water Clim. Change 8, 715–729. https://doi.org/10.2166/wcc.2017.179

Ballard, S., Porro, J., Trommsdorff, C., 2018. The roadmap to a Low Carbon Urban Water Utility: An international guide to the WaCCliM approach. IWA Publishing, Place of publication not identified.

Baresel, C., Andersson, S., Yang, J., Andersen, M.H., 2016. Comparison of nitrous oxide (N2O) emissions calculations at a Swedish wastewater treatment plant based on water concentrations versus off-gas concentrations. Adv. Clim. Change Res., Including special topic on atmospheric black carbon and its effects on cryosphere 7, 185–191. https://doi.org/10.1016/j.accre.2016.09.001

Batstone, D.J., Hülsen, T., Mehta, C.M., Keller, J., 2015. Platforms for energy and nutrient recovery from domestic wastewater: A review. Chemosphere 140, 2–11.

Bellandi, G., Porro, J., Senesi, E., Caretti, C., Caffaz, S., Weijers, S., Nopens, I., Gori, R., 2018. Multi-point monitoring of nitrous oxide emissions in three full-scale conventional activated sludge tanks in Europe. Water Sci. Technol. 77, 880–890. https://doi.org/10.2166/wst.2017.560

Blomberg, K., Kosse, P., Mikola, A., Kuokkanen, A., Fred, T., Heinonen, M., Mulas, M., Lübken, M., Wichern, M., Vahala, R., 2018. Development of an Extended ASM3 Model for Predicting the Nitrous Oxide Emissions in a Full-Scale Wastewater Treatment Plant. Environ. Sci. Technol. 52, 5803–5811. https://doi.org/10.1021/acs.est.8b00386

Blum, J.-M., Jensen, M.M., Smets, B.F., 2018. Nitrous oxide production in intermittently aerated Partial Nitritation-Anammox reactor: oxic N2O production dominates and relates with ammonia removal rate. Chem. Eng. J. 335, 458–466. https://doi.org/10.1016/j.cej.2017.10.146

Bollon, J., Filali, A., Fayolle, Y., Guerin, S., Rocher, V., Gillot, S., 2016a. Full-scale post denitrifying biofilters: sinks of dissolved N2O? Sci. Total Environ. 563–564, 320–328. https://doi.org/10.1016/j.scitotenv.2016.03.237

Bollon, J., Filali, A., Fayolle, Y., Guerin, S., Rocher, V., Gillot, S., 2016b. N2O emissions from full-scale nitrifying biofilters. Water Res. 102, 41–51. https://doi.org/10.1016/j.watres.2016.05.091

Breiman, L., 2001. Random Forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324

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.

Byvatov, E., Schneider, G., 2003. Support vector machine applications in bioinformatics. Appl. Bioinformatics 2, 67–77.

Cadwallader Adam, VanBriesen Jeanne M., 2017. Incorporating Uncertainty into Future Estimates of Nitrous Oxide Emissions from Wastewater Treatment. J. Environ. Eng. 143, 04017029. https://doi.org/10.1061/(ASCE)EE.1943-7870.0001231

Caivano, M., Bellandi, G., Mancini, I.M., Masi, S., Brienza, R., Panariello, S., Gori, R., Caniani, D., 2017. Monitoring the aeration efficiency and carbon footprint of a medium-sized WWTP: experimental results on oxidation tank and aerobic digester. Environ. Technol. 38, 629–638. https://doi.org/10.1080/09593330.2016.1205150

Caranto, J.D., Vilbert, A.C., Lancaster, K.M., 2016. Nitrosomonas europaea cytochrome P460 is a direct link between nitrification and nitrous oxide emission. Proc. Natl. Acad. Sci. 113, 14704–14709. https://doi.org/10.1073/pnas.1611051113

Castellano-Hinojosa, A., Maza-Márquez, P., Melero-Rubio, Y., González-López, J., Rodelas, B., 2018. Linking nitrous oxide emissions to population dynamics of nitrifying and denitrifying prokaryotes in four full-scale wastewater treatment plants. Chemosphere 200, 57–66. https://doi.org/10.1016/j.chemosphere.2018.02.102

Castro-Barros, C.M., Daelman, M.R.J., Mampaey, K.E., van Loosdrecht, M.C.M., Volcke, E.I.P., 2015. Effect of aeration regime on N2O emission from partial nitritation-anammox in a full-scale granular sludge reactor. Water Res. 68, 793–803. https://doi.org/10.1016/j.watres.2014.10.056

Cawley, G.C., Talbot, N.L., Girolami, M., 2007. Sparse Multinomial Logistic Regression via Bayesian L1 Regularisation, in: Schölkopf, B., Platt, J.C., Hoffman, T. (Eds.), Advances in Neural Information Processing Systems 19. MIT Press, pp. 209–216.

Çelik, M., Dadaşer-Çelik, F., Dokuz, A.Ş., 2011. Anomaly detection in temperature data using DBSCAN algorithm, in: 2011 International Symposium on Innovations in Intelligent Systems and Applications. Presented at the 2011 International Symposium on Innovations in Intelligent Systems and Applications, pp. 91–95. https://doi.org/10.1109/INISTA.2011.5946052

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

Chen, W.-H., Yang, J.-H., Yuan, C.-S., Yang, Y.-H., 2016. Toward better understanding and feasibility of controlling greenhouse gas emissions from treatment of industrial wastewater with activated sludge. Environ. Sci. Pollut. Res. 23, 20449–20461. https://doi.org/10.1007/s11356-016-7183-2

Colominas, M.A., Schlotthauer, G., Torres, M.E., Flandrin, P., 2012. Noise-assisted emd methods in action. Adv. Adapt. Data Anal. 04, 1250025. https://doi.org/10.1142/S1793536912500252

COM, 2017. Ninth Report on the implementation status and the programmes for implementation (as required by Article 17) of Council Directive 91/271/EEC concerning urban waste water treatment (No. 749). European Commission.

Conthe, M., Parchen, C., Stouten, G., Kleerebezem, R., van Loosdrecht, M.C.M., 2018. O2 versus N2O respiration in a continuous microbial enrichment. Appl. Microbiol. Biotechnol. 102, 8943–8950. https://doi.org/10.1007/s00253-018-9247-3

Cordell, D., Drangert, J.-O., White, S., 2009. The story of phosphorus: global food security and food for thought. Glob. Environ. Change 19, 292–305.

Cornejo, P.K., Zhang, Q., Mihelcic, J.R., 2016. How Does Scale of Implementation Impact the Environmental Sustainability of Wastewater Treatment Integrated with Resource Recovery? Environ. Sci. Technol. 50, 6680–6689. https://doi.org/10.1021/acs.est.5b05055

Corominas, L., Larsen, H.F., Flores-Alsina, X., Vanrolleghem, P.A., 2013. Including Life Cycle Assessment for decision-making in controlling wastewater nutrient removal systems. J. Environ. Manage. 128, 759–767. https://doi.org/10.1016/j.jenvman.2013.06.002

Corominas, L., Villez, K., Aguado, D., Rieger, L., Rosén, C., Vanrolleghem, P.A., 2011. Performance evaluation of fault detection methods for wastewater treatment processes. Biotechnol. Bioeng. 108, 333–344. https://doi.org/10.1002/bit.22953

Corominas, Ll., Garrido-Baserba, M., Villez, K., Olsson, G., Cortés, U., Poch, M., 2018. Transforming data into knowledge for improved wastewater treatment operation: A critical review of techniques. Environ. Model. Softw., Special Issue on Environmental Data Science. Applications to Air quality and Water cycle 106, 89–103. https://doi.org/10.1016/j.envsoft.2017.11.023

Cortes, C., Vapnik, V., 1995. Support-vector networks. Mach. Learn. 20, 273–297. https://doi.org/10.1007/BF00994018

Crill, P.M., Butler, J.H., Cooper, D.J., Novelli, P.C., 1995. Standard analytical methods for measuring trace gases in the environment. Biog. Trace Gases Meas. Emiss. Soil Water 164–205.

Curtis, T.P., 2010. Low-Energy Wastewater Treatment: Strategies and Technologies, in: Environmental Microbiology. John Wiley & Sons, Ltd, pp. 301–318. https://doi.org/10.1002/9780470495117.ch13

Daelman, M.R.J., van Loosdrecht, M.C., Volcke, E.I., De Baets, B., 2013a. Influence of sampling strategies on the estimated nitrous oxide emission from wastewater treatment plants. Water Res. 47, 3120–3130.

Daelman, M.R.J., van Voorthuizen, E.M., Van Dongen, L., Volcke, E.I.P., Van Loosdrecht, M.C.M., 2013b. Methane and nitrous oxide emissions from municipal wastewater treatment–results from a long-term study. Water Sci. Technol. 67, 2350–2355.

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 N2O 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

Daniel, J.S., Fleming, E.L., Portmann, R.W., Velders, G.J.M., Jackman, C.H., Ravishankara, A.R., 2010. Options to accelerate ozone recovery: ozone and climate benefits. Atmospheric Chem. Phys. 10, 7697–7707. https://doi.org/10.5194/acp-10-7697-2010

de Faria, A.B., Spérandio, M., Ahmadi, A., Tiruta-Barna, L., 2015. Evaluation of new alternatives in wastewater treatment plants based on dynamic modelling and life cycle assessment (DM-LCA). Water Res. 84, 99–111.

Delre, A., Mønster, J., Samuelsson, J., Fredenslund, A.M., Scheutz, C., 2018. Emission quantification using the tracer gas dispersion method: The influence of instrument, tracer gas species and source simulation. Sci. Total Environ. 634, 59–66. https://doi.org/10.1016/j.scitotenv.2018.03.289

Delre, A., Mønster, J., Scheutz, C., 2017. Greenhouse gas emission quantification from wastewater treatment plants, using a tracer gas dispersion method. Sci. Total Environ. 605–606, 258–268. https://doi.org/10.1016/j.scitotenv.2017.06.177

Denmead, O.T., 2008. Approaches to measuring fluxes of methane and nitrous oxide between landscapes and the atmosphere. Plant Soil 309, 5–24. https://doi.org/10.1007/s11104-008-9599-z

Desloover, J., De Clippeleir, H., Boeckx, P., Du Laing, G., Colsen, J., Verstraete, W., Vlaeminck, S.E., 2011. Floc-based sequential partial nitritation and anammox at full scale with contrasting N2O emissions. Water Res. 45, 2811–2821. https://doi.org/10.1016/j.watres.2011.02.028

Desloover, J., Vlaeminck, S.E., Clauwaert, P., Verstraete, W., Boon, N., 2012. Strategies to mitigate N2O emissions from biological nitrogen removal systems. Curr. Opin. Biotechnol. 23, 474–482. https://doi.org/10.1016/j.copbio.2011.12.030

Ding, X., Zhao, J., Hu, B., Li, X., Ge, G., Gao, K., Chen, Y., 2017. Mathematical modeling of nitrous oxide (N2O) production in anaerobic/anoxic/oxic processes: Improvements to published N2O models. Chem. Eng. J. 325, 386–395. https://doi.org/10.1016/j.cej.2017.05.082

Domingo-Félez, C., Calderó-Pascual, M., Sin, G., Plósz, B.G., Smets, B.F., 2017a. Calibration of the comprehensive NDHA-N2O dynamics model for nitrifier-enriched biomass using targeted respirometric assays. Water Res. 126, 29–39. https://doi.org/10.1016/j.watres.2017.09.013

Domingo-Félez, C., Mutlu, A.G., Jensen, M.M., Smets, B.F., 2014. Aeration Strategies To Mitigate Nitrous Oxide Emissions from Single-Stage Nitritation/Anammox Reactors. Environ. Sci. Technol. 48, 8679–8687. https://doi.org/10.1021/es501819n

Domingo-Félez, C., Pellicer-Nàcher, C., Petersen, M.S., Jensen, M.M., Plósz, B.G., Smets, B.F., 2017b. Heterotrophs are key contributors to nitrous oxide production in activated sludge under low C-to-N ratios during nitrification-Batch experiments and modeling. Biotechnol. Bioeng. 114, 132–140. https://doi.org/10.1002/bit.26062

Domingo-Félez, C., Smets, F.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

Duan, H., Ye, L., Erler, D., Ni, B.-J., Yuan, Z., 2017. Quantifying nitrous oxide production pathways in wastewater treatment systems using isotope technology – A critical review. Water Res. 122, 96–113. https://doi.org/10.1016/j.watres.2017.05.054

Duchemin, E., Lucotte, M., Canuel, R., 1999. Comparison of Static Chamber and Thin Boundary Layer Equation Methods for Measuring Greenhouse Gas Emissions from Large Water Bodies§. Environ. Sci. Technol. 33, 350–357. https://doi.org/10.1021/es9800840

Duda, R.O., Hart, P.E., Stork, D.G., 2001. Pattern classification 2nd ed. John Willey Sons Inc.

Ekici, S., 2009. Classification of power system disturbances using support vector machines. Expert Syst. Appl. 36, 9859–9868. https://doi.org/10.1016/j.eswa.2009.02.002

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

Environment Agency, 2010. Transforming wastewater treatment to reduce carbon emissions. Environment Agency, Bristol.

Eskicioglu, C., Galvagno, G., Cimon, C., 2018. Approaches and processes for ammonia removal from side-streams of municipal effluent treatment plants. Bioresour. Technol. 268, 797–810. https://doi.org/10.1016/j.biortech.2018.07.020

Ester, M., Kriegel, H.-P., Sander, J., Xu, X., 1996. A density-based algorithm for discovering clusters in large spatial databases with noise., in: Kdd. pp. 226–231.

European Commission, 2012a. Commission staff working documen. European Overview (1/2) Accompanying the document REPORT FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT AND THE COUNCIL on the Implementation of the Water Framework Directive (2000/60/EC) River Basin Management Plans.

European Commission, 2012b. Directive 2012/27/EU of the European Parliament and of the Council of 25 October 2012 on energy efficiency, amending Directives

2009/125/EC and 2010/30/EU and repealing Directives 2004/8/EC and 2006/32/EC Text with EEA relevance, 315.

European Commission, 1991. Council Directive of 21. May 1991 concerning urban waste water treatment (91/271/EEC).

European Communities, 2000. Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for community action in the field of water policy. Official journal of the European Communities L327(43), 1–72.

Fang, F., Ni, B., Li, W., Sheng, G., Yu, H., 2011. A simulation-based integrated approach to optimize the biological nutrient removal process in a full-scale wastewater treatment plant. Chem. Eng. J. 174, 635–643. https://doi.org/10.1016/j.cej.2011.09.079

Fayyad, U., Piatetsky-Shapiro, G., Smyth, P., 1996. The KDD process for extracting useful knowledge from volumes of data. Commun. ACM 39, 27–34. https://doi.org/10.1145/240455.240464

Filali, A., Fayolle, Y., Peu, P., Philippe, L., Nauleau, F., Gillot, S., 2013. Aeration control in a full-scale activated sludge wastewater treatment plant: impact on performances, energy consumption and N2O emission. Presented at the 11ème Conférence IWA sur l'instrumentation, le contrôle et l'automatisation. ICA2013, p. 4 p.

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

Flores-Alsina, X., Corominas, L., Snip, L., Vanrolleghem, P.A., 2011. Including greenhouse gas emissions during benchmarking of wastewater treatment plant control strategies. Water Res. 45, 4700–4710. https://doi.org/10.1016/j.watres.2011.04.040

Flores-Alsina, X., Gallego, A., Feijoo, G., Rodriguez-Roda, I., 2010. Multiple-objective evaluation of wastewater treatment plant control alternatives. J. Environ. Manage. 91, 1193–1201. https://doi.org/10.1016/j.jenvman.2010.01.009

Flowers, J.J., Cadkin, T.A., McMahon, K.D., 2013. Seasonal bacterial community dynamics in a full-scale enhanced biological phosphorus removal plant. Water Res. 47, 7019–7031. https://doi.org/10.1016/j.watres.2013.07.054

Foley, J., de Haas, D., Yuan, Z., Lant, P., 2010. Nitrous oxide generation in full-scale biological nutrient removal wastewater treatment plants. Water Res. 44, 831–844. https://doi.org/10.1016/j.watres.2009.10.033

Foley, J., Yuan, Z., Keller, J., Senante, E., Chandran, K., Willis, J., Shah, A., Loosdrecht, M.C.M. van, Voorthuizen, E. van, 2015. N2O and CH4 Emission from Wastewater Collection and Treatment Systems: State of the Science Report and Technical Report. IWA Publishing.

Frison, N., Katsou, E., Malamis, S., Oehmen, A., Fatone, F., 2015. Development of a novel process integrating the treatment of sludge reject water and the production of polyhydroxyalkanoates (PHAs). Environ. Sci. Technol. 49, 10877–10885.

Frutos, O.D., Quijano, G., Aizpuru, A., Muñoz, R., 2018. A state-of-the-art review on nitrous oxide control from waste treatment and industrial sources. Biotechnol. Adv. 36, 1025–1037. https://doi.org/10.1016/j.biotechadv.2018.03.004

Gabarró, J., González-Cárcamo, P., Ruscalleda, M., Ganigué, R., Gich, F., Balaguer, M.D., Colprim, J., 2014. Anoxic phases are the main N2O contributor in partial nitritation reactors treating high nitrogen loads with alternate aeration. Bioresour. Technol. 163, 92–99. https://doi.org/10.1016/j.biortech.2014.04.019

Ghoneim, W.A.M., Helal, A.A., Wahab, M.A., 2016. Minimizing energy consumption in Wastewater Treatment Plants, in: Renewable Energies for Developing Countries (REDEC), 2016 3rd International Conference On. IEEE, pp. 1–8.

Granitto, P.M., Furlanello, C., Biasioli, F., Gasperi, F., 2006. Recursive feature elimination with random forest for PTR-MS analysis of agroindustrial products. Chemom. Intell. Lab. Syst. 83, 83–90. https://doi.org/10.1016/j.chemolab.2006.01.007

Gruber, W., Villez, K., Kipf, M., Wunderlin, P., Siegrist, H., Vogt, L., Joss, A., 2019. N2O emission in full-scale wastewater treatment: Proposing a refined monitoring strategy. Sci. Total Environ. 134157. https://doi.org/10.1016/j.scitotenv.2019.134157

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, G., Wang, Y., Hao, T., Wu, D., Chen, G.-H., 2017. Enzymatic nitrous oxide emissions from wastewater treatment. Front. Environ. Sci. Eng. 12, 10. https://doi.org/10.1007/s11783-018-1021-3

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

Gustavsson, D.J.I., la Cour Jansen, J., 2011. Dynamics of nitrogen oxides emission from a full-scale sludge liquor treatment plant with nitritation. Water Sci. Technol. J. Int. Assoc. Water Pollut. Res. 63, 2838–2845.

Guyon, I., Weston, J., Barnhill, S., Vapnik, V., 2002. Gene Selection for Cancer Classification using Support Vector Machines. Mach. Learn. 46, 389–422. https://doi.org/10.1023/A:1012487302797

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

Hadjimichael, A., Comas, J., Corominas, L., 2016. Do machine learning methods used in data mining enhance the potential of decision support systems? A review for the urban water sector. AI Commun. 29, 747–756. https://doi.org/10.3233/AIC-160714

Hahsler, M., Piekenbrock, M., Arya, S., Mount, D., 2017. dbscan: Density Based Clustering of Applications with Noise (DBSCAN) and Related Algorithms. R Package Version 1–0.

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

Harper, W.F., Takeuchi, Y., Riya, S., Hosomi, M., Terada, A., 2015. Novel abiotic reactions increase nitrous oxide production during partial nitrification: Modeling and experiments. Chem. Eng. J. 281, 1017–1023. https://doi.org/10.1016/j.cej.2015.06.109

Harris, E., Joss, A., Emmenegger, L., Kipf, M., Wolf, B., Mohn, J., Wunderlin, P., 2015. Isotopic evidence for nitrous oxide production pathways in a partial nitritation-anammox reactor. Water Res. 83, 258–270. https://doi.org/10.1016/j.watres.2015.06.040

Haslinger, J., Lindtner, S., Krampe, J., 2016. Operating costs and energy demand of wastewater treatment plants in Austria: benchmarking results of the last 10 years. Water Sci. Technol. 74, 2620–2626. https://doi.org/10.2166/wst.2016.390

He, Z., Yu, W., 2010. Stable feature selection for biomarker discovery. Comput. Biol. Chem. 34, 215–225. https://doi.org/10.1016/j.compbiolchem.2010.07.002

Henderson, S.L., Dandie, C.E., Patten, C.L., Zebarth, B.J., Burton, D.L., Trevors, J.T., Goyer, C., 2010. Changes in Denitrifier Abundance, Denitrification Gene mRNA Levels, Nitrous Oxide Emissions, and Denitrification in Anoxic Soil Microcosms Amended with Glucose and Plant Residues. Appl Env. Microbiol 76, 2155–2164. https://doi.org/10.1128/AEM.02993-09

Hensen, A., Skiba, U., Famulari, D., 2013. Low cost and state of the art methods to measure nitrous oxide emissions. Environ. Res. Lett. 8, 025022.

Henze, M., Grady Jr, C.L., Gujer, W., Marais, G., Matsuo, T., 1987. Activated Sludge model no. 1. IAWQ Scientific and Technical Report No. 1 IAWQ. Lond. UK.

Henze, M., Gujer, W., Mino, T., Van Loosdrecht, M.C.M., 2000. Activated sludge models ASM1, ASM2, ASM2d and ASM3. IWA publishing.

Hiatt, W.C., Grady, C.P.L., 2008. An Updated Process Model for Carbon Oxidation, Nitrification, and Denitrification [WWW Document]. https://doi.org/info:doi/10.2175/106143008X304776

Hochstein, L.I., Tomlinson, G.A., 1988. The Enzymes Associated with Denitrification. Annu. Rev. Microbiol. 42, 231–261. https://doi.org/10.1146/annurev.mi.42.100188.001311

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.

Hooper, A.B., 1968. A nitrite-reducing enzyme from Nitrosomonas europaea Preliminary characterization with hydroxylamine as electron donor. Biochim. Biophys. Acta BBA - Bioenerg. 162, 49–65. https://doi.org/10.1016/0005-2728(68)90213-2

Hooper, A.B., Terry, K.R., 1979. Hydroxylamine oxidoreductase of Nitrosomonas: Production of nitric oxide from hydroxylamine. Biochim. Biophys. Acta BBA - Enzymol. 571, 12–20. https://doi.org/10.1016/0005-2744(79)90220-1

Huang Norden E., Shen Zheng, Long Steven R., Wu Manli C., Shih Hsing H., Zheng Quanan, Yen Nai-Chyuan, Tung Chi Chao, Liu Henry H., 1998. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proc. R. Soc. Lond. Ser. Math. Phys. Eng. Sci. 454, 903–995. https://doi.org/10.1098/rspa.1998.0193

Hwang, K.-L., Bang, C.-H., Zoh, K.-D., 2016. Characteristics of methane and nitrous oxide emissions from the wastewater treatment plant. Bioresour. Technol. 214, 881–884. https://doi.org/10.1016/j.biortech.2016.05.047

IEA, 2016. Water-energy nexus World Energy Outlook 2016 (Paris: International Energy Agency).

INTERNATIONAL FERTILIZER INDUSTRY ASSOCIATION, 2001. Global estimates of gaseous emissions of NH3, NO and N2O from agricultural land. Food and Agriculture Organization of the United Nations (FAO).

IPCC, 2019. Refinement to the 2006 IPCC Guidelines for National Greenhouse gas Inventories. Chapter 5.6 wastewater treatment and discharge.

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.

IPCC, 2006. 2006 IPCC Guidelines for National Greenhouse Gas Inventories. Intergovernmental Panel on Climate Change.

Ishii, S., Song, Y., Rathnayake, L., Tumendelger, A., Satoh, H., Toyoda, S., Yoshida, N., Okabe, S., 2014. Identification of key nitrous oxide production pathways in aerobic partial nitrifying granules. Environ. Microbiol. 16, 3168–3180. https://doi.org/10.1111/1462-2920.12458

Itokawa, H., Hanaki, K., Matsuo, T., 2001. Nitrous oxide production in high-loading biological nitrogen removal process under low cod/n ratio condition. Water Res. 35, 657–664. https://doi.org/10.1016/S0043-1354(00)00309-2

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

Jalori, S., Reddy, T.A., 2015. A new clustering method to identify outliers and diurnal schedules from building energy interval data, in: ASHRAE Transactions. Presented at the 2015 ASHRAE Annual Conference, Amer. Soc. Heating, Ref. Air-Conditoning Eng. Inc., pp. 33–44.

James, N.A., Matteson, D.S., 2013. ecp: An R Package for Nonparametric Multiple Change Point Analysis of Multivariate Data. ArXiv13093295 Stat.

Janus, H.M., van der Roest, H.F., 1997. Don't reject the idea of treating reject water. Water Sci. Technol. 35, 27–34. https://doi.org/10.2166/wst.1997.0351

Jaramillo, F., Orchard, M., Muñoz, C., Antileo, C., Sáez, D., Espinoza, P., 2018. On-line estimation of the aerobic phase length for partial nitrification processes in SBR based on features extraction and SVM classification. Chem. Eng. J. 331, 114–123. https://doi.org/10.1016/j.cej.2017.07.185

Jeppsson, U., Rosen, C., Alex, J., Copp, J., Gernaey, K.V., Pons, M.N., Vanrolleghem, P.A., 2006. Towards a benchmark simulation model for plant-wide control strategy performance evaluation of WWTPs. Water Sci. Technol. J. Int. Assoc. Water Pollut. Res. 53, 287–295.

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.

Joss, A., Salzgeber, D., Eugster, J., König, R., Rottermann, K., Burger, S., Fabijan, P., Leumann, S., Mohn, J., Siegrist, H., 2009. Full-Scale Nitrogen Removal from Digester Liquid with Partial Nitritation and Anammox in One SBR. Environ. Sci. Technol. 43, 5301–5306. https://doi.org/10.1021/es900107w

Kalousis, A., Prados, J., Hilario, M., 2007. Stability of feature selection algorithms: a study on high-dimensional spaces. Knowl. Inf. Syst. 12, 95–116. https://doi.org/10.1007/s10115-006-0040-8

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, 2008a. 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

Kampschreur, M.J., van der Star, W.R.L., Wielders, H.A., Mulder, J.W., Jetten, M.S.M., van Loosdrecht, M.C.M., 2008b. Dynamics of nitric oxide and nitrous oxide emission during full-scale reject water treatment. Water Res. 42, 812–826. https://doi.org/10.1016/j.watres.2007.08.022

Karatzoglou, A., Smola, A., Hornik, K., Zeileis, A., 2004. kernlab - An S4 Package for Kernel Methods in R. J. Stat. Softw. 11, 1–20.

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

Kohavi, R., John, G.H., 1997. Wrappers for feature subset selection. Artif. Intell., Relevance 97, 273–324. https://doi.org/10.1016/S0004-3702(97)00043-X

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

Kosse, P., Lübken, M., Schmidt, T.C., Wichern, M., 2017. Quantification of nitrous oxide in wastewater based on salt-induced stripping. Sci. Total Environ. 601–602, 83–88. https://doi.org/10.1016/j.scitotenv.2017.05.053

Kotta, J., Herkül, K., Jaagus, J., Kaasik, A., Raudsepp, U., Alari, V., Arula, T., Haberman, J., Järvet, A., Kangur, K., Kont, A., Kull, A., Laanemets, J., Maljutenko, I., Männik, A., Nõges, P., Nõges, T., Ojaveer, H., Peterson, A., Reihan, A., Rõõm, R., Sepp, M., Suursaar, Ü., Tamm, O., Tamm, T., Tõnisson, H., 2018. Linking atmospheric, terrestrial and aquatic environments: Regime shifts in the Estonian climate over the past 50 years. PLOS ONE 13, e0209568. https://doi.org/10.1371/journal.pone.0209568

Kuhn, M., 2008. Building Predictive Models in R Using the caret Package. J. Stat. Softw. 28. https://doi.org/10.18637/jss.v028.i05

Kursa, M.B., 2014. Robustness of Random Forest-based gene selection methods. BMC Bioinformatics 15, 8. https://doi.org/10.1186/1471-2105-15-8

Lackner, S., Gilbert, E.M., Vlaeminck, S.E., Joss, A., Horn, H., van Loosdrecht, M.C., 2014. Full-scale partial nitritation/anammox experiences–an application survey. Water Res. 55, 292–303.

Lane, J.L., de Haas, D.W., Lant, P.A., 2015. The diverse environmental burden of city-scale urban water systems. Water Res. 81, 398–415. https://doi.org/10.1016/j.watres.2015.03.005

Law, Y., Lant, P., Yuan, Z., 2013. The Confounding Effect of Nitrite on N2O Production by an Enriched Ammonia-Oxidizing Culture. Environ. Sci. Technol. 47, 7186–7194. https://doi.org/10.1021/es4009689

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

Lee, D.S., Jeon, C.O., Park, J.M., Chang, K.S., 2002. Hybrid neural network modeling of a full-scale industrial wastewater treatment process. Biotechnol. Bioeng. 78, 670–682. https://doi.org/10.1002/bit.10247

Li, C., Wang, T., Zheng, N., Zhang, J., Ngo, H.H., Guo, W., Liang, S., 2013a. Influence of organic shock loads on the production of N2O in denitrifying phosphorus removal process. Bioresour. Technol., Challenges in Environmental Science and Engineering (CESE-2012) 141, 160–166. https://doi.org/10.1016/j.biortech.2013.03.117

Li, C., Zhang, J., Liang, S., Ngo, H.H., Guo, W., Zhang, Y., Zou, Y., 2013b. Nitrous oxide generation in denitrifying phosphorus removal process: main causes and control measures. Environ. Sci. Pollut. Res. 20, 5353–5360. https://doi.org/10.1007/s11356-013-1530-3

Li, H., Peng, D., Liu, W., Wei, J., Wang, Z., Wang, B., 2016. N2O generation and emission from two biological nitrogen removal plants in China. Desalination Water Treat. 57, 11800–11806. https://doi.org/10.1080/19443994.2015.1046145

Li, Y., Liu, Y., Zhao, L., Hastings, A., Guo, H., 2015. Exploring change of internal nutrients cycling in a shallow lake: A dynamic nutrient driven phytoplankton

model. Ecol. Model. 313, 137–148. https://doi.org/10.1016/j.ecolmodel.2015.06.025

Liu, H., Motoda, H., 1998. Feature Extraction, Construction and Selection: A Data Mining Perspective. Springer Science & Business Media.

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 (N2O) 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

Liu, Y.-J., Gu, J., Liu, Y., 2018. Energy self-sufficient biological municipal wastewater reclamation: Present status, challenges and solutions forward. Bioresour. Technol. 269, 513–519. https://doi.org/10.1016/j.biortech.2018.08.104

Longo, S., d'Antoni, B.M., Bongards, M., Chaparro, A., Cronrath, A., Fatone, F., Lema, J.M., Mauricio-Iglesias, M., Soares, A., Hospido, A., 2016. Monitoring and diagnosis of energy consumption in wastewater treatment plants. A state of the art and proposals for improvement. Appl. Energy 179, 1251–1268. https://doi.org/10.1016/j.apenergy.2016.07.043

Lorke, A., Bodmer, P., Noss, C., Alshboul, Z., Koschorreck, M., Somlai-Haase, C., Bastviken, D., Flury, S., Mcginnis, D.F., Maeck, A., Müller, D., Premke, K., 2015. Technical note: drifting versus anchored flux chambers for measuring greenhouse gas emissions from running waters. Biogeosciences 12, 7013–7024.

Luts, J., Ojeda, F., Van de Plas, R., De Moor, B., Van Huffel, S., Suykens, J.A.K., 2010. A tutorial on support vector machine-based methods for classification problems in chemometrics. Anal. Chim. Acta 665, 129–145. https://doi.org/10.1016/j.aca.2010.03.030

Ma, C., Jensen, M.M., Smets, B.F., Thamdrup, B., 2017. Pathways and Controls of N2O Production in Nitritation–Anammox Biomass. Environ. Sci. Technol. 51, 8981–8991. https://doi.org/10.1021/acs.est.7b01225

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

Maktabifard, M., Zaborowska, E., Makinia, J., 2018. Achieving energy neutrality in wastewater treatment plants through energy savings and enhancing renewable energy production. Rev. Environ. Sci. Biotechnol. 17, 655–689. https://doi.org/10.1007/s11157-018-9478-x

Mampaey, K.E., Beuckels, B., Kampschreur, M.J., Kleerebezem, R., van Loosdrecht, M.C. m., Volcke, E.I. p., 2013. Modelling nitrous and nitric oxide emissions by autotrophic ammonia-oxidizing bacteria. Environ. Technol. 34, 1555–1566. https://doi.org/10.1080/09593330.2012.758666

Mampaey, K.E., De Kreuk, M.K., van Dongen, U.G.J.M., van Loosdrecht, M.C.M., Volcke, E.I.P., 2016. Identifying N2O 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

Mampaey, K.E., Spérandio, M., van Loosdrecht, M.C.M., Volcke, E.I.P., 2019. Dynamic simulation of N2O emissions from a full-scale partial nitritation reactor. Biochem. Eng. J. 152, 107356. https://doi.org/10.1016/j.bej.2019.107356

Mampaey, K.E., van Dongen, U.G., van Loosdrecht, M.C., Volcke, E.I., 2015. Novel method for online monitoring of dissolved N2O concentrations through a gas stripping device. Environ. Technol. 36, 1680–1690.

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, 254–270.

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.

Massara, T.M., Solís, B., Guisasola, A., Katsou, E., Baeza, J.A., 2018. Development of an ASM2d-N2O model to describe nitrous oxide emissions in municipal WWTPs under dynamic conditions. Chem. Eng. J. 335, 185–196.

Matthews, C.J.D., St.Louis, V.L., Hesslein, R.H., 2003. Comparison of Three Techniques Used To Measure Diffusive Gas Exchange from Sheltered Aquatic Surfaces. Environ. Sci. Technol. 37, 772–780. https://doi.org/10.1021/es0205838

Mello, W.Z. de, Ribeiro, R.P., Brotto, A.C., Kligerman, D.C., Piccoli, A. de S., Oliveira, J.L.M., 2013. Nitrous oxide emissions from an intermittent aeration activated sludge system of an urban wastewater treatment plant. Quím. Nova 36, 16–20. https://doi.org/10.1590/S0100-40422013000100004

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.

Mo, W., Zhang, Q., 2013. Energy–nutrients–water nexus: integrated resource recovery in municipal wastewater treatment plants. J. Environ. Manage. 127, 255–267.

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

Mountrakis, G., Im, J., Ogole, C., 2011. Support vector machines in remote sensing: A review. ISPRS J. Photogramm. Remote Sens. 66, 247–259. https://doi.org/10.1016/j.isprsjprs.2010.11.001

Mulder, A., 2003. The quest for sustainable nitrogen removal technologies. Water Sci. Technol. 48, 67–75.

Nanopoulos, A., Alcock, R., Manolopoulos, Y., 2001. Feature-based classification of time-series data. Int. J. Comput. Res. 10, 49–61.

Newhart, K.B., Holloway, R.W., Hering, A.S., Cath, T.Y., 2019. Data-driven performance analyses of wastewater treatment plants: A review. Water Res. 157, 498–513. https://doi.org/10.1016/j.watres.2019.03.030

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

Ni, B.-J., Peng, L., Law, Y., Guo, J., Yuan, Z., 2014. Modeling of Nitrous Oxide Production by Autotrophic Ammonia-Oxidizing Bacteria with Multiple Production Pathways. Environ. Sci. Technol. 48, 3916–3924. https://doi.org/10.1021/es405592h

Ni, B.-J., Ye, L., Law, Y., Byers, C., Yuan, Z., 2013. Mathematical Modeling of Nitrous Oxide (N2O) Emissions from Full-Scale Wastewater Treatment Plants. Environ. Sci. Technol. 47, 7795–7803. https://doi.org/10.1021/es4005398

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.

Palut, M.P.J., Canziani, O.F., 2007. Contribution of working group II to the fourth assessment report of the intergovernmental panel on climate change. Cambridge University Press.

Pan, Y., Ni, B.-J., Bond, P.L., Ye, L., Yuan, Z., 2013. Electron competition among nitrogen oxides reduction during methanol-utilizing denitrification in wastewater treatment. Water Res. 47, 3273–3281. https://doi.org/10.1016/j.watres.2013.02.054

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

Peng, H., Long, F., Ding, C., 2005. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. IEEE Trans. Pattern Anal. Mach. Intell. 1226–1238.

Peng, L., Carvajal-Arroyo, J.M., Seuntjens, D., Prat, D., Colica, G., Pintucci, C., Vlaeminck, S.E., 2017. Smart operation of nitritation/denitritation virtually abolishes nitrous oxide emission during treatment of co-digested pig slurry centrate. Water Res. 127, 1–10. https://doi.org/10.1016/j.watres.2017.09.049

Peng, L., Ni, B.-J., Erler, D., Ye, L., Yuan, Z., 2014. The effect of dissolved oxygen on N2O production by ammonia-oxidizing bacteria in an enriched nitrifying sludge. Water Res. 66, 12–21. https://doi.org/10.1016/j.watres.2014.08.009

Peng, L., Ni, B.-J., Ye, L., Yuan, Z., 2015. The combined effect of dissolved oxygen and nitrite on N 2 O production by ammonia oxidizing bacteria in an enriched nitrifying sludge. Water Res. 73, 29–36.

Pes, B., Dessì, N., Angioni, M., 2017. Exploiting the ensemble paradigm for stable feature selection: A case study on high-dimensional genomic data. Inf. Fusion 35, 132–147. https://doi.org/10.1016/j.inffus.2016.10.001

Pijuan, M., Torà, J., Rodríguez-Caballero, A., César, E., Carrera, J., Pérez, J., 2014. Effect of process parameters and operational mode on nitrous oxide emissions from a nitritation reactor treating reject wastewater. Water Res. 49, 23–33. https://doi.org/10.1016/j.watres.2013.11.009

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.

Porro, J., Kampschreur, M.J., Pijuan, M., Volcke, E., Daelman, M., Guo, L., Nopens, I., Vanrolleghem, P.A., Yuan, Z., Chandran, K., others, 2014. Measuring nitrous oxide emissions from biological wastewater treatment, art or science?, in: IWA Specialist Conference on Global Challenges: Sustainable Wastewater Treatment and Resource Recovery. International Water Association (IWA).

Portmann, R.W., Daniel, J.S., Ravishankara, A.R., 2012. Stratospheric ozone depletion due to nitrous oxide: influences of other gases. Phil Trans R Soc B 367, 1256–1264. https://doi.org/10.1098/rstb.2011.0377

Poth, M., Focht, D.D., 1985. 15N Kinetic Analysis of N2O Production by Nitrosomonas europaea: an Examination of Nitrifier Denitrification. Appl Env. Microbiol 49, 1134–1141.

Puchongkawarin, C., Gomez-Mont, C., Stuckey, D.C., Chachuat, B., 2015. Optimization-based methodology for the development of wastewater facilities for energy and nutrient recovery. Chemosphere 140, 150–158.

Puyol, D., Batstone, D.J., Hülsen, T., Astals, S., Peces, M., Krömer, J.O., 2017. Resource Recovery from Wastewater by Biological Technologies: Opportunities, Challenges, and Prospects. Front. Microbiol. 7. https://doi.org/10.3389/fmicb.2016.02106

Qiao, J., Zhang, W., 2018. Dynamic multi-objective optimization control for wastewater treatment process. Neural Comput. Appl. 29, 1261–1271. https://doi.org/10.1007/s00521-016-2642-8

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/. (accessed 12.12.17).

Raghavendra, N.S., Deka, P.C., 2014. Support vector machine applications in the field of hydrology: A review. Appl. Soft Comput. 19, 372–386. https://doi.org/10.1016/j.asoc.2014.02.002

Rapson, T.D., Dacres, H., 2014. Analytical techniques for measuring nitrous oxide. TrAC Trends Anal. Chem. 54, 65–74. https://doi.org/10.1016/j.trac.2013.11.004

Rathnayake, R.M., Oshiki, M., Ishii, S., Segawa, T., Satoh, H., Okabe, S., 2015. Effects of dissolved oxygen and pH on nitrous oxide production rates in autotrophic partial nitrification granules. Bioresour. Technol. 197, 15–22.

Ravishankara, A.R., Daniel, J.S., Portmann, R.W., 2009. Nitrous oxide (N2O): the dominant ozone-depleting substance emitted in the 21st century. science 326, 123–125.

Reino, C., van Loosdrecht, M.C.M., Carrera, J., Pérez, J., 2017. Effect of temperature on N2O emissions from a highly enriched nitrifying granular sludge performing partial nitritation of a low-strength wastewater. Chemosphere 185, 336–343. https://doi.org/10.1016/j.chemosphere.2017.07.017

Ren, Y. g., Wang, J. h., Li, H. f., Zhang, J., Qi, P. y., Hu, Z., 2013. Nitrous oxide and methane emissions from different treatment processes in full-scale municipal wastewater treatment plants. Environ. Technol. 34, 2917–2927. https://doi.org/10.1080/09593330.2012.696717

Ren, Y., Ngo, H.H., Guo, W., Ni, B.-J., Liu, Y., 2019. Linking the nitrous oxide production and mitigation with the microbial community in wastewater treatment: A review. Bioresour. Technol. Rep. 7, 100191. https://doi.org/10.1016/j.biteb.2019.100191

Ribeiro, R.P., Bueno, R.F., Piveli, R.P., Kligerman, D.C., de Mello, W.Z., Oliveira, J.L.M., 2017. The response of nitrous oxide emissions to different operating conditions in activated sludge wastewater treatment plants in Southeastern Brazil. Water Sci. Technol. 76, 2337–2349. https://doi.org/10.2166/wst.2017.399

Ribera-Guardia, A., Marques, R., Arangio, C., Carvalheira, M., Oehmen, A., Pijuan, M., 2016. Distinctive denitrifying capabilities lead to differences in N2O production by denitrifying polyphosphate accumulating organisms and denitrifying glycogen

accumulating organisms. Bioresour. Technol. 219, 106–113. https://doi.org/10.1016/j.biortech.2016.07.092

Ritchie, G. a. F., Nicholas, D.J.D., 1972. Identification of the sources of nitrous oxide produced by oxidative and reductive processes in Nitrosomonas europaea. Biochem. J. 126, 1181–1191. https://doi.org/10.1042/bj1261181

Robertson, G.P., Tiedje, J.M., 1987. Nitrous oxide sources in aerobic soils: Nitrification, denitrification and other biological processes. Soil Biol. Biochem. 19, 187–193. https://doi.org/10.1016/0038-0717(87)90080-0

Rodriguez-Caballero, A., Aymerich, I., Marques, R., Poch, M., Pijuan, M., 2015. Minimizing N2O emissions and carbon footprint on a full-scale activated sludge sequencing batch reactor. Water Res. 71, 1–10. https://doi.org/10.1016/j.watres.2014.12.032

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

Rodriguez-Galiano, V.F., Luque-Espinar, J.A., Chica-Olmo, M., Mendes, M.P., 2018. Feature selection approaches for predictive modelling of groundwater nitrate pollution: An evaluation of filters, embedded and wrapper methods. Sci. Total Environ. 624, 661–672. https://doi.org/10.1016/j.scitotenv.2017.12.152

Rodriguez-Garcia, G., Frison, N., Vázquez-Padín, J.R., Hospido, A., Garrido, J.M., Fatone, F., Bolzonella, D., Moreira, M.T., Feijoo, G., 2014. Life cycle assessment of nutrient removal technologies for the treatment of anaerobic digestion supernatant and its integration in a wastewater treatment plant. Sci. Total Environ. 490, 871–879. https://doi.org/10.1016/j.scitotenv.2014.05.077

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.

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

Saeedi Emadi, H., Mazinani, S.M., 2018. A Novel Anomaly Detection Algorithm Using DBSCAN and SVM in Wireless Sensor Networks. Wirel. Pers. Commun. 98, 2025–2035. https://doi.org/10.1007/s11277-017-4961-1

Saeys, Y., Abeel, T., Van de Peer, Y., 2008. Robust Feature Selection Using Ensemble Feature Selection Techniques, in: Daelemans, W., Goethals, B., Morik, K. (Eds.), Machine Learning and Knowledge Discovery in Databases, Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 313–325.

Saeys, Y., Inza, I., Larrañaga, P., 2007. A review of feature selection techniques in bioinformatics. Bioinformatics 23, 2507–2517. https://doi.org/10.1093/bioinformatics/btm344

Salcedo-Sanz, S., Cornejo-Bueno, L., Prieto, L., Paredes, D., García-Herrera, R., 2018. Feature selection in machine learning prediction systems for renewable energy applications. Renew. Sustain. Energy Rev. 90, 728–741. https://doi.org/10.1016/j.rser.2018.04.008

Samuelsson, J., Delre, A., Tumlin, S., Hadi, S., Offerle, B., Scheutz, C., 2018. Optical technologies applied alongside on-site and remote approaches for climate gas emission quantification at a wastewater treatment plant. Water Res. 131, 299–309. https://doi.org/10.1016/j.watres.2017.12.018

Sarni, W., White, C., Webb, R., Cross, K., Glotzbach, R., 2019. Industry leaders chart the transformation journey. International Water Association (IWA) and Xylem White Paper.

Schaubroeck, T., De Clippeleir, H., Weissenbacher, N., Dewulf, J., Boeckx, P., Vlaeminck, S.E., Wett, B., 2015. Environmental sustainability of an energy self-sufficient sewage treatment plant: Improvements through DEMON and co-digestion. Water Res. 74, 166–179. https://doi.org/10.1016/j.watres.2015.02.013

Schmidt, C.E., 1994. Theory and operation of the US EPA Surface Emission Isolation Flux Chamber at POTWs. One-day training course for Bureau of Sanitation. City Los Angel.

Schulthess, R.V., Gujer, W., 1996. Release of nitrous oxide (N2O) 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. v., Kühni, M., Gujer, W., 1995. Release of nitric and nitrous oxides from denitrifying activated sludge. Water Res. 29, 215–226. https://doi.org/10.1016/0043-1354(94)E0108-I

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.

Seshan, H., Goyal, M.K., Falk, M.W., Wuertz, S., 2014. Support vector regression model of wastewater bioreactor performance using microbial community diversity indices: Effect of stress and bioaugmentation. Water Res. 53, 282–296. https://doi.org/10.1016/j.watres.2014.01.015

Severn Trent Plc, 2013. Changing Course through the sustainable implementation of the Water Framework Directive. [WWW Document]. URL https://www.severntrent.com/content/dam/stw/ST_Corporate/About_us/Docs/ChangingCourse-WaterFrameworkDirectiveNov2013.pdf (accessed 9.26.19).

Severn Trent Plc, 2012. Sustainable Growth: Annual Report and Accounts 2012. Severn Trent Plc, Coventry, UK.

Sharma, M., Bajracharya, S., Gildemyn, S., Patil, S.A., Alvarez-Gallego, Y., Pant, D., Rabaey, K., Dominguez-Benetton, X., 2014. A critical revisit of the key parameters used to describe microbial electrochemical systems. Electrochimica Acta 140, 191–208. https://doi.org/10.1016/j.electacta.2014.02.111

Shi, S., Xu, G., 2018. Novel performance prediction model of a biofilm system treating domestic wastewater based on stacked denoising auto-encoders deep learning network. Chem. Eng. J. 347, 280–290. https://doi.org/10.1016/j.cej.2018.04.087

Shurkhovetskyy, G., Andrienko, N., Andrienko, G., Fuchs, G., 2018. Data Abstraction for Visualizing Large Time Series. Comput. Graph. Forum 37, 125–144. https://doi.org/10.1111/cgf.13237

Singh, D.P., Maurya, N.S., 2016. A Comparative Study of the Various Methodologies for Estimation of Green House Gas Emission from Wastewater Treatment Systems (A Review). Orient. J. Chem. 32, 1373–1380.

Singh, K.P., Basant, N., Gupta, S., 2011. Support vector machines in water quality management. Anal. Chim. Acta 703, 152–162. https://doi.org/10.1016/j.aca.2011.07.027

Soler-Jofra, A., Stevens, B., Hoekstra, M., Picioreanu, C., Sorokin, D., van Loosdrecht, M.C.M., Pérez, J., 2016. Importance of abiotic hydroxylamine conversion on nitrous oxide emissions during nitritation of reject water. Chem. Eng. J. 287, 720–726. https://doi.org/10.1016/j.cej.2015.11.073

Song, K., Suenaga, T., Hamamoto, A., Satou, K., Riya, S., Hosomi, M., Terada, A., 2014. Abundance, transcription levels and phylogeny of bacteria capable of nitrous oxide reduction in a municipal wastewater treatment plant. J. Biosci. Bioeng. 118, 289–297. https://doi.org/10.1016/j.jbiosc.2014.02.028

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

Spérandio, M., Pocquet, M., Guo, L., Ni, B.-J., Vanrolleghem, P.A., Yuan, Z., 2016. Evaluation of different nitrous oxide production models with four continuous long-term wastewater treatment process data series. Bioprocess Biosyst. Eng. 39, 493–510. https://doi.org/10.1007/s00449-015-1532-2

Spinelli, M., Eusebi, A.L., Vasilaki, V., Katsou, E., Frison, N., Cingolani, D., Fatone, F., 2018. Critical analyses of nitrous oxide emissions in a full scale activated sludge

system treating low carbon-to-nitrogen ratio wastewater. J. Clean. Prod. 190, 517–524. https://doi.org/10.1016/j.jclepro.2018.04.178

Stenström, F., Tjus, K., Jansen, J. la C., 2014. Oxygen-induced dynamics of nitrous oxide in water and off-gas during the treatment of digester supernatant. Water Sci. Technol. 69, 84–91. https://doi.org/10.2166/wst.2013.558

Su, Q., Domingo-Félez, C., Zhang, Z., Blum, J.-M., Jensen, M.M., Smets, B.F., 2019. The effect of pH on N2O production in intermittently-fed nitritation reactors. Water Res. 156, 223–231. https://doi.org/10.1016/j.watres.2019.03.015

Sun, S., Bao, Z., Li, R., Sun, D., Geng, H., Huang, X., Lin, J., Zhang, P., Ma, R., Fang, L., Zhang, X., Zhao, X., 2017. Reduction and prediction of N2O emission from an Anoxic/Oxic wastewater treatment plant upon DO control and model simulation. Bioresour. Technol. 244, 800–809. https://doi.org/10.1016/j.biortech.2017.08.054

Sun, S., Bao, Z., Sun, D., 2015. Study on emission characteristics and reduction strategy of nitrous oxide during wastewater treatment by different processes. Environ. Sci. Pollut. Res. 22, 4222–4229. https://doi.org/10.1007/s11356-014-3654-5

Sun, S., Cheng, X., Sun, D., 2013. Emission of N2O 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

Sun, Y., Guan, Y., Pan, M., Zhan, X., Hu, Z., Wu, G., 2017. Enhanced biological nitrogen removal and N2O emission characteristics of the intermittent aeration activated sludge process. Rev. Environ. Sci. Biotechnol. 16, 761–780. https://doi.org/10.1007/s11157-017-9444-z

SWD, 2017. Commision Staff Working Document Accompanying the document. Ninth Report on the implementation status and the programmes for implementation (as required by Article 17) of Council Directive 91/271/EEC concerning urban waste water treatment. (No. 445). European Commission.

Sweetapple, C., Fu, G., Butler, D., 2014. Multi-objective optimisation of wastewater treatment plant control to reduce greenhouse gas emissions. Water Res. 55, 52–62. https://doi.org/10.1016/j.watres.2014.02.018

Tallec, G., Garnier, J., Billen, G., Gousailles, M., 2006. Nitrous oxide emissions from secondary activated sludge in nitrifying conditions of urban wastewater treatment plants: Effect of oxygenation level. Water Res. 40, 2972–2980. https://doi.org/10.1016/j.watres.2006.05.037

Tang, J., Alelyani, S., Liu, H., 2014. Feature Selection for Classification: A Review, in: Data Classification: Algorithms and Applications. https://doi.org/10.1201/b17320-3

Terada, A., Sugawara, S., Hojo, K., Takeuchi, Y., Riya, S., Harper, W.F., Yamamoto, T., Kuroiwa, M., Isobe, K., Katsuyama, C., Suwa, Y., Koba, K., Hosomi, M., 2017. Hybrid Nitrous Oxide Production from a Partial Nitrifying Bioreactor:

Hydroxylamine Interactions with Nitrite. Environ. Sci. Technol. 51, 2748–2756. https://doi.org/10.1021/acs.est.6b05521

Tesoriero, A.J., Gronberg, J.A.M., Juckem, P.F., Miller, M.P., Austin, B.P., 2017. Predicting redox-sensitive contaminant concentrations in groundwater using random forest classification. Water Resour. Res. 53, 73167331. https://doi.org/10.1002/2016WR020197

Thaler, K.M., Berger, C., Leix, C., Drewes, J., Niessner, R., Haisch, C., 2017. Photoacoustic Spectroscopy for the Quantification of N2O in the Off-Gas of Wastewater Treatment Plants. Anal. Chem. 89, 3795–3801. https://doi.org/10.1021/acs.analchem.7b00491

Tippett, J., 2005. The value of combining a systems view of sustainability with a participatory protocol for ecologically informed design in river basins. Environ. Model. Softw., Policies and Tools for Sustainable Water Management in the European Union 20, 119–139. https://doi.org/10.1016/j.envsoft.2003.12.016

Tomperi, J., Koivuranta, E., Kuokkanen, A., Leiviskä, K., 2017. Modelling effluent quality based on a real-time optical monitoring of the wastewater treatment process. Environ. Technol. 38, 1–13.

Torres, M.E., Colominas, M.A., Schlotthauer, G., Flandrin, P., 2011. A complete ensemble empirical mode decomposition with adaptive noise, in: 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Presented at the 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 4144–4147. https://doi.org/10.1109/ICASSP.2011.5947265

Townsend-Small, A., Pataki, D.E., Tseng, L.Y., Tsai, C.-Y., Rosso, D., 2011. Nitrous Oxide Emissions from Wastewater Treatment and Water Reclamation Plants in Southern California. J. Environ. Qual. 40, 1542–1550. https://doi.org/10.2134/jeq2011.0059

Toyoda, S., Suzuki, Y., Hattori, S., Yamada, K., Fujii, A., Yoshida, N., Kouno, R., Murayama, K., Shiomi, H., 2011. Isotopomer Analysis of Production and Consumption Mechanisms of N2O and CH4 in an Advanced Wastewater Treatment System. Environ. Sci. Technol. 45, 917–922. https://doi.org/10.1021/es102985u

Tumendelger, A., Toyoda, S., Yoshida, N., 2014. Isotopic analysis of N2O produced in a conventional wastewater treatment system operated under different aeration conditions. Rapid Commun. Mass Spectrom. 28, 1883–1892. https://doi.org/10.1002/rcm.6973

US EPA, O., 2016. U.S. Greenhouse Gas Inventory Report: 1990-2014 [WWW Document]. URL https://www.epa.gov/ghgemissions/us-greenhouse-gas-inventory-report-1990-2014 (accessed 1.25.17).

Vachon, D., Prairie, Y.T., Cole, J.J., 2010. The relationship between near-surface turbulence and gas transfer velocity in freshwater systems and its implications for floating chamber measurements of gas exchange. Limnol. Oceanogr. 55, 1723–1732. https://doi.org/10.4319/lo.2010.55.4.1723

Van Hulle, S.W.H., Vandeweyer, H.J.P., Meesschaert, B.D., Vanrolleghem, P.A., Dejans, P., Dumoulin, A., 2010. Engineering aspects and practical application of autotrophic nitrogen removal from nitrogen rich streams. Chem. Eng. J. 162, 1–20. https://doi.org/10.1016/j.cej.2010.05.037

van Loosdrecht, M.C., Nielsen, P.H., Lopez-Vazquez, C.M., Brdjanovic, D., 2016. Experimental Methods in Wastewater Treatment. IWA Publishing.

van Loosdrecht, M.C.M., Salem, S., 2006. Biological treatment of sludge digester liquids. Water Sci. Technol. J. Int. Assoc. Water Pollut. Res. 53, 11–20.

Vanrolleghem, P.A., Lee, D.S., 2003. On-line monitoring equipment for wastewater treatment processes: state of the art. Water Sci. Technol. J. Int. Assoc. Water Pollut. Res. 47, 1–34.

Vasilaki, V., Massara, T.M., Stanchev, P., Fatone, F., Katsou, E., 2019. A decade of nitrous oxide (N2O) monitoring in full-scale wastewater treatment processes: A critical review. Water Res. 161, 392–412. https://doi.org/10.1016/j.watres.2019.04.022

Vasilaki, V., Volcke, E.I.P., Nandi, A.K., van Loosdrecht, M.C.M., Katsou, E., 2018. Relating N2O emissions during biological nitrogen removal with operating conditions using multivariate statistical techniques. Water Res. 140, 387–402. https://doi.org/10.1016/j.watres.2018.04.052

Wan, X., Baeten, J.E., Volcke, E.I.P., 2019. Effect of operating conditions on N2O emissions from one-stage partial nitritation-anammox reactors. Biochem. Eng. J. 143, 24–33. https://doi.org/10.1016/j.bej.2018.12.004

Wang, H., Ren, Z.J., 2014. Bioelectrochemical metal recovery from wastewater: a review. Water Res. 66, 219–232.

Wang, J., Zhang, J., Wang, Jing, Qi, P., Ren, Y., Hu, Z., 2011. Nitrous oxide emissions from a typical northern Chinese municipal wastewater treatment plant. Desalination Water Treat. 32, 145–152. https://doi.org/10.5004/dwt.2011.2691

Wang, X., Liu, J., Ren, N.-Q., Yu, H.-Q., Lee, D.-J., Guo, X., 2012. Assessment of Multiple Sustainability Demands for Wastewater Treatment Alternatives: A Refined Evaluation Scheme and Case Study. Environ. Sci. Technol. 46, 5542–5549. https://doi.org/10.1021/es300761x

Wang, X.-H., Wang, X., Huppes, G., Heijungs, R., Ren, N.-Q., 2015. Environmental implications of increasingly stringent sewage discharge standards in municipal wastewater treatment plants: case study of a cool area of China. J. Clean. Prod. 94, 278–283. https://doi.org/10.1016/j.jclepro.2015.02.007

Wang, Y., Fang, H., Zhou, D., Han, H., Chen, J., 2016a. Characterization of nitrous oxide and nitric oxide emissions from a full-scale biological aerated filter for secondary nitrification. Chem. Eng. J. 299, 304–313. https://doi.org/10.1016/j.cej.2016.04.050

Wang, Y., Geng, J., Ren, Z., He, W., Xing, M., Wu, M., Chen, S., 2011. Effect of anaerobic reaction time on denitrifying phosphorus removal and N2O production. Bioresour. Technol. 102, 5674–5684. https://doi.org/10.1016/j.biortech.2011.02.080

Wang, Y., Lin, X., Zhou, D., Ye, L., Han, H., Song, C., 2016b. Nitric oxide and nitrous oxide emissions from a full-scale activated sludge anaerobic/anoxic/oxic process. Chem. Eng. J. 289, 330–340. https://doi.org/10.1016/j.cej.2015.12.074

Wang, Yinghong, Wang, Yuesi, Ling, H., 2010. A new carrier gas type for accurate measurement of N2O by GC-ECD. Adv. Atmospheric Sci. 27, 1322–1330.

Wang, Z., Meng, Y., Fan, T., Du, Y., Tang, J., Fan, S., 2015. Phosphorus removal and N2O production in anaerobic/anoxic denitrifying phosphorus removal process: Long-term impact of influent phosphorus concentration. Bioresour. Technol. 179, 585–594. https://doi.org/10.1016/j.biortech.2014.12.016

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

Weissenbacher, N., Takacs, I., Murthy, S., Fuerhacker, M., Wett, B., 2010. Gaseous Nitrogen and Carbon Emissions from a Full-Scale Deammonification Plant. Water Environ. Res. 82, 169–175. https://doi.org/10.2175/106143009X447867

Wheeler, D.C., Nolan, B.T., Flory, A.R., DellaValle, C.T., Ward, M.H., 2015. Modeling groundwater nitrate concentrations in private wells in Iowa. Sci. Total Environ. 536, 481–488. https://doi.org/10.1016/j.scitotenv.2015.07.080

White, C.J., Lehnert, N., 2016. Is there a pathway for N2O production from hydroxylamine oxidoreductase in ammonia-oxidizing bacteria? Proc. Natl. Acad. Sci. 113, 14474–14476. https://doi.org/10.1073/pnas.1617953114

Wold, S., Kettaneh, N., Fridén, H., Holmberg, A., 1998. Modelling and diagnostics of batch processes and analogous kinetic experiments. Chemom. Intell. Lab. Syst. 44, 331–340. https://doi.org/10.1016/S0169-7439(98)00162-2

Wu, Z., Huang, N.E., 2009. Ensemble Empirical Mode Decomposition: a Noise-Assisted Data Analysis Method. Adv. Adapt. Data Anal. 1, 1–41. https://doi.org/10.1142/S1793536909000047

Wu, Z., Huang, N.E., Chen, X., 2009. The multi-dimensional ensemble empirical mode decomposition method. Adv. Adapt. Data Anal. 01, 339–372. https://doi.org/10.1142/S1793536909000187

Wu, Z., Huang, N.E., Long, S.R., Peng, C.-K., 2007. On the trend, detrending, and variability of nonlinear and nonstationary time series. Proc. Natl. Acad. Sci. 104, 14889–14894. https://doi.org/10.1073/pnas.0701020104

Wunderlin, P., Mohn, J., Joss, A., Emmenegger, L., Siegrist, H., 2012. Mechanisms of N2O 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

Xie, W.-M., Zhang, R., Li, W.-W., Ni, B.-J., Fang, F., Sheng, G.-P., Yu, H.-Q., Song, J., Le, D.-Z., Bi, X.-J., Liu, C.-Q., Yang, M., 2011. Simulation and optimization of a full-scale Carrousel oxidation ditch plant for municipal wastewater treatment. Biochem. Eng. J. 56, 9–16. https://doi.org/10.1016/j.bej.2011.04.010

Yan, X., Li, L., Liu, J., 2014. Characteristics of greenhouse gas emission in three full-scale wastewater treatment processes. J. Environ. Sci. 26, 256–263. https://doi.org/10.1016/S1001-0742(13)60429-5

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

Yoshida, H., Mønster, J., Scheutz, C., 2014. Plant-integrated measurement of greenhouse gas emissions from a municipal wastewater treatment plant. Water Res. 61, 108–118. https://doi.org/10.1016/j.watres.2014.05.014

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

Zhang, R., Xie, W.-M., Yu, H.-Q., Li, W.-W., 2014. Optimizing municipal wastewater treatment plants using an improved multi-objective optimization method. Bioresour. Technol. 157, 161–165. https://doi.org/10.1016/j.biortech.2014.01.103

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, X., Zhang, X., Zhang, Z., Liu, Y., 2018. Full nitration-denitration versus partial nitration-denitration-anammox for treating high-strength ammonium-rich organic wastewater. Bioresour. Technol. 261, 379–384. https://doi.org/10.1016/j.biortech.2018.04.049

Zhou, Y., Pijuan, M., Zeng, R.J., Yuan, Z., 2008. Free Nitrous Acid Inhibition on Nitrous Oxide Reduction by a Denitrifying-Enhanced Biological Phosphorus Removal Sludge. Environ. Sci. Technol. 42, 8260–8265. https://doi.org/10.1021/es800650j

Zhu, X., Chen, Y., 2011. Reduction of N2O and NO Generation in Anaerobic−Aerobic (Low Dissolved Oxygen) Biological Wastewater Treatment Process by Using Sludge Alkaline Fermentation Liquid. Environ. Sci. Technol. 45, 2137–2143. https://doi.org/10.1021/es102900h

Zhu-Barker, X., Cavazos, A.R., Ostrom, N.E., Horwath, W.R., Glass, J.B., 2015. The importance of abiotic reactions for nitrous oxide production. Biogeochemistry 126, 251–267. https://doi.org/10.1007/s10533-015-0166-4

List of publications

The thesis is based on the following publications and conference presentations:

Publications

- Vasilaki, V., Massara, T.M., Stanchev, P., Fatone, F. and Katsou, E., (2019). A decade of nitrous oxide (N2O) monitoring in full-scale wastewater treatment processes: A critical review. Water research, 161, p.392.
- Vasilaki, V., E. I. P. Volcke, A. K. Nandi, M. C. M. van Loosdrecht, and E. Katsou.. (2018). Relating N2O emissions during biological nitrogen removal with operating conditions using multivariate statistical techniques. Water research, 140, pp.387-402.
- Spinelli, M., Eusebi, A.L., Vasilaki, V., Katsou, E., Frison, N., Cingolani, D. and Fatone, F., (2018). Critical analyses of nitrous oxide emissions in a full scale activated sludge system treating low carbon-to-nitrogen ratio wastewater. Journal of cleaner production, 190, pp.517-524.
- Danishvar, M., Vasilaki, V., Huang, Z., Katsou, E. and Mousavi, A., 2018, July. Application of Data Driven techniques to Predict N 2 O Emission in Full-scale WWTPs. In 2018 IEEE 16th International Conference on Industrial Informatics (INDIN) (pp. 993-997). IEEE.
- Vasilaki, V., Danishvar, M., Huang, Z., Mousavi, A., & Katsou, E. (2017). Application of Event-Based Real-Time Analysis for Long-Term N 2 O Monitoring in Full-Scale WWTPs. In Frontiers International Conference on Wastewater Treatment and Modelling (pp. 436-443). Springer, Cham.
- Vasilaki, V., Conca, V., Frison, N., Eusebi, A.L., Fatone, F., Katsou, E., 2020. A knowledge discovery framework to predict the N2O emissions in the wastewater sector. Water Research 178, 115799.
- Vasilaki, V., Danishvar, S., Mousavi, A., Katsou, E., 2020. Data-driven versus conventional N2O EF quantification methods in wastewater; how can we quantify reliable annual EFs? Computers & Chemical Engineering 141, 106997.

Conference presentations

- Vasilaki, V.; Conca, V.; Frison, N.; Fatone, F.; Katsou, E. 2019. Trade-offs between environmental and operational parameters in SCENA process, 3rd IWA Resource Recovery Conference, Venice, Italy, 8-12 September 2019.

- Vasilaki, V.; Conca, V.; Frison, N.; Fatone, F.; Katsou, E. 2019, June. Multivariate Analysis for Behavioural System Analysis in The SCENA Process. Poster pitch presentation at the 16th IWA Leading Edge Conference on Water and Wastewater Technologies (LET2019), Edinburgh, UK, 10-14 Jun2 2019
- Vasilaki V., Mattias T., Angadi V.C., Sousa P., Mousavi A., Katsou E. 2018 Prediction of wastewater N2O emissions using artifcial Neural Networks, 6th International Conference on Sustainable Solid Waste Management, Naxos Island, Greece, 13-16 June 2018
- Vasilaki V., Angadi V.C., Frison N., Mousavi A., Fatone F., Katsou E.. 2018. Assessment of wastewater N2O generation using multivariate techniques, at the 6th International Conference on Sustainable Solid Waste Management, Naxos Island, Greece, 13-16 June 18
- Vasilaki V., Volcke E.I.P., Nandi A.K., van Loosdrecht M.C.M., Katsou E.. 2018 Integrating sustainability indicators in the monitoring and control strategy of WWTPs - Moving towards a smart digital water society, 12th European Waste Water Management Conference, Manchester, UK, 17-18 July 2018
- Vasilaki V., Danishvar M., Huang Z., Mousavi A., Katsou E. 2017. Application of Event-based Real-Time Analysis for Long-term N2O Monitoring in Full-scale WWTPs. FICWTM2017, Palermo, Italy, 21-24 May 2017

Appendix A

Table A.1: Studies considered in Fig. 2.1-2.6. All studies considered in Fig. 2.1 can be found in Tables A.3 and A.4.

Study	Figure
Castro-Barros, C.M., Daelman, M.R.J., Mampaey, K.E., van Loosdrecht, M.C.M., Volcke, E.I.P., 2015. Effect of aeration regime on N_2O emission from partial nitritation-anammox in a full-scale granular sludge reactor. Water Res. 68, 793–803. https://doi.org/10.1016/j.watres.2014.10.056	2.1, 2.2, 2.3
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, 2009. 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	2.1, 2.2, 2.3
Kampschreur, M.J., van der Star, W.R.L., Wielders, H.A., Mulder, J.W., Jetten, M.S.M., van Loosdrecht, M.C.M., 2008. Dynamics of nitric oxide and nitrous oxide emission during full-scale reject water treatment. Water Res. 42, 812–826. https://doi.org/10.1016/j.watres.2007.08.022	2.1, 2.2, 2.3
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	2.1, 2.2, 2.3
Stenström, F., Tjus, K., Jansen, J. la C., 2014. Oxygen-induced dynamics of nitrous oxide in water and off-gas during the treatment of digester supernatant. Water Sci. Technol. 69, 84–91. https://doi.org/10.2166/wst.2013.558	2.1, 2.2, 2.3
Weissenbacher, N., Takacs, I., Murthy, S., Fuerhacker, M., Wett, B., 2010. Gaseous Nitrogen and Carbon Emissions from a Full-Scale Deammonification Plant. Water Environ. Res. 82, 169–175. https://doi.org/10.2175/106143009X447867	2.1, 2.2, 2.3
Ahn, J.H., Kim, S., Park, H., Rahm, B., Pagilla, K., Chandran, K., 2010. N_2O Emissions from Activated Sludge Processes, 2008−2009: Results of a National Monitoring Survey in the United States. Environ. Sci. Technol. 44, 4505–4511. https://doi.org/10.1021/es903845y	2.1, 2.2, 2.3, 2.4, 2.5, 2.6
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.	2.1, 2.2, 2.3, 2.4, 2.5, 2.6

Environ. Sci. Pollut. Res. 22, 11840–11849. https://doi.org/10.1007/s11356-015-4467-x

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 2.1, 2.2, 2.3, 2.4, 2.5, 2.6

Mello, W.Z. de, Ribeiro, R.P., Brotto, A.C., Kligerman, D.C., Piccoli, A. de S., Oliveira, J.L.M., 2013. Nitrous oxide emissions from an intermittent aeration activated sludge system of an urban wastewater treatment plant. Quím. Nova 36, 16–20. https://doi.org/10.1590/S0100-40422013000100004 2.1, 2.2, 2.3, 2.4, 2.6

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 2.1, 2.2, 2.3, 2.4, 2.5, 2.6

Ren, Y. g., Wang, J. h., Li, H. f., Zhang, J., Qi, P. y., Hu, Z., 2013. Nitrous oxide and methane emissions from different treatment processes in full-scale municipal wastewater treatment plants. Environ. Technol. 34, 2917–2927. https://doi.org/10.1080/09593330.2012.696717 2.1, 2.2, 2.3, 2.4, 2.5, 2.6

Samuelsson, J., Delre, A., Tumlin, S., Hadi, S., Offerle, B., Scheutz, C., 2018. Optical technologies applied alongside on-site and remote approaches for climate gas emission quantification at a wastewater treatment plant. Water Res. 131, 299–309. https://doi.org/10.1016/j.watres.2017.12.018 2.1, 2.2, 2.3, 2.4, 2.6

Spinelli, M., Eusebi, A.L., Vasilaki, V., Katsou, E., Frison, N., Cingolani, D., Fatone, F., 2018. Critical analyses of nitrous oxide emissions in a full scale activated sludge system treating low carbon-to-nitrogen ratio wastewater. J. Clean. Prod. 190, 517–524. https://doi.org/10.1016/j.jclepro.2018.04.178 2.1, 2.2, 2.3, 2.4, 2.5, 2.6

Sun, S., Bao, Z., Li, R., Sun, D., Geng, H., Huang, X., Lin, J., Zhang, P., Ma, R., Fang, L., Zhang, X., Zhao, X., 2017. Reduction and prediction of N_2O emission from an Anoxic/Oxic wastewater treatment plant upon DO control and model simulation. Bioresour. Technol. 244, 800–809. https://doi.org/10.1016/j.biortech.2017.08.054 2.1, 2.2, 2.3, 2.4, 2.5, 2.6

Wang, Y., Fang, H., Zhou, D., Han, H., Chen, J., 2016a. Characterization of nitrous oxide and nitric oxide emissions from a full-scale biological aerated filter for secondary nitrification. Chem. Eng. J. 299, 304–313. https://doi.org/10.1016/j.cej.2016.04.050 2.1, 2.2, 2.3, 2.4, 2.5, 2.6

Wang, Y., Lin, X., Zhou, D., Ye, L., Han, H., Song, C., 2016b. Nitric oxide and nitrous oxide emissions from a full-scale activated sludge anaerobic/anoxic/oxic process. Chem. Eng. J. 289, 330–340. https://doi.org/10.1016/j.cej.2015.12.074 2.1, 2.2, 2.3, 2.4, 2.5, 2.6

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 2.1, 2.2, 2.3, 2.4, 2.5, 2.6

Filali, A., Fayolle, Y., Peu, P., Philippe, L., Nauleau, F., Gillot, S., 2013. Aeration control in a full-scale activated sludge wastewater treatment plant: impact on performances, energy consumption and N_2O emission. Presented at the 11ème Conférence IWA sur l'instrumentation, le contrôle et l'automatisation. ICA2013, p. 4 p. 2.1, 2.2, 2.3, 2.4

Rodriguez-Caballero, A., Aymerich, I., Marques, R., Poch, M., Pijuan, M., 2015. Minimizing N_2O emissions and carbon footprint on a full-scale activated sludge sequencing batch reactor. Water Res. 71, 1–10. https://doi.org/10.1016/j.watres.2014.12.032 2.1, 2.2, 2.3

Sun, S., Bao, Z., Sun, D., 2015. Study on emission characteristics and reduction strategy of nitrous oxide during wastewater treatment by different processes. Environ. Sci. Pollut. Res. 22, 4222–4229. https://doi.org/10.1007/s11356-014-3654-5 2.1, 2.2, 2.3, 2.4, 2.5

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 2.1, 2.2, 2.3,

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 2.1, 2.2, 2.3, 2.4, 2.5

Baresel, C., Andersson, S., Yang, J., Andersen, M.H., 2016. Comparison of nitrous oxide (N_2O) emissions calculations at a Swedish wastewater treatment plant based on water concentrations versus off-gas concentrations. Advances in Climate Change Research, Including special topic on atmospheric black carbon and its effects on cryosphere 7, 185–191. https://doi.org/10.1016/j.accre.2016.09.001 2.1, 2.2, 2.3, 2.4, 2.5

Foley, J., de Haas, D., Yuan, Z., Lant, P., 2010. Nitrous oxide generation in full-scale biological nutrient removal wastewater treatment plants. Water Research 44, 831–844. https://doi.org/10.1016/j.watres.2009.10.033 2.1, 2.2, 2.3, 2.44

&

Foley, J., Yuan, Z., Keller, J., Senante, E., Chandran, K., Willis, J., Shah, A., Loosdrecht, M.C.M. van, Voorthuizen, E. van, 2015. N_2O and CH4 Emission from Wastewater Collection and Treatment Systems: State of the Science Report and Technical Report. IWA Publishing.

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. Science of The Total Environment 493, 384–391. https://doi.org/10.1016/j.scitotenv.2014.06.015 2.1, 2.2, 2.3, 2.4, 2.5, 2.6

Yan, X., Li, L., Liu, J., 2014. Characteristics of greenhouse gas emission in three full-scale wastewater treatment processes. J. Environ. Sci. 26, 256–263. https://doi.org/10.1016/S1001-0742(13)60429-5 2.1, 2.2, 2.3, 2.4, 2.5

Table A.2: Conversion of N_2O EF to % of TN-load

Paper	Process	TN (mg/l) influent	TN removal efficiency (%)	EF (mgN_2O/L wastewater)	EF TN removed (%)	EF TN Effluent (%)	EF TN influent (%)
Yan et al., 2014	A^2/O	69.55	85	0.06	0.08		0.068
Ren et al., 2013	A^2/O		81		0.13		0.104
Weissenbacher et al., 2010	Anammox		87			1.3	0.169
Ren et al., 2013	OD		70		0.14		0.098
Yan et al., 2014	OD	73.40	53	0.2	0.36		0.191
Ren et al., 2013	r- A^2/O		80		0.11		0.091
Yan et al., 2014	r- A^2/O	69.55	48	0.07	0.23		0.110

Table A.3: Performance, operational & monitoring strategy and monitoring period of sidestream configurations with respect to the N_2O EF. Studies with (*) have been considered in the development of Fig 2.2 and 2.3 with EFs under normal operating conditions. All studies have been considered in Fig. 2.1. Additionally, details are provided in the supplementary excel file

Configuration		N_2O **EF**	**Performance & Operational Strategy**				**Monitoring period & strategy**		
Source	**Process**	**EF (with respect to TN inf.)**	**N-load (kg d^{-1})**	**NH_4^+ eff. (mg L^{-1})**	**NO_2^- eff. (mg L^{-1})**	**Operational strategy**	**Period**	**C/GS**	**Monitoring strategy**
Castro-Barros et al., 2015*	One-stage PNA granular[a]	2%	1,053[g]	24.5-16.6	4.7-6.6	• Continuous intermittent aeration • Fresh air controlled by NH_4^+ & NO_2^- concentration	3.08[b]	C	• Summer campaign • Off-gas equal to blower flow-rate • Covered
Mampaey et al., 2016*	One-stage SHARON granular reactor	3.8%	894	nm	nm	• Aerobic retention time=1.35 d • 2-h cycles intermittent aeration; aerated period (DO setpoint=2 mg L^{-1}); non-aerated period based on varying influent flow-rates	9.5[d]	C	• Summer campaign • Off-gas equal to blower flow-rate • Covered
		1.8%	1,193	nm	nm	• Short cycles (1 h) • NO_3^- concentration increased from 8 to 15 g N m^{-3}	1[e]	C	
		1.5%	965	nm	nm	• Continuous aerobic conditions (DO setpoint=2 mg L^{-1})	1[f]	C	
		3.9%	914	nm	nm	• DO concentration=1 g O_2 m^{-3}	<1[h]	C	
		18.5%	323	nm	nm	• DO concentration=0.6 g O_2 m^{-3}	<1[h]	C	
Kampschreur et al., 2009 *		1.23%	1,200[g]	1-2	6-8	• DO setpoint=5 mg L^{-1}	20 h	GS	

	One-stage PNA granular [a]	~1.23%[j]	nm	nm	nm	• A factor 1.5 aeration rate increase (from 2,000 to 3,000 N m^{-3} h^{-1}) leading directly to a factor 1.5 increase in NO emission	<1	GS	• Continuous off-gas collection (at 1 L min^{-1}) during the measurement campaign • Off-gas equal to blower flow-rate
		~0	nm	nm	nm	• DO<0.5 mg L^{-1}	<1	GS	
		>1.23%[k]	nm	nm	nm	• NO_2^- accumulation up to 23 mg NO_2-N L^{-1}	<1	GS	
Ahn et al., 2010*	Nitritation reactor	0.24-0.54[l]	8,650[m]	710[n]	nm	• No details provided	1	C	• SEIFC • Flow-rate assessed 5 times d^{-1}
Kampschreur et al., 2008*	Two-reactor nitritation-anammox process (anammox reactor)	0.6	980[o]	78.1-115	4.9-7.2	• Off-gas recycled to the bottom of the reactor • Designed for a conversion of 500 kg N d^{-1}	3.13	GS	• Off-gas equal to blower flow-rate • Covered
	Two-reactor nitritation-anammox process (nitritation SHARON reactor)	1.7				• DO=2.5 mg L^{-1} • Discontinuous aeration to maintain (solid) aerobic retention time=1.4 d • Aeration cycles of 2 h	3.13	GS	
Weissenbacher et al., 2010*	Single deammonification reactor	0.17	2 89	136	1.75	• pH-controlled aerobic/anoxic sequence of 6 h • DO=0.3 mg L^{-1} • Continuous influent flow-rate=2.5 L s^{-1} • Settling & discharge phase of 2 h • HRT=3.3 d	ns	GS	• Cylinder used for gas collection • 15-min GS • Off-gas equal to blower flow-rate

Joss et al., 2009	Nitritation-anammox (continuous aeration)	0.4 %[p]	625	30	<0.2	• DO<1 mg L^{-1} • The simultaneous anammox activity during aeration already depleting most of the NO_2^- • Better monitoring of the process for avoidance of NO_2^- accumulation	ns	C	• Off-gas equal to blower flow-rate • Covered
	Nitritation-anammox (intermittent aeration)	0.6%[p]	625	30	<0.2	• 45 min aeration & 15 min stirring • Both control strategies performing comparably in terms of NO_2^- oxidizing activity	ns	C	
Gustavsson and la Cour Jansen, 2011	Nitritation SBR	3.8%[g]	4,000	52.9	600-850	• 6-h SBR cycle • Simultaneous filling (174 m^3 h^{-1}), a settling & a decantation phase • pH controlled to 6.8 by addition of NaOH • DO setpoint=1.3 mg L^{-1} • Aerobic HRT=1.4 d • Total HRT=2.3 d	5[c]	C	• Off-gas collected by a long vertical pipe the end of which placed just below the water surface at the lowest water level • Non-covered
Stenström et al., 2014*	Nitrification-denitrification SBR	5.1%	84[q]	30-90	10–25	• 8-h SBR cycle • Anoxic feeding with ethanol dosage/oxic/settling phase • HRT= 5.0 ± 0.5 d. • SRT=15 ± 3 d • DO=2 mgL^{-1}	2	C	• Flow-rate assessed by mass flow-meter and rotameter

[a.] Effluent form potato-processing UASB plant & AD reject water from municipal WWTP
[b.] Normal operation, summer period; the total monitoring campaign lasted 7 days
[c.] Winter campaign: 19th to 20th January (4 cycles) & 14th to 16th February (8 cycles)
[d.] 114 cycles considered as standard operation, accounting for 228 h & a total of 21 d
[e.] 21 cycles
[f.] 2.8 h of continuous aerobic conditions (DO setpoint=2 mg L^{-1}), equivalent to the duration of 11 cycles
[g.] NH_4-N load
[h.] Result of 2 cycles
[i.] Conversion capacity

[j] No significant change in the N_2O emissions probably caused by a slower response of the N_2O off-gas concentration due to mass transfer limitation combined with higher N_2O solubility

[k] NO_2^- accumulation when high oxygen concentration applied after oxygen limitation; the latter leading to higher nitrification rates compared to the anammox conversion rates

[l] Influent TKN

[m] TKN

[n] TN load (g N m^3)

[o] Kj-N

[p] N_2O EF with respect to the N-load removed

[q] Per cycle

Table A.4: Performance, operational & monitoring strategy and monitoring period of mainstream configurations with respect to the N_2O EF. (P: monitoring period, C: Continuous, GS: Grab-samples, nm: not mentioned/specified). Studies with (*) have been considered in the development of Fig 2.2-2.5 and with average EFs under normal operating conditions. Specific details on the studies considered in each Fig. All studies have been considered in Fig. 2.1. Additionally, details are provided in the supplementary excel file

Source	Process	Control	P.E. (10^3)	Q ($ML\ d^{-1}$)	Inf. N-content ($mg\ L^{-1}$)	Eff. NH_4^+ ($mg\ L^{-1}$)	Eff. NO_3^- ($mg\ L^{-1}$)	P (d)	Monitoring	EF with respect to inf. TN	C/ GS
Aboobakar et al., 2013*	Plug-flow	• Small anoxic zone followed by aerated zones (3 passes) • DO control=1.5 mg L^{-1} • SRT=10 d • HRT=8 h • T=16-19 ℃	210	4.4	15.3	0.25	12	56	• 7 days per pass • 8 equally distributed sampling points • Floating chamber • Off-gas equal to blower flow-rate	0.04%	C
Pan et al., 2016*	Plug-flow step-feed	• SRT=12 d (aerobic = 8 d) • HRT=12 h • 2 steps (anoxic/aerobic) • Influent in both anoxic zones • RAS sent to path 1 • DO control=1 mg L^{-1}	nm	50	47.4	0.3	12.1	49	• N_2O monitored: beginning, middle, end of oxic zones • Floating chamber: modified from plastic commercial hopper tank • Oxygen meter to calculate Qgas	1.9%[a]; 0.7%[a]; 3.5%[a]	C
Rodriguez-Caballero et al., 2014*	Plug-flow	• Anoxic/aerobic/short anoxic/aerobic • DO oxic 1=2.1 mg L^{-1} [c] • DO oxic 2=1.4 mg L^{-1} [c] • DO oxic 3=0.7 mg L^{-1} [c] • HRT≈0.6 d • SRT≈10 d • T=25 ℃	112	21.8	47.9	5.1[b]	nm	70[d]	• 4 days per sampling location (SL) • SLs separated based on different air diffusers & flow turbulence • Off-gas equal to blower flow-rate	0.12%[b]	C

									• Commercial flux chamber (AC'SCENT® Flux Hood)		
Ahn et al., 2010*	Plug-flow	• 4 passes • DO=3.1 mg L^{-1} [e] • T=11 ℃	nm	15	22[b]	12.2[b]	nm	1	• SEIFC • Air flow-rate: 5 times per day	0.4%[b]	C
		• 4 passes • DO = 0.9 mg L^{-1} [e] • T=23 ℃			26.4 [b]	10.7[b]	nm	1		0.4%[b]	
Ahn et al., 2010*	Plug-flow	• 2 passes • DO=1.7 mg L^{-1} [e] • T=11 ℃	nm	8.7	15.9 [b]	10.1[b]	nm	1	• SEIFC • Air flow-rate: 5 times per day	0.6%[b]	C
		• 2 passes • DO=0.9 mg L^{-1} [e] • T=22 ℃			22[b]	13.3[b]	nm	1		0.1%[b]	
Wang et al., 2016b*	A^2/O	• Plug-flow pattern • HRT=7.7-10.3 h • T=13-34 ℃ • Diurnal DO=0.6-6.8 mg L^{-1} • No automatic DO control	200	48	10-30	2-10	nm	365	• Monitoring frequency: once per month • Investigation of seasonal effects • Location: oxic starting point • Chamber: closed floating equipped with liquid level meter, temperature meter, pressure bag & fan	0.1-3.4%	C
Ren et al., 2013*	A-A^2/O	• T=12.5-23.5 ℃	260	100	73.9	2.1	0.4	61	• SEIFC & gas bag • Off-gas equal to blower flow-rate • N_2O calculated based on gas bag concentration change • Gas samples collected every 10 min for a total of 30 min • March to June	0.104%[f]	GS

Yan et al., 2014*	A^2/O	• HRT=15 h • SRT = 14 d • Temperature/DO not specified	nm	230	38.8 - 91.3	ns[j]	ns[j]	9	• Gas bag • 9-month (from March to November) once per month • Investigation of seasonal effects • Various sampling locations[i]	0.1%[k]	GS
Hwang et al., 2016	A^2/O	• T=15-35 ℃	3,220	1,710	12–1,550[l]	nm	nm	10	• Modified forced-draught chamber & gas bag • Gas flow-rate measured	0.16%[m]	GS
Foley et al., 2010	A^2/O	• Four-stage bioreactor (similar to A^2/O configuration) with diffused aeration & supplemental COD dosing by primary sludge	nm	25	55-85[p]	2.9-3.1[p]	nm	2	• Floating gas hood & gas bags • 2-4 h of intensive sampling	1.4%[y]	GS
Wang et al., 2011	A^2/O + WWTP	• High density settler tanks and high efficiency fiber filter beds after final settling tanks • T=12-24 ℃	1,500	300	37.5	< 10[p]	nm	90	• Sampling 3 times per week 9:00 am – 05:00 pm • Emission isolation gas hood and gas-bag • Sampling positions determined by DO change and water surface area	0.1-0.13%[v]	GS
Toyoda et al., 2011	A^2O	• HRT = 21.6 • T = 19.7	160	75	nm	nm	nm	1	• Gas samples were collected at exhaust duct over the oxic tank via stainless steel and glass bottles • Concentration and isotopomer ratios of N_2O in the water and gas samples measured on an isotope-ratio monitoring mass spectrometer	nm	-
Li et al., 2016*	Reversed A^2/O	• Fine bubble aeration	nm	250	34	8	nm	nm	• Sampling bag collected every 2-3 d	6-10%[b]	GS

		• Temperature/DO not specified							• Pre-experiment to select sampling points representative to reactor zones (3 oxic zones, 1 anoxic & 1 anaerobic) • Hand-crafted SEIFC • Summer (May-August) and winter (November-January) • Investigation of seasonal effects		
Ren et al., 2013*	Reversed A^2/O	• T=12.5-23.5 °C	160	50	104.1	2.4	4.3	61	• SEIFC & gas bag • Off-gas equal to blower flow-rate • N_2O calculated based on gas bag concentration change • Gas samples collected every 10 min for a total of 30 min • March to June (frequency not clear)	0.091%[h]	GS
Yan et al., 2014*	Reversed A^2/O	• HRT=15 h • SRT=8 d • Temperature/DO not specified	nm	230	38.8 - 91.3	nm[j]	nm[j]	90	• Gas bag • 9-month (from March to November); once per month • Investigation of seasonal effects • Various sampling locations[i]	0.11%[n]	GS
Daelman et al., 2015*	Plug-flow & 2 carrousels	• DO, NO_3^-, NH_4^+ & TSS data available from SCADA system at 10-min intervals • Average influent COD=328 mg L^{-1} (87% average removal efficiency) • MLSS controlled by operators depending on	360	Infl. flow-rate available from SCADA	42	8	nm	487	• Covered system • Off-gas continuously pumped to a Servomex infrared gas analyzer to measure N_2O • 16-month monitoring period (October 2010-January 2012; 1-month interruption	2.8%	C

		temperature; SRT resulting from the amount of wasted sludge to maintain the target MLSS		syste m at 10-min interv als					in October 2011 due to technical failure)		
Brotto et al., 2015*	Non-BNR AS	• Average HRT=10 h (range:7-13 h) • SRT=3 d • Continuous aeration system with no intentional anoxic zones • DO ranging from <1 to >7 mg L^{-1} • T=26-31 ℃	50	14.7	31	16.2	0.9	183	• Sampling from 7 distinct points • January to July (once per month) • Gas sampling: upturned funnel device • Investigation of seasonal effects • Off-gas equal to blower flow-rate	0.02%-0.3%	GS
Ribeiro et al., 2017	Extended aeration non-BNR AS	• Located at research institution • Raw wastewater flow linked to institution's work regime (not continuous) • HRT=8 h • SRT=25 d	2.5	1.1	nm	nm	nm	3	• Floating chamber	ns	C
Filali et al., 2013*	AS	• Annular type reactor with a central anaerobic zone & an alternate aeration mode in the outer ring • Conventional aeration control system (O_2/ORP)	230	nm	nm	nm	nm	1	• Floating chamber • 2 sampling locations representative of the measured gas/liquid mass transfer parameters	0.004%[b]	C
Chen et al., 2016*	AS	• Influent wastewater mainly produced from various metal-processing, metal surface & chemical industries	nm	18	15.5 - 16.3	nm	nm	14	• 7 d, 4 times per d, twice (winter/summer) • Samples collected by a QT-2B air sampler (1m above reactor surface)	0.0872%[w]; 0.017[s]	GS

		• T=30.5 °C (winter) & 30.8 °C(summer) • HRT=5 h									
Bellandi et al., 2018	AS	• No pre-denitrification • Aerated tank equipped with EPDM membrane disk diffusers • 1st half of aerated tank: 56.6% of diffusers; 2nd half: 43.4% of diffusers • Fixed air flowrate adjusted once per day according to manual DO measurement & AS characteristics • No online monitoring/ logging	300	nm	50	nm	nm	1	• 3 floating hoods distributed along the length of 1 of aeration tanks	0.06[r]	C
Mello et al., 2013 *	AS	• Intermittent aeration (60 min aerated & 30 min non-aerated) • Air flow-rate=640 m^3 h^{-1}	2	0.6	27.5	13.4	0.9	6	• Duration monitoring: 90 min per d • Sampling method: upturned 30-cm diameter plastic funnel	0.1%	GS
Tumendelger et al., 2014*	AS	• DO controlled at 1.5, 2.0 and 2.5 mg L–1 • HRT = 11.1 – 11.5 h • SRT = 11.5–13.5 days • T = 17.1–18.3 °C	447	14	nm	1.3	59 - 76	3	• Aerobic tanks separated into three zones further partitioned into three sections • Sampling between 9:00 am – 05:00 pm • Gas samples collected from the exhaust duct • Dissolved N_2O measured via isotope ratio mass spectrometer	0.03-0.14%	-

Castellano-Hinojosa et al., 2018	Sequential AS	• pre-denitrification/nitrification • 4 full-scale municipal WWTPs	10-425	2-488	41.1 - 93.6	14-52.3	nm	10	• Biweekly 5-month campaign (December-April) • 50mL of freshly collected AS placed in 125-mL glass bottles: closed with serum rubber caps to allow injection & withdrawal of gas samples, & incubated at 25 °C during 3 h to achieve maximum N_2O emission • Varian 4900 Gas Chromatograph with a PoraPlot Qcolumn equipped with an electron capture detector with N_2 as carrier gas	Average N_2O emission in AS samples: varying from 0.1± 0.05 to 6.5± 8.9 mg N_2O h^{-1} L^{-1} AS	C
Ahn et al., 2010*	Step-feed non-BNR	• T=17 °C	nm	322.8	26.6 [b]	13.6[b]	nm	1	• Covered • Air flow-rate: 5 times per day	0.2[b]	C
Ahn et al., 2010*	Step-feed non-BNR	• T=26 °C	nm	422.8	21.1 [b]	9.9[b]	nm	1	• Covered • Air flow-rate: 5 times per day	1.8[b]	C
Ahn et al., 2010*	Modified Ludzack-Ettinger	• T=26 °C • DO=3.2 mg L^{-1} [e]	nm	18.2	37.4 [b]	10.4[b]	nm	1	• SEIFC • Air flow-rate: 5 times per day	0.07[b]	C
Ahn et al., 2010*	Modified Ludzack-Ettinger	• T=26 °C	nm	18.6[b]	37[b]	8	nm	1	• SEIFC • Air flow-rate: 5 times per day	0.06[b]	C
Spinelli et al., 2018*	Modified Ludzack-Ettinger	• Fine-bubble diffusers • Air supply: 1,870-9,210 m^3 h^{-1} • Blowers automatically controlled	80	30	25.1	0.2	12	46	• Aerobic basin: sampling point placed at the head of aerobic reactor for 46 d & at the end for 7 d • Fixed & floating chambers tested	0.001-0.005%	C

		• DO: 3 different operating settings: 0.4, 1 & 3 mg L^{-1} • DO=4 mg L^{-1} (measured) • SRT=10 d • T=17.7 °C							• Off-gas equal to blower flow-rate • Calibration tests to select sampling methodology (based on types/dimensions of applied chambers)		
Bellandi et al., 2018	Modified Ludzack-Ettinger	• Fine-bubble diffusers • Decreasing aerator density towards tank outlet: 44%, 30.5% & 25.5% of aerators in each section • NH_4^+-DO cascade control • Limited NH_4^+ concentration at plant entrance due to constant groundwater infiltration in the sewer	600	200	<5	nm	nm	1	• 5 floating hoods distributed along the length of 1 of the aeration tanks	0.01[r]	C
Townsend-Small et al., 2011	Modified Ludzack-Ettinger	• Anoxic zone representing≈ 33% of total reactor • Methanol addition • Mean cell retention time (MCRT)=8.5 d	5,700	35[s]	<10[s]	nm	nm	nm	• Air samples above the treatment tanks collected in pre-evacuated stainless-steel canisters using a vacuum pump through a magnesium perchlorate water trap (5 cm above the wastewater surface) • Samples collected at mid-day on weekdays (wastewater & N-loading average levels)	0.86[t]	GS
Caivano et al., 2017*	Modified Ludzack-Ettinger	• Submerged aerators coarse-bubble diffusers • Entire oxidation system αSOTE in oxidation tank=1.9% • DO=0.2 mg L^{-1} (measured)	15	3.8	45-55.6	34.5-43.5	0.003-0.4	3	• Custom-made floating chamber & super-inert multi foil bags • Monitoring of gas flow-rate • Sampling over a geometrically representative	0.03[b]	GS

									area of the whole basin on at least 2% of the total tank surface (15 different locations)		
Foley et al., 2010	Modified Ludzack-Ettinger	• Diffused aeration	nm	63	50[p]	11.8[p]	nm	1	• Floating gas hood & gas bags • 2-4 h of intensive sampling	2.7%[y]	GS
Foley et al., 2010*	Modified Ludzack-Ettinger	• Diffused aeration	nm	49	69-103 [p]	8-11.8[p]	nm	2	• Floating gas hood & gas bags • 2-4 h of intensive sampling	5.1%[y]	GS
Foley et al., 2010*	Modified Ludzack-Ettinger	• Diffused aeration	nm	20	66-85[p]	12.4-15.5[p]	nm	2	• Floating gas hood & gas bags • 2-4 h of intensive sampling	9%[y]	GS
Ahn et al., 2010*	Separate-stage BNR	• T=15 °C • DO=3.2 mg L^{-1}	nm	104.6	17.2 [b]	3.4[b]	nm	1	• SEIFC • Air flow-rate: 5 times per day	0.03[b]	C
Ahn et al., 2010*	Separate-stage BNR	• T=23 °C • DO=4.6 mg L^{-1}	nm	122.8	18.7 [b]	3.5[b]	nm	1	• SEIFC • Air flow-rate: 5 times per day	0.01[b]	C
Ahn et al., 2010*	Step-feed BNR	• T=29 °C • DO=5.2 mg L^{-1}	nm	131.8	16.7 [b]	22[b]	nm	1	• SEIFC • Air flow-rate: 5 times per day	1.5[b]	C
Ahn et al., 2010*	Step-feed BNR	• T=24 °C	nm	390	20[b]	22[b]	nm	1	• SEIFC • Air flow-rate: 5 times per day	0.05[b]	C
Ahn et al., 2010*	Step-feed BNR	• T=19 °C • DO=1.9 mg L^{-1}	nm	131.2	23.5 [b]	10.6[b]	nm	1	• SEIFC • Air flow-rate: 5 times per day	1.6[b]	C
Ahn et al., 2010*	Step-feed BNR	• T=25 °C • DO=1.5 mg L^{-1}	nm	136.1	21.3 [b]	6.9[b]	nm	1	• SEIFC • Air flow-rate: 5 times per day	0.62[b]	C

Filali et al., 2013*	SND BNR	• Annular type reactor with a central anaerobic zone & an alternate aeration mode in the outer ring • Control based on continuous monitoring of NH_4^+ & O_2 concentrations • Aeration energy consumption by ≈30% for this control	230	nm	nm	nm	nm	1	• Floating chamber • 2 sampling locations representative of the measured gas/liquid mass transfer parameters	0.004%[b]	C
Sun et al., 2017*	A/O	• Total HRT=9 h (7.5 h for aerobic zone & 1.5 h for anoxic zone) • SRT = 25 d • T= 14.2–25.7 °C	1,200	500	34.2 (24.8-43.7)	0.9 (0.2-1.7)	nm	48	• Twice per month during a 2-year period • Gas bag fixed onto a foam board using steel hoops • Off-gas equal to blower flow-rate	1.6%	GS
Foley et al., 2010	Johannesburg WWTP	• Submerged aspirating aerators	nm	10	85-114[p]	4.6-5.4[p]	nm	2	• Floating gas hood & gas bags • 2-4 h of intensive sampling	1.5%[y]	GS[S]
Bellandi et al., 2018	UCT	• 3 concentric rings: 1 covered anaerobic tank (inner ring), 1 covered anoxic tank (middle ring) & 1 open air aerobic/anoxic tank (outer ring) • Plate aerators evenly distributed	750	250	nm	0-4	nm	1	• 3 floating hoods on tank surface (anoxic zone before entering the aerated compartment, beginning & end of aerated zone)	0.1%[r]	C
Baresel et al., 2016*	UCT	• Sweden's 3rd largest municipal WWTP situated underground on Lidingo island • Serving 11 municipalities in Stockholm region	nm	nm	nm	≤10	nm	60	• Measurement campaign during 3 months in 1 of newer treatment lines (block BB11) • BB11: 7 zones, 3 aerated in the middle	0.3%	C

		• Varying DO concentrations applied in the process							• N_2O-concentration measurements using Teledyne analytical instrument (Model GFC-7002E) • Dissolved N_2O: Clark-type microsensors (by Unisense Environment A/S) at different locations in the basin (both aerated & non-aerated zones)		
Ahn et al., 2010*	4-stage Bardenpho	• T=14 °C • DO=4.4 mg L^{-1}	nm	34.5	24.3[b]	4.8[b]	nm	1	• SEIFC • Air flow-rate: 5 times per day	0.16[b]	C
Ahn et al., 2010*	4-stage Bardenpho	• T=23 °C • DO=2.3 mg L^{-1}	nm	36.8	20.1[b]	2.1[b]	nm	1	• SEIFC • Air flow-rate: 5 times per day	0.6[b]	C
Sun et al., 2015*	OD	• 4 corridors • 3 aeration brushes set at equal intervals in each corridor • DO concentration (1 m under water surface) dropping below 0.5 mg L^{-1} at 10 m behind the aeration brush	500	200	40-60	nm	nm	730	• Bimonthly sampling (48 d) • Gas bag • 6 gas samples collected from each aerated sampling point during 2 parallel experiments	0.25%	GS
Li et al., 2016*	OD	• Surface aerators • Temperature/DO not specified	nm	150,000	20	8	nm	nm	• Sampling bag collected every 2-3 d • Pre-experiment to select sampling points representative to reactor zones (3 oxic zones, 1 anoxic & 1 anaerobic) • Hand-crafted SEIFC	6-10%[b]	GS

									• Summer (May-August) and winter (November-January) • Investigation of seasonal effects		
Ren et al., 2013*	OD	• T=12.5-23.5 ℃	260	100	47.4	1.7	11.1	nm	• SEIFC & gas bag • Off-gas equal to blower flow-rate • N_2O calculated based on gas bag concentration change • Gas samples collected every 10 min for a total of 30 min • March to June	0.098%[g]	GS
Ahn et al., 2010*	OD	• T=19 ℃ • DO=0.6 mg L^{-1} [e]	nm	15.5	25.2 [b]	2.8[b]	nm	1	• SEIFC • Air flow-rate: 5 times per day	0.03[b]	C
Foley et al., 2010	OD	• Extended aeration with diffusers	nm	38	75-103[p]	2.1-3.3[p]	nm	2	• Floating gas hood & gas bags • 2-4 h of intensive sampling	0.8%[y]	GS
Yan et al., 2014*	OD (domestic & industrial)	• HRT=10 h • SRT=12 d • Temperature/DO not specified	nm	70	43-87.1	nm[j]	nm[j]	90	• Gas bag • 9 months (from March to November); once per month • Investigation of seasonal effects • Various sampling locations[i]	0.08%[o]	GS
Sun et al., 2013*	SBR	• Feeding, aeration, settling, decanting (1 h each) • 6 continuous 4-hour operating cycles in the SBR tank per day	nm	80	60-90[p]	nm	nm	365	• Bimonthly sampling • Gas-bag • Seasonality integrated in the EF	5.6%	GS
Rodriguez-Caballero et al., 2015	SBR	• Cycle: reaction phase (~130 min), settling (~65 min) & decanting (~65 min)	48	nm	38.7	3.3	~0	33	• Commercial floating chamber (AC'SCENT® Flux Hood)	6.8%[q]	C

		• NH_4^+ setpoint automatically regulated • Continuous wastewater treatment • T=15.4 ℃ • HRT=10.3 h • SRT=26.3 d • DO=0-2.5 mg L^{-1}							• Off-gas equal to blower flow-rate		
Foley et al., 2010	SBR	• 4 compartments with diffused aeration & bio-selector zone	nm	137	47-58[p]	12.1-18[p]	nm	2	• Floating gas hood & gas bags • 2-4 h of intensive sampling	2.3%[y]	GS
Townsend-Small et al., 2011	Packed-bed biofilm reactor	• Designed only for carbon oxidation • Intense aeration (odor management) & warm temperatures resulting in N-removal • No supplemental carbon addition for denitrification • Fixed-speed centrifugal blowers	nm	87	40[s]	26[s]	nm	nm	• Air samples collected directly from the blowers flushing air from the trickling filters • Samples collected at mid-day on weekdays (average N-loading & wastewater levels)	0.25%[t]	GS
Quinn Brannon et al., 2017	Aerated IFAS	• Combined • sewer from domestic & industrial sources • 10 identical open • air tanks, each with the following 4 main zones: pre-anoxic, aerated IFAS, post-anoxic & reaeration • Aerated IFAS providing • additional surface area for biofilm growth • T=20.3 ℃ • DO=0.3-2.8 mg L^{-1}	226	145	3.3	nm	nm	16	• Measurements once in June & once in October • N_2O emission measurements • & water samples collected in each of the four zones • Water samples collected just below the water surface within 3 h of the emission measurements	0.02-0.04%[v]	C

Bollon et al., 2016a	full-scale nitrifying biofiltration	• Receiving ≈80% of the Paris (France) wastewater flow • 84 Biostyr® filters for nitrification • Continuous aeration through perforated tubes located under the support material • Aeration flow-rate is automatically controlled outer ammonium concentration. • Washing procedure automatically triggered (every 20-30 min lasting for 30 min)	5,000	1,700	Summer: 1.5 [x] Winter: 1.2 [x]	Summer: 1.2 [x] Winter 1 [x]	nm	21	• 1 campaign 7 days 1 campaign 14 days (summer, winter) • Floating chamber for gaseous samples • Gas flow measured using a precision mass flowmeter (Brooks Instrument®, SLA 5860S, USA) • Off-gas directed to infrared analyzer (AP2E, ProCeas®, France) to measure N_2O concentration online • Seasonal EF investigated	Summer: 2.26% Winter: 4.86% Average: 3.56% [w]	C
Bollon et al., 2016b	Full-scale post-denitrifying filter	• Receiving ≈80% of the Paris (France) wastewater flow • The monitored denitrifying biofilters operated downstream from the biological treatment (post-denitrification) • 84 Biostyr® filters for nitrification followed by 18 Biostyr® & 12 Biofor® filters for post-denitrification with methanol as external carbon source • Biostyr® filters monitored because of treating main part of NO_3^- load	5,000	1,700	Summer: 7.1; Winter: 6.7	nm	NO_3^- removal rate= 2.24 kg N m^{-3} d^{-1} (summer) / 2.66 kg N m^{-3} d^{-1} (winter)	21	• 1 summer (September 2014) & 1 winter campaign (late January, early February 2015) • Online dissolved N_2O measurement with microsensors (Unisense A/S, Denmark)	Summer: 1.3±2% of NO_3^- uptake; Winter: 0.2±0.3 % of NO_3^- uptake	C

		• Temperature=22.5 ℃ (summer)/14.7 ℃ (winter)									
Wang et al., 2016a*	Full-scale biological aerated filter (BAF) for secondary nitrification	• Municipal WWTP in Shanghai (China) • Biostyr® & Biofor® filters for post-denitrification with methanol as external carbon source • After sedimentation ≈50% of treated wastewater • enters the BAF for secondary nitrification to polish effluent • 8 BAFs • BAF HRT=3 h	200	24	0.05 -5.9	0-1.5	nm	365	• Investigation of diurnal & seasonal variations • Seasonal monitoring conducted monthly from May to April the next year (generally from 9 am to 11 am the following day) • Floating hoods • Online N_2O analyzer (AO2020 Uras26, ABB Automation GmbH, Germany)	0.02-0.83%	C
Kosonen et al., 2016 27*	Full-scale WWTP (BNR)	• WWTP in Helsinki (Finland) • Wastewater: 15% industrial & 85% domestic • DO=3.5 mg L^{-1} • HRT=7.5 h • Aeration air flow=9-10 m^3 s^{-1}	840	310	42.1	4.6	nm	365	• Total N_2O emissions continuously measured via an online Fourier transform Infrared (FT-IR) analyzing unit situated in the effluent air channel • Measurement data available at 1-min intervals covering all WWTP operations • Monitoring campaign: 1st July 2012-30th June 2013	1.9%	C
Samuelsson et al., 2018*	AS	• 6% from industry • high-loaded activated sludge; pre-denitrification, simultaneous phosphorus precipitation, BOD • removal	806	404	22.9[z]	5.6[z]		1	• 1 monitoring hour with FTIR	0.016%	

Samuelsson et al., 2018*	Trickling filter	• 6% from industry • Nitrification	806	404	22.9[z]	5.6[z]		5	• 5 hours total at different dates (1h/d) with FTIR	0.71%
Samuelsson et al., 2018*	MBBR	• Post-denitrification • Methanol addition	806	404	22.9[z]	5.6[z]	nm	1	• 1 hour with FTIR	0.021%

a. Total: 1.9±0.25% of the influent N-load; 1st step:0.7± 0.1%, 2nd step: 3.5± 0.5%
b. TKN
c. Temperature/DO data from online sensors
d. Online for 48-72 h per week for 10 weeks between June & October 2013
e. Data from online sensors
f. 0.13± 0.01%: total N_2O emission from BNR process with respect to TN-removed from BNR process; for conversion to TN in the influent see supplementary material
g. 0.14±0.01%: total N_2O emission from BNR process with respect to TN-removed from BNR process; for conversion to TN in the influent see supplementary material
h. 0.114±0.006%: total N_2O emission from BNR process with respect to TN-removed from BNR process; for conversion to TN in the influent see supplementary material
i. Not specified
j. Significant variation in the removal efficiencies from <10% to >90%
k. 0.6-5.9% calculated based on g N_2O per kg TN-removed; for conversion to TN in the influent see supplementary material
l. Range of TN-removed
m. % TN-removed. Average: 1.6 g N_2O per kg TN-removed
n. Range:0.6-5.9 g N_2O/kg TN-removed; average: 2.3 g N_2O/kg TN-removed; for conversion to TN in the influent see supplementary material
o. Range: 0.8-8.8 g N_2O/kg TN-removed; average: 3.6 g N_2O/kg TN-removed; for conversion to TN in the influent see supplementary material
p. TN
q. With respect to the influent NH_4^+ load (%) (different configurations)
r. NH_4^+ influent
s. TIN
t. % TIN influent
u. 0.4% of N-removed released as N_2O=0.25% of N_2O released per kg N-influent; detailed calculations provided in the supplementary material
v. % N-removed
w. % NH_4-N removed
x. Applied and removed N-load respectively ($kgN/m^3/d$)
y. % N_2O–N kgN^{-1} denitrified
z. Plant influent/effluent concentrations

Appendix B

Table B.1: List of abbreviations for the sensors of the system

Abbreviation	Unit	Description
N_2O PF	$kg\ N_2O\ h^{-1}$	N_2O emission from plug flow reactor – sampling point 1
N_2O C	$kg\ N_2O\ h^{-1}$	N_2O emission from northern carrousel reactor
NH_4-N PF	$mg\ NH_4 - N\ l^{-1}$	Ammonium concentration in plug flow reactor
NO_3-N PF	$mg\ \ NO_3 - N\ l^{-1}$	Nitrate concentration in plug flow reactor
DO PF	$mg\ O_2\ l^{-1}$	Dissolved oxygen concentration in plug flow reactor
Influent	m^3h^{-1}	Influent flow-rate
NH_4-N C	$mg\ NH_4 - N\ l^{-1}$	Ammonium concentration in northern carrousel reactor
NO_3-N C	$mg\ \ NO_3 - N\ l^{-1}$	Nitrate concentration in northern carrousel reactor
NO_2-N C	$mg\ NO_2 - N\ l^{-1}$	Nitrite concentration in northern carrousel reactor
DO1	$mg\ O_2\ l^{-1}$	Dissolved oxygen concentration of sensor 1 in northern carrousel
DO2	$mg\ O_2\ l^{-1}$	Dissolved oxygen concentration of sensor 1 in northern carrousel
DO3	$mg\ O_2\ l^{-1}$	Dissolved oxygen concentration of sensor 1 in northern carrousel
T	$°C$	Water temperature in northern carrousel
TSS	$mg\ TSS\ l^{-1}$	Total suspended solids in northern carrousel

Table B.2: List with the R packages used in the analysis

Method	Package	Source
Timeseries processing	xts	(Ryan et al., 2017)
Data imputation	imputeTS	(Moritz, 2017)
Changepoint detetion	changepoint	(Killick and Eckley, 2014; Killick et al., 2016)
Correlogram Spearman's correlation	corrplot	(Wei et al., 2017)
Number of clusters	NBclust	(Charrad et al., 2014)
Data visualization	ggplot2	(Wickham, 2009; Wickham et al., 2016)
hierarchical k-means clustering and PCA		
PCA biplot	factoextra	(Kassambara and Mundt, 2017)

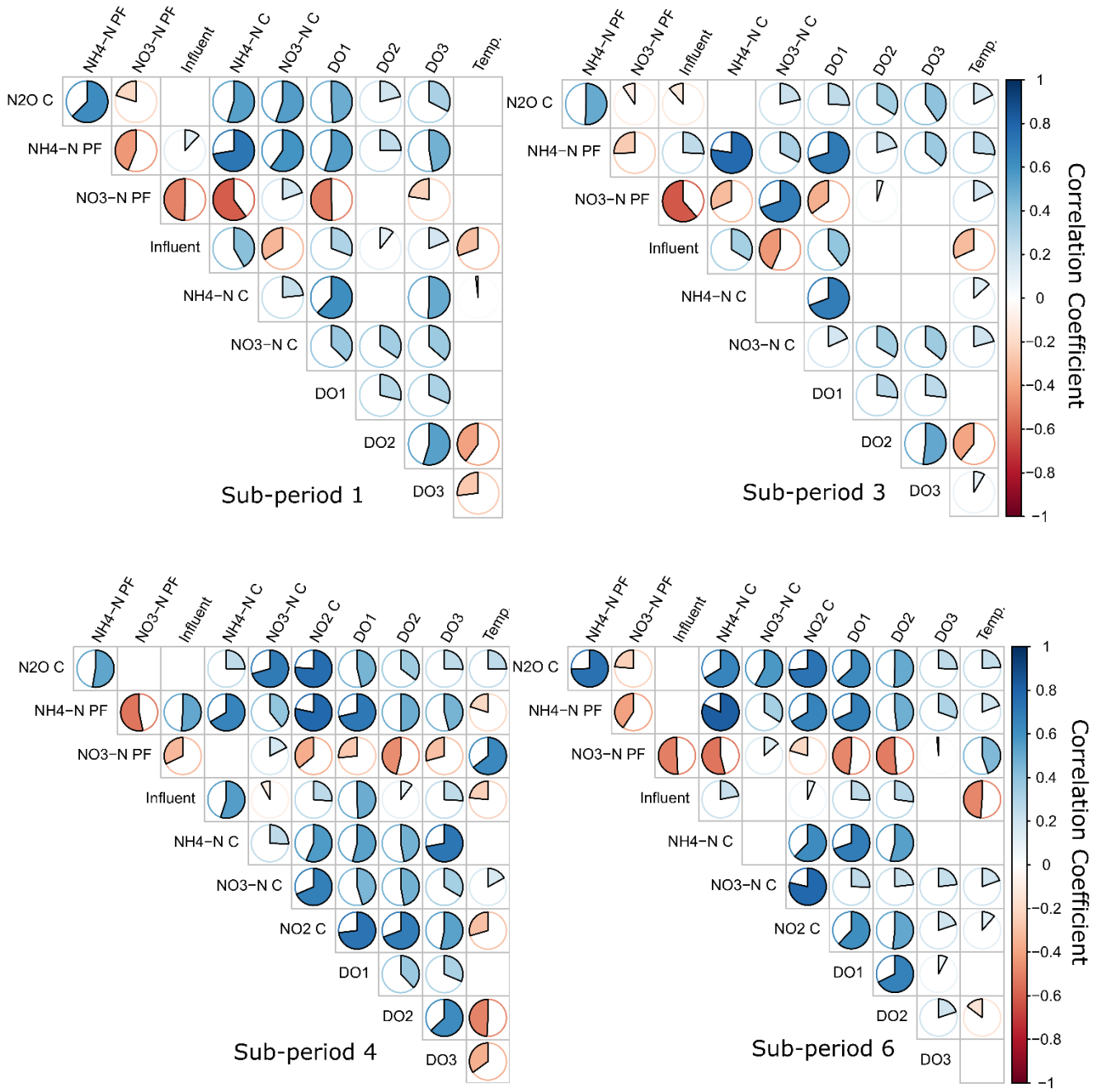


Figure B.1: Spearman's rank correlation coefficient for sensor signals for sub-periods 1, 3, 4, 6 (Red: negative correlation, blue: positive correlation, the colored part of the circles is proportional to the correlation coefficient, only results with p-value<0.01 are shown)

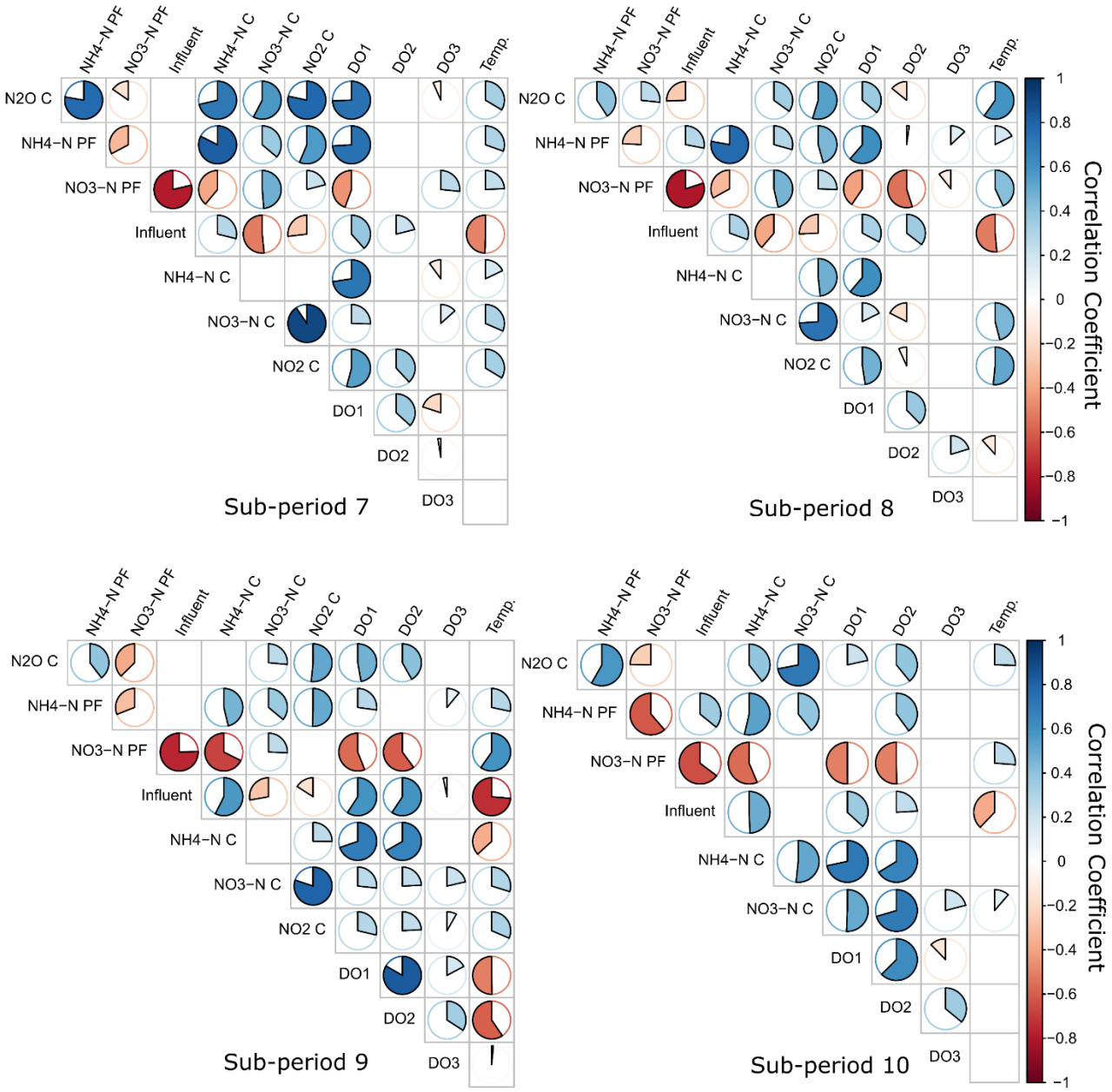


Figure B.2: Spearman's rank correlation coefficient for sensor signals for sub-periods 7, 8, 9, 10 (Red: negative correlation, blue: positive correlation, the colored part of the circles is proportional to the correlation coefficient, only results with p-value<0.01 are shown)

Table B.3: Average of offline monitored operating variables for sub-periods 1-9

	Influent COD		Influent BOD5		Influent TKN		Influent TP		Effluent COD		Effluent BOD5		Effluent TKN		Effluent total P		Effluent pH	
Sub-	(mg/L)		(mg/L)		(mg/L)		(mg/L)		(mg/L)		(mg/L)		(mg/L)		(mg/L)			
period	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
1	215.4	41.0	67.9	21.8	39.9	5.6	5.9	0.7	35.4	3.9	5.4	1.3	3.3	1.6	1.0	0.5	7.88	0.1
2	238.9	25.6	77.8	21.4	39.9	3.2	5.7	0.5	36.7	3.8	6.5	2.1	3.2	1.1	1.3	0.3	7.90	0.1
3	245.1	41.5	76.9	18.5	44.0	5.3	6.1	1.1	39.7	4.1	7.5	1.4	2.8	0.6	1.6	0.4	8.00	0.1
4	244.8	43.6	70.8	19.4	48.0	5.9	7.1	1.5	45.0	2.4	7.8	0.6	2.4	0.3	0.9	0.3	7.96	0.1
5	302.7	95.1	78.5	20.4	52.7	4.7	9.5	2.1	43.3	8.5	6.5	2.2	3.0	1.2	1.1	0.9	8.12	0.2
6	208.3	91.8	65.3	35.2	34.3	9.6	6.9	2.3	32.2	6.5	3.4	2.4	2.5	1.5	1.0	0.6	8.02	0.1
7	151.0	21.0	41.5	2.5	30.5	6.5	5.4	1.2	33.0	1.0	2.4	0.8	2.8	1.0	1.3	0.2	8.10	0.0
8	177.7	23.2	54.0	8.3	29.3	3.4	4.9	0.5	29.7	4.6	3.0	0.6	2.3	0.7	0.6	0.1	7.93	0.1
9	250.5	95.2	80.3	41.7	44.5	12.8	7.6	2.1	35.0	3.5	6.3	1.8	3.2	0.7	1.0	0.3	7.94	0.2

Table B.4: Similarity of the clustering results in the Carrousel reactor when NH_4-N PF and NO_3-N PF are not included in the analysis (the percentage represents the data rows assigned to the same cluster)

Sub-period	Similarity of Clustering results
1	90%
2	75%
3	100%
4	85%
5	87.5%
6	95%
7	60%
8	63%
9	76%

Table B.5: Average of operating variables for all clusters defined by hierarchical k-means clustering

	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	**T**
		kg/h	mg/l	mg/l	m^3/h	mg/l	mg/l	mg/l	mg/l	mg/l	mg/l	°C
1	1	0.09	14.13	1.48	3883	1.47	8.66	1.04	0.78	1.72		15.7
	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		11.2
	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		11.5
	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	0.95	2.41	2.14	4.10	12.9
	11	3.83	10.82	0.77	2747	0.79	1.80	0.05	1.20	1.51	1.40	
	12	6.89	25.45	0.48	6375	10.86	3.62	2.12	2.34	1.98	4.28	
5	13	2.23	10.33	3.26	2485	0.97	3.59	0.11	0.34	1.96	0.67	18.2
	14	4.95	18.15	1.91	3542	3.20	7.28	1.60	0.69	2.03	2.58	
6	15	2.54	17.66	0.75	5922	5.00	5.07	1.30	0.73	2.34	1.08	20
	16	0.51	8.20	2.84	3811	0.98	2.64	0.10	0.10	2.21	0.35	
7	17	0.66	19.67	0.10	9877	13.40	3.38	1.05	0.37	2.03	0.61	20
	18	0.52	14.14	1.68	6228	2.00	10.44	1.15	0.41	2.55	1.10	
	19	0.02	7.07	1.81	6389	1.01	2.40	0.12	0.17	2.67	0.23	
	20	0.13	7.72	5.66	3636	1.03	11.32	0.07	0.34	2.81	0.80	
8	21	0.14	8.90	5.55	3834	1.01	10.06	0.19	0.17	2.74	0.74	19.6
	22	0.09	16.57	0.57	9410	8.21	7.79	1.28	0.55	2.43	0.96	
9	23	0.05	17.91	2.98	3047	1.72	10.57	0.77	0.75	1.76	2.32	
	24	0.00	7.31	8.60	2027	1.06	8.77	0.03	0.12	1.32	1.68	
	25	0.04	11.17	1.50	5379	1.27	3.18	0.53	0.62	1.42	1.05	
	26	0.65	16.64	1.67	4442	2.37	9.17	1.92	1.58	1.58	2.80	
	27	0.07	11.66	1.31	5372	2.20	9.27	2.93	2.50	1.70	1.94	
	28	0.05	16.06	0.29	10785	6.97	6.25	1.96	1.78	1.42	1.72	
10	29	0.58	16.32	1.15	4080	1.50	3.68	0.18	0.78	1.62		13
	30	2.18	19.18	1.13	3695	1.61	9.13	0.79	1.49	1.57		
	31	0.60	11.35	2.57	2982	1.36	3.70	0.00	0.02	1.14		
	32	2.77	19.16	0.30	8666	2.85	8.05	1.75	1.64	1.45		
	33	0.18	9.08	2.65	2911	1.70	12.25	4.49	4.03	2.49		

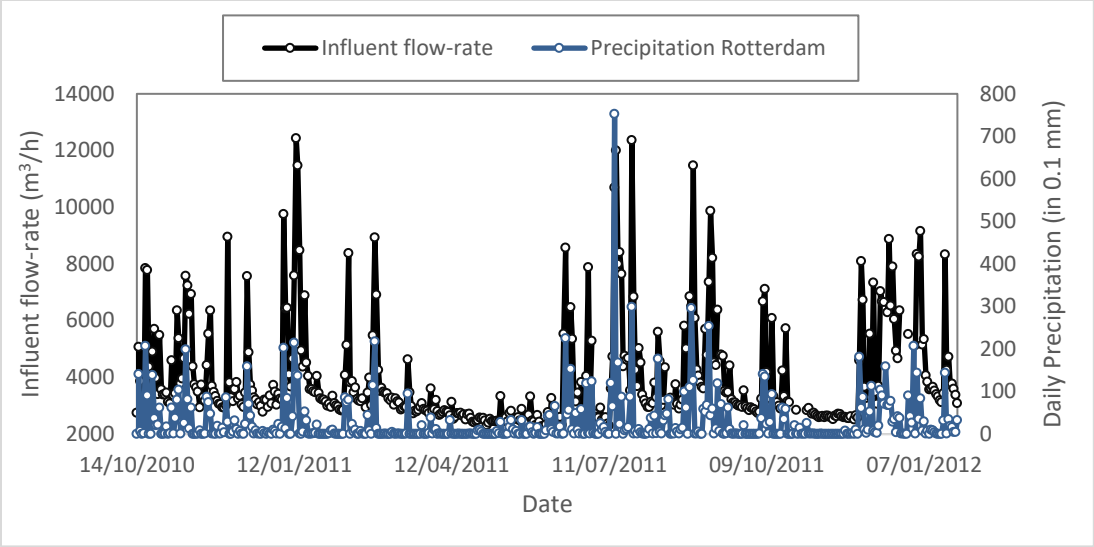


Figure B.3: Daily influent flow-rate and daily precipitation data extracted from the Royal Netherlands meteorological institute (http://www.sciamachy-validation.org/climatology/daily_data/selection.cgi station 344 Rotterdam).

Sub-period 1

Table B.6: PCA loadings sub-period 1, Carrousel reactor

	PC1	PC2	PC3	PC4
NH_4-N PF	0.45	-0.26	0.08	-0.31
NO_3-N PF	-0.18	0.45	-0.08	0.46
Influent	0.35	-0.31	0.26	0.36
NH_4-N C	0.35	-0.27	0.11	0.57
NO_3-N C	0.25	0.46	0.13	-0.12
DO1	0.44	0.19	-0.10	-0.31
DO2	0.44	0.27	-0.09	0.00
DO3	0.19	0.46	0.09	0.27
Temperature	0.16	0.00	-0.65	-0.02

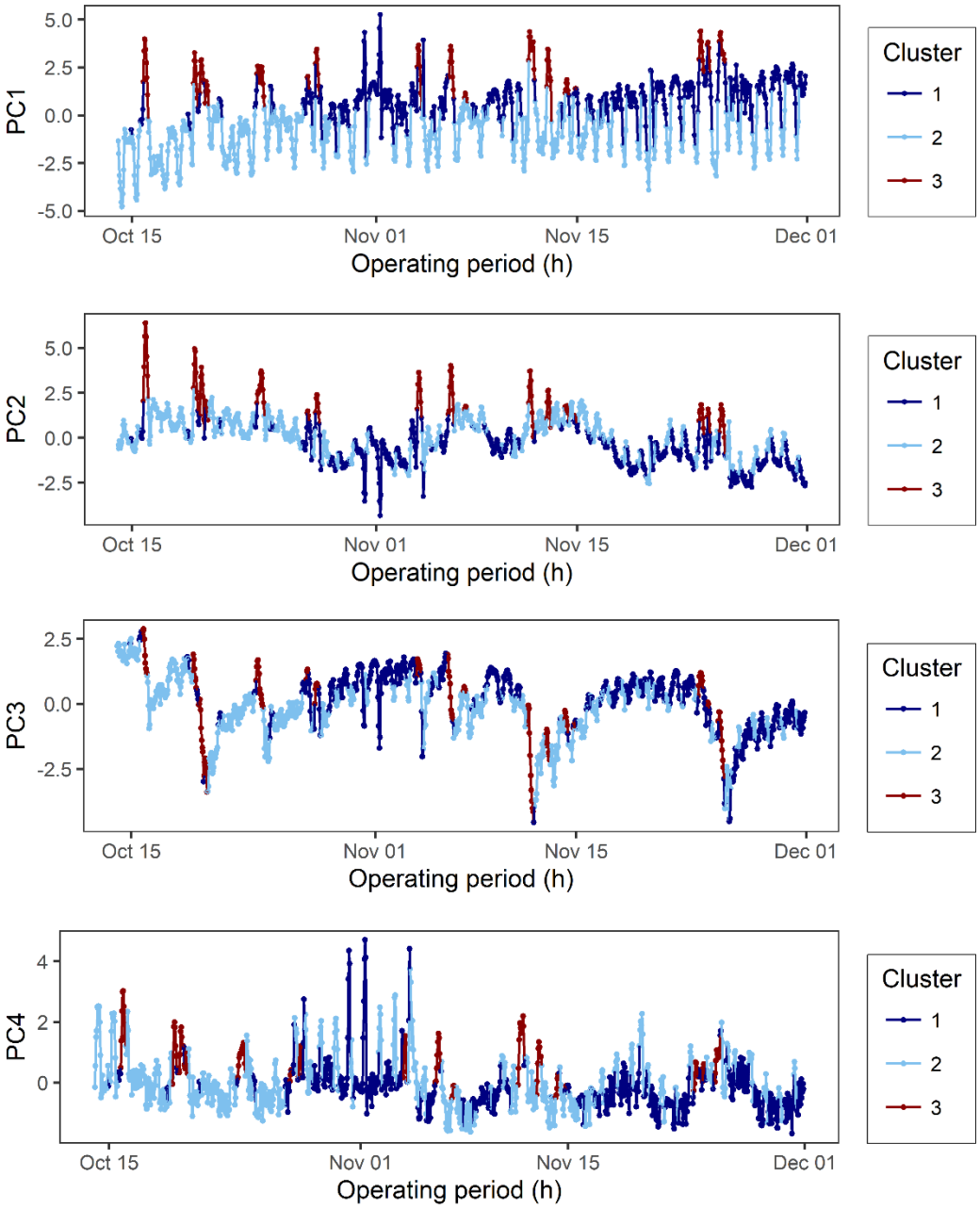


Figure B.3: PC scores for sub-period 1

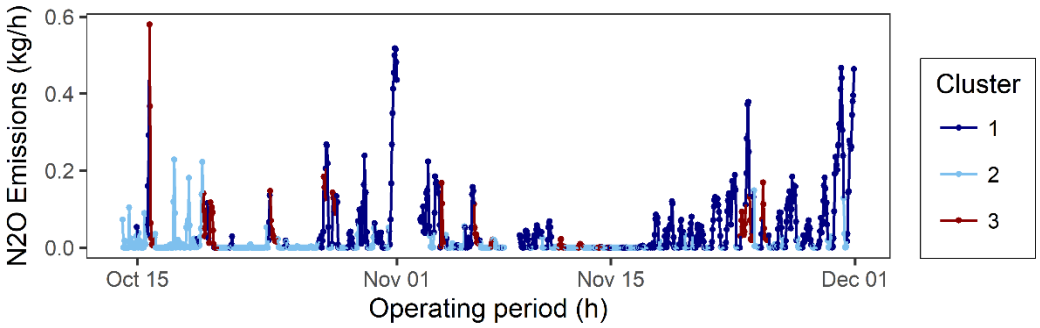


Figure B.4: Profile of N_2O emissions; colored points indicate the respective clusters for sub-period 1.

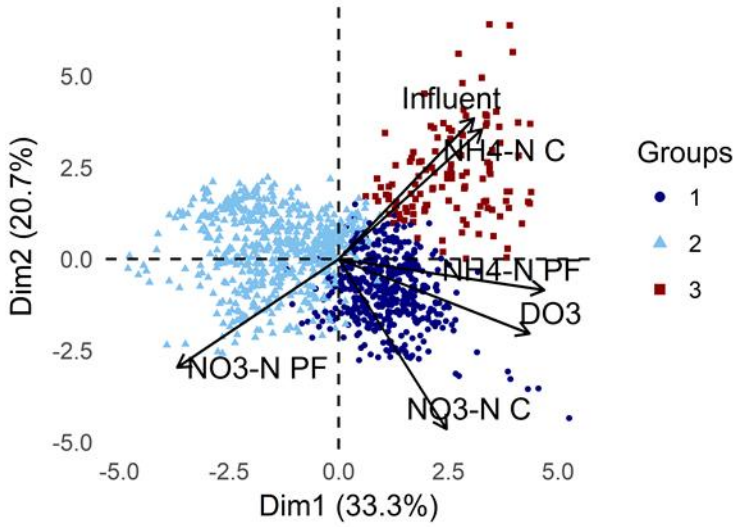


Figure B.5: Score of the first two principal components sub-period 1

Sub-period 3

Table B.7: PCA loadings sub-period 3, Carrousel reactor

	PC1	PC2	PC3	PC4
NH_4-N PF	-0.53	0.09	-0.23	0.06
NO_3-N PF	0.23	0.46	0.10	-0.56
Influent	-0.23	-0.44	0.27	-0.33
NH_4-N C	-0.40	-0.19	-0.21	-0.46
NO_3-N C	-0.10	0.58	0.08	-0.32
DO1	-0.50	-0.01	0.08	-0.16
DO2	-0.21	0.24	0.60	0.20
DO3	-0.37	0.27	0.20	0.41
Temperature	-0.12	0.30	-0.64	0.16

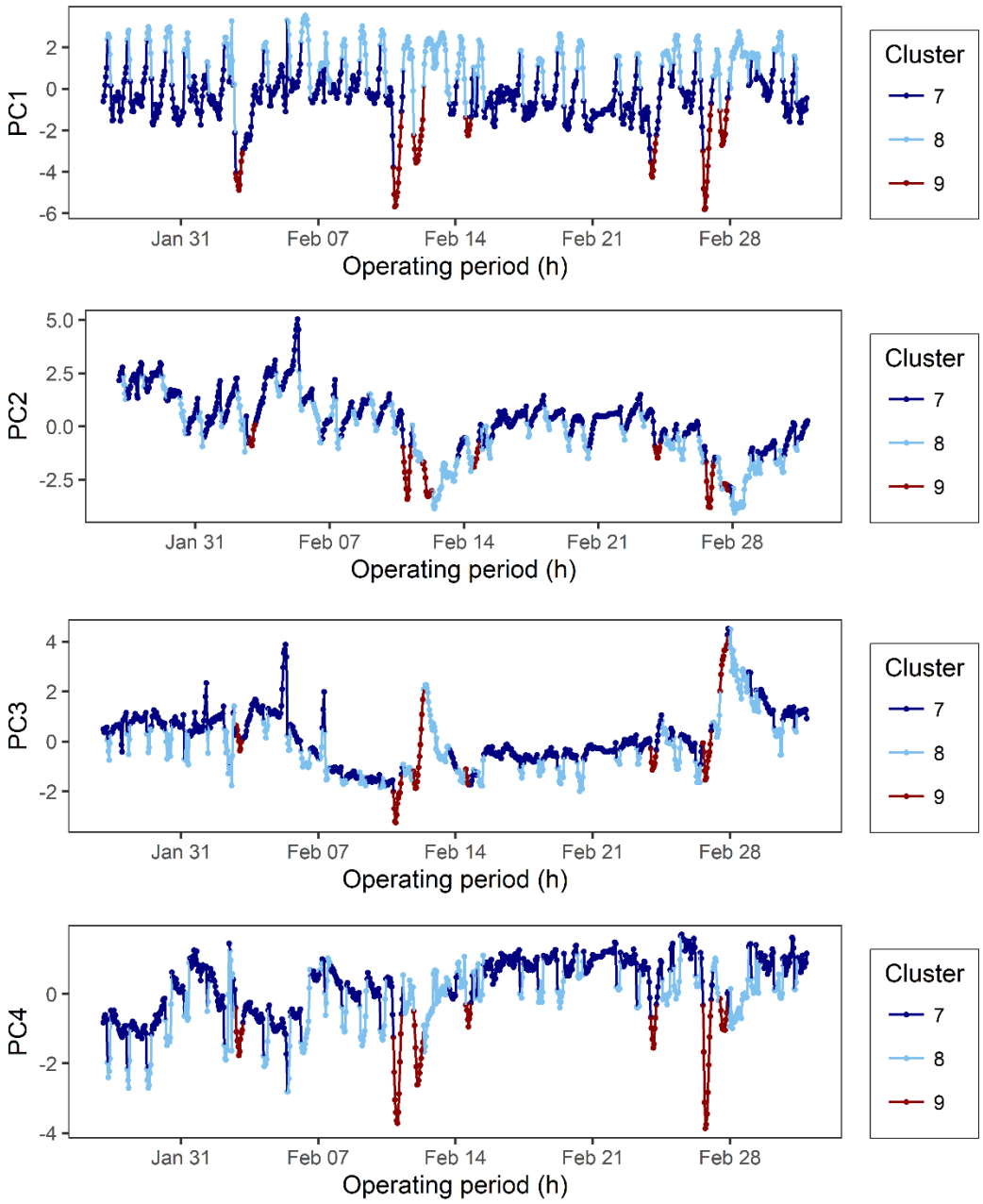


Figure B.6: PC scores for sub-period 3; colored points indicate the respective clusters for sub-period 3.

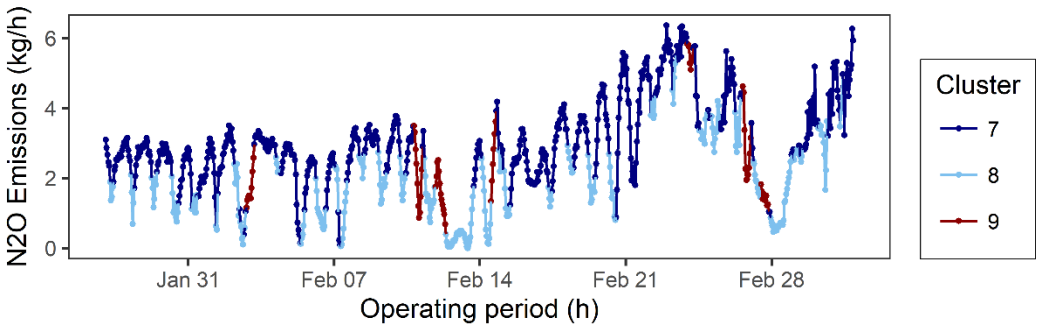


Figure B.7: Profile of N_2O emissions; colored points indicate the respective clusters for sub-period 3

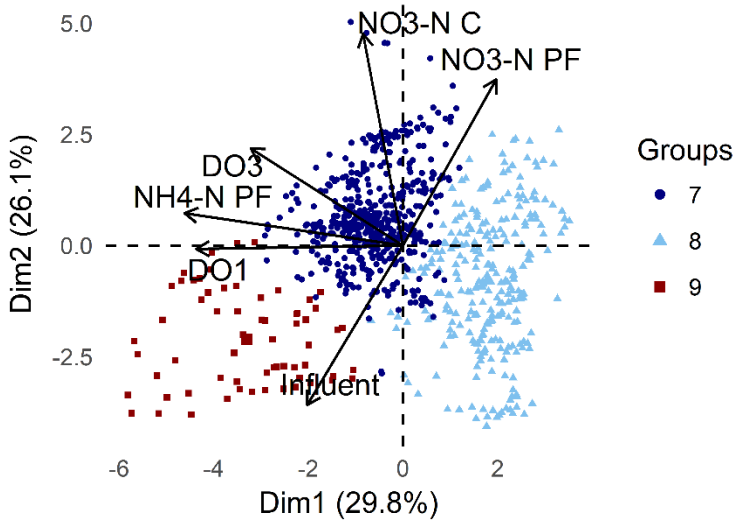


Figure B.8: Score of the first two principal components sub-period 3

- PC1 again represents the behavior of the system based on the DO control and the diurnal fluctuation of NH_4-N PF.
- PC2 represents how influent flow-rate affects the nitrates in the system (when influent flow-rate increases they decrease).
- PC3 shows an increase of DO with the decrease of temperature (even though these variables are not directly related it is possible temperature to affect the activities and therefore the consumption of DO)
- PC2 decreases after 10 of February and remains relatively low. This coincides with a significant reduction in the NO_3-N PF and NO_3-N C. Emissions start increasing 5 days later. Around 10 of February a precipitation event is observed followed by some days with diluted streams, reduced nitates and increased flow-rate. After this event the system nitrates do not recover to their previous levels.

Sub-period 4

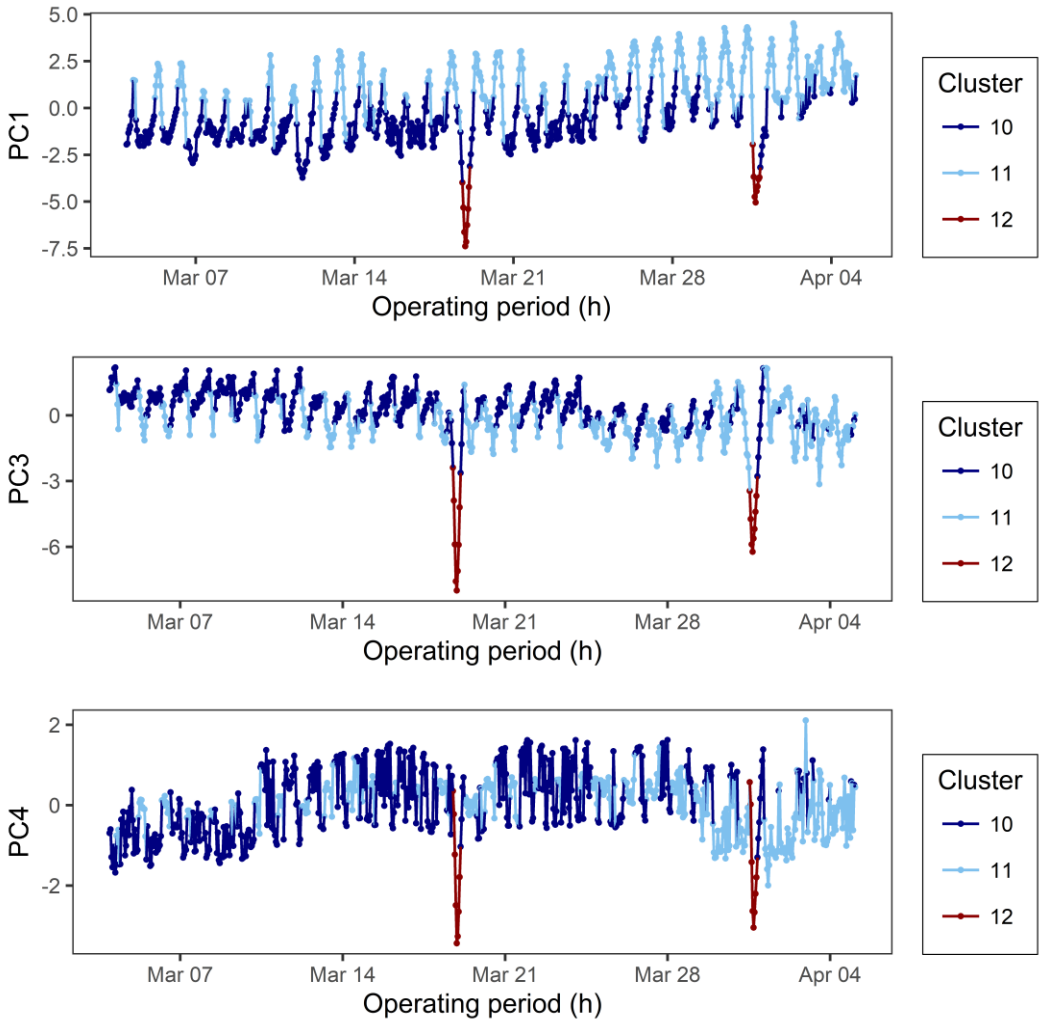


Figure B.9: PC scores for sub-period 4; colored points indicate the respective clusters for sub-period 4.

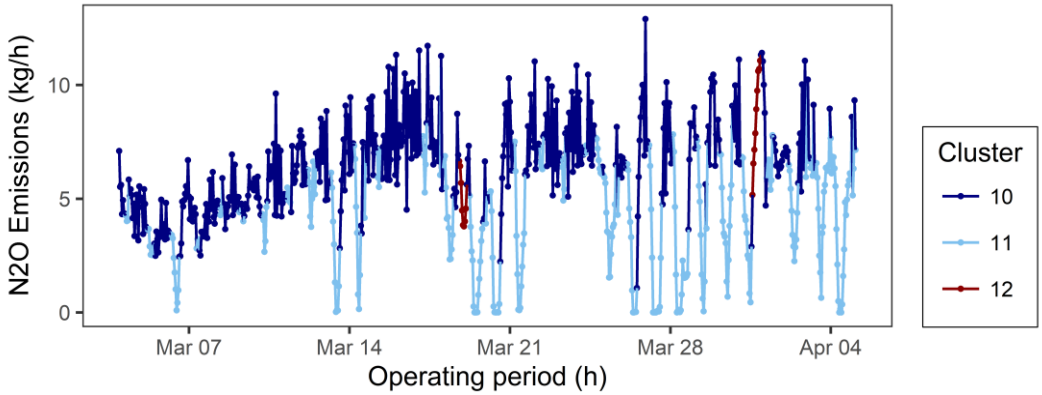


Figure B.10: Profile of N_2O emissions; colored points indicate the respective clusters for sub-period 4

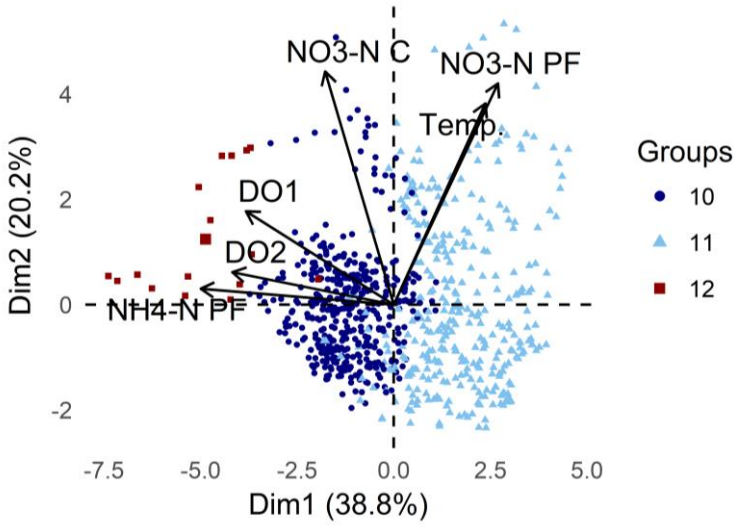


Figure B.11: Score of the first two principal components sub-period 4

Sub-period 5

Table B.8: PCA loadings sub-period 5, Carrousel reactor

	PC1	PC2	PC3	PC4
NH_4-N PF	0.45	-0.26	0.08	-0.31
NO_3-N PF	-0.18	0.45	-0.08	0.46
Influent	0.35	-0.31	0.26	0.36
NH_4-N C	0.35	-0.27	0.11	0.57
NO_3-N C	0.25	0.46	0.13	-0.12
NO_2-N C	0.44	0.19	-0.10	-0.31
DO1	0.44	0.27	-0.09	0.00
DO2	0.19	0.46	0.09	0.27
DO3	0.16	0.00	-0.65	-0.02
Temperature	0.06	-0.16	-0.67	0.22

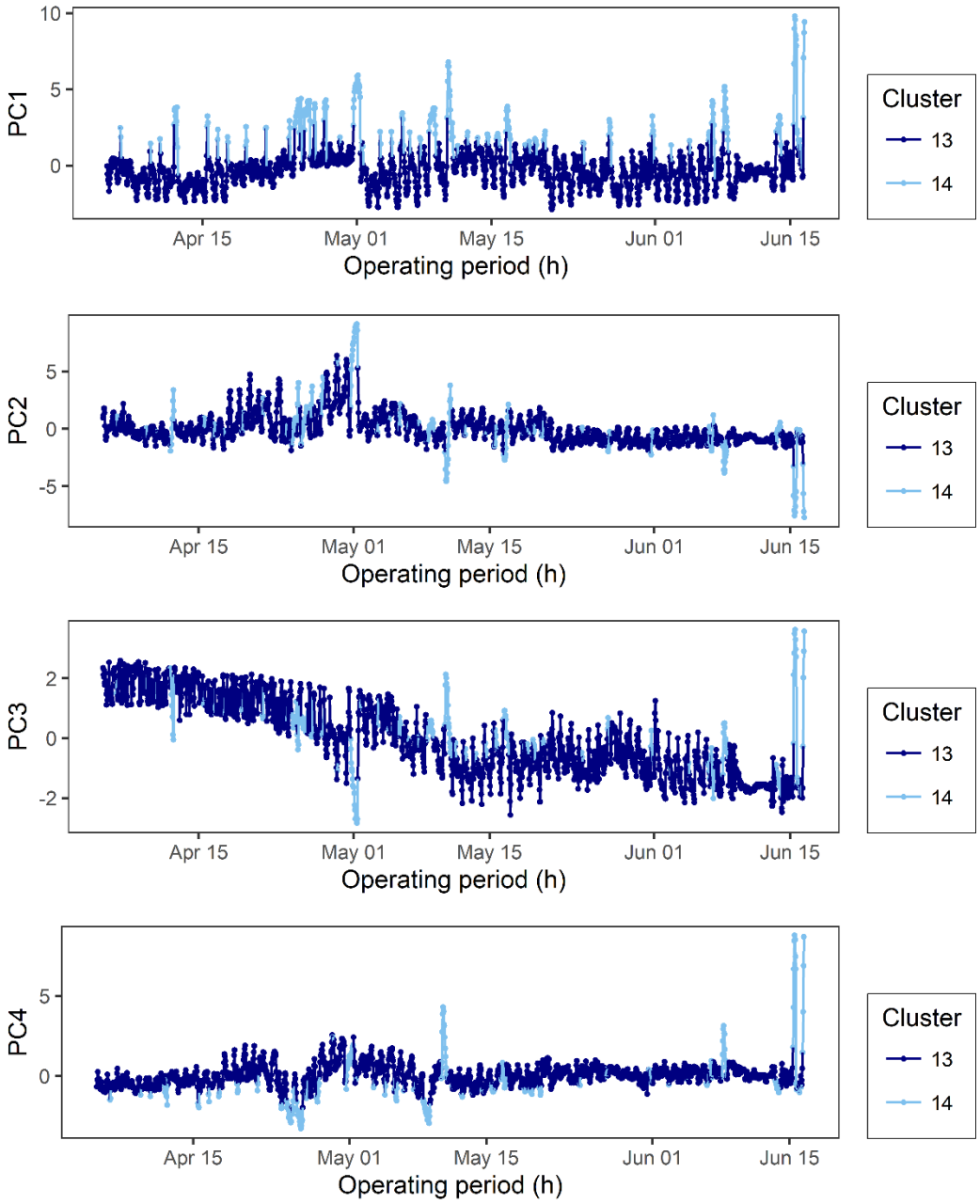


Figure B.12: PC scores for sub-period 5; colored points indicate the respective clusters for sub-period 5.

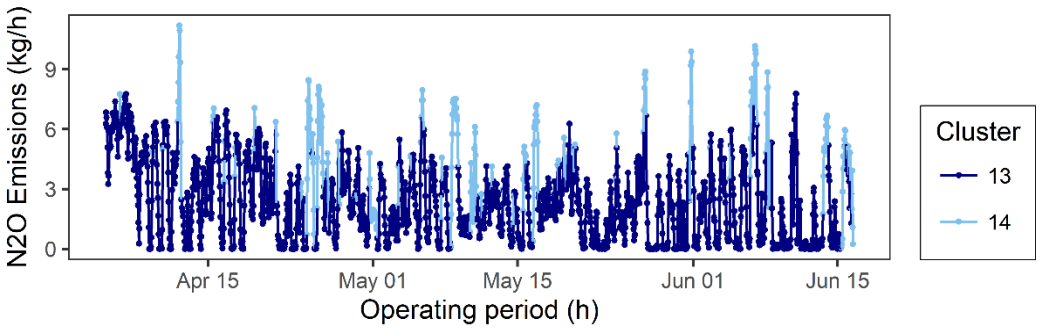


Figure B.13: Profile of N_2O emissions; colored points indicate the respective clusters for sub-period 5

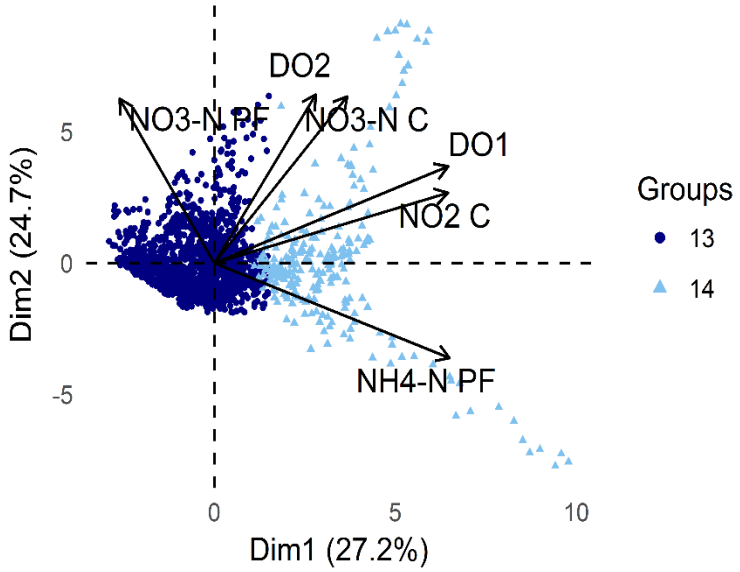


Figure B.14: Score of the first two principal components sub-period 5

- In period 5 a shift is observed in Cluster 14 around 10 of May and again NO_3-N PF and NO_3-N C are lower than the previous period (PC2 cluster 14 becomes negative from positive – score plot and bi-plot). However, it never becomes as low as in sub-period 4. Additionally, the decrease of NO_3-N C coincides with an increase in TSS in the Carrousel reactor and therefore can represent higher efficiency in NO_3-N C removal.
- In this sub-period temperature increases significantly (from 15 Oc to more than 20). DO3 at the end of the reactor increases with the temperature**.**

Sub-period 6

Table B.9: PCA loadings sub-period 6, Carrousel reactor

	PC1	PC2	PC3	PC4
NH_4-N PF	-0.39	-0.19	0.27	-0.35
NO_3-N PF	0.32	-0.36	-0.03	0.23
Influent	-0.23	0.45	-0.11	0.26
NH_4-N C	-0.34	0.11	0.49	0.02
NO_3-N C	-0.22	-0.43	-0.43	0.32
NO_2-N C	-0.38	-0.34	-0.16	0.30
DO1	-0.46	-0.01	0.13	0.06
DO2	-0.39	0.11	-0.16	0.11
DO3	-0.16	-0.11	-0.48	-0.73
Temperature	0.04	-0.54	0.43	-0.09

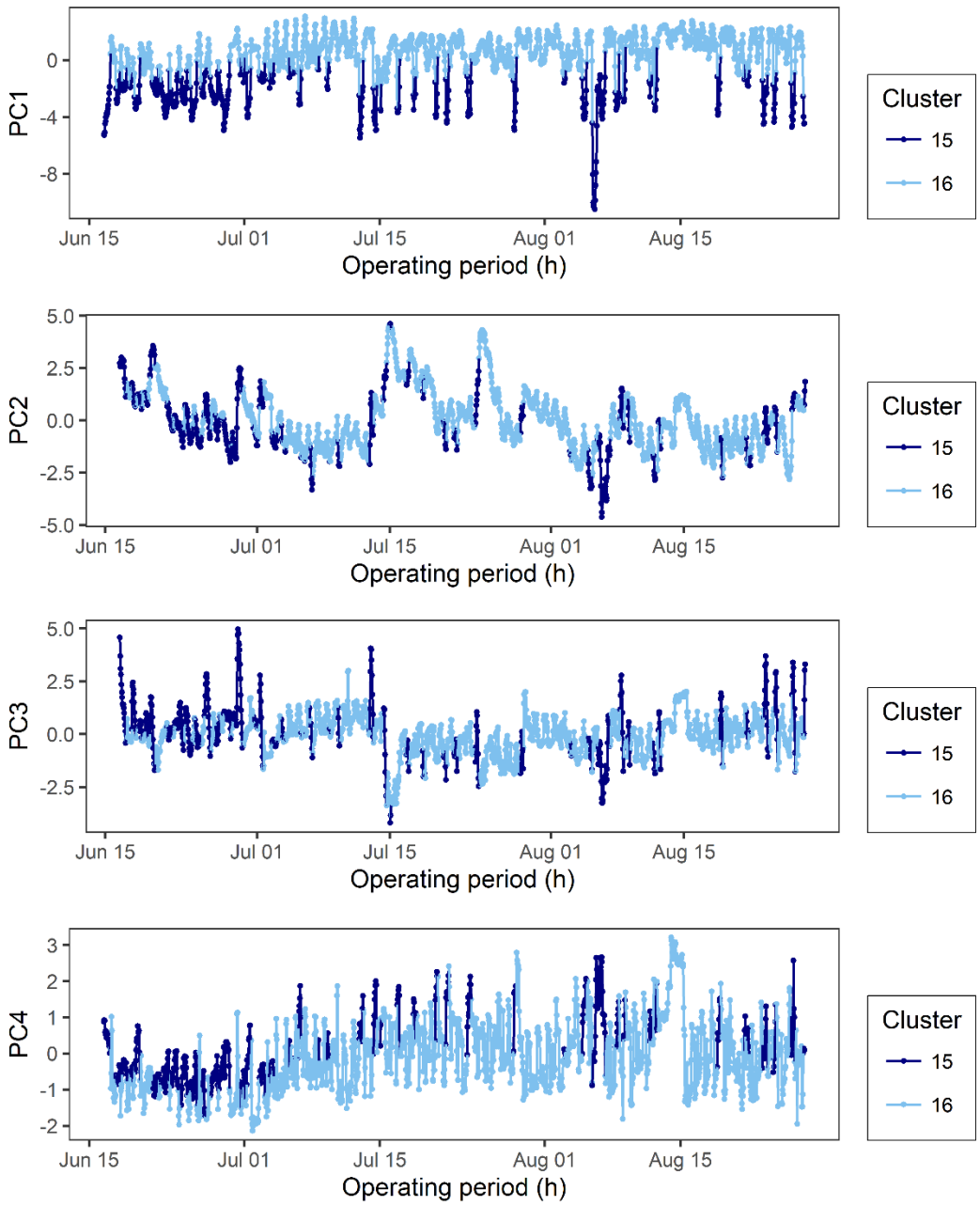


Figure B.15: PC scores for sub-period 6; colored points indicate the respective clusters for sub-period 6.

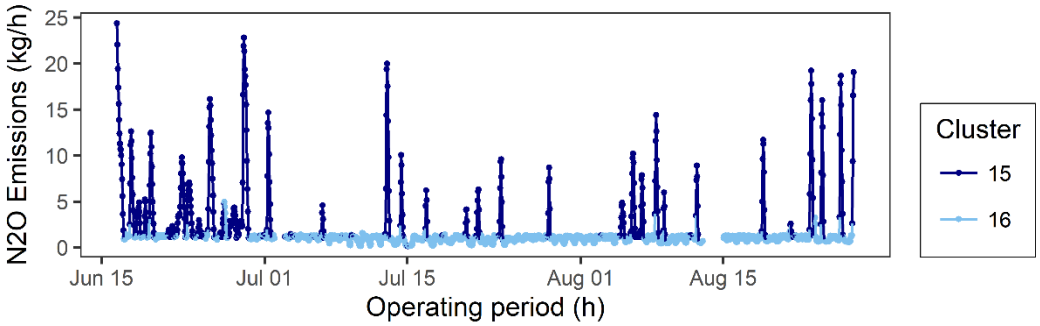


Figure B.16: Profile of N_2O emissions; colored points indicate the respective clusters for sub-period 6

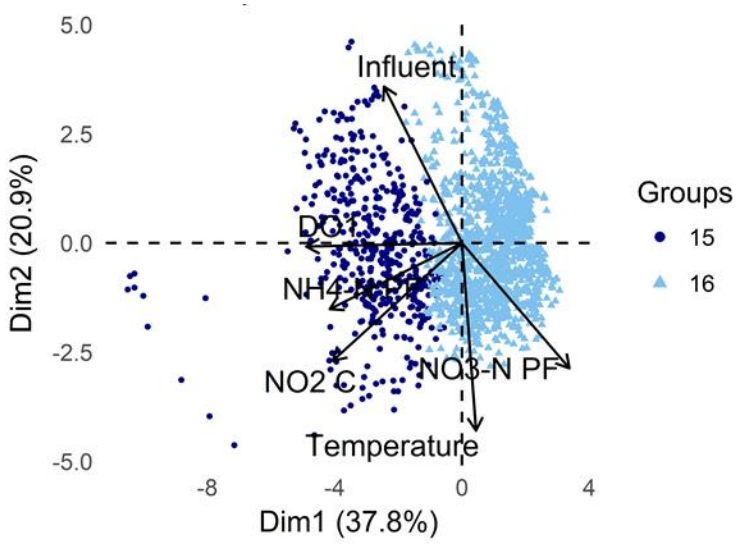


Figure B.17: Score of the first two principal components sub-period 6

- In sub-period 6 we clearly see that N_2O emission peaks coincide with peaks in NH_4-N C. In total 176 data points have NH_4-N C> 4mg/l. Out of them 156 have N_2O > 1 and 20 N_2O <1. DO1 also clearly follows the same patterns. Generally, emissions decrease again (high are up to 2.5 kg/h)
- Correlation between PC1 and N_2O emissions is equal to -0.64.

Sub-period 7

Table B.10: PCA loadings sub-period 7, Carrousel reactor

	PC1	PC2	PC3	PC4
NH_4-N PF	-0.20	0.47	-0.18	0.16
NO_3-N PF	-0.35	-0.37	-0.03	0.00
Influent	0.37	0.33	0.24	-0.04
NH_4-N C	0.05	0.41	-0.24	0.48
NO_3-N C	-0.52	0.03	0.21	-0.24
NO_2-N C	-0.47	0.25	0.21	-0.25
DO1	-0.10	0.50	0.14	-0.22
DO2	0.02	0.11	0.66	0.18
DO3	-0.10	-0.20	0.47	0.61
Temperature	-0.43	0.05	-0.31	0.41

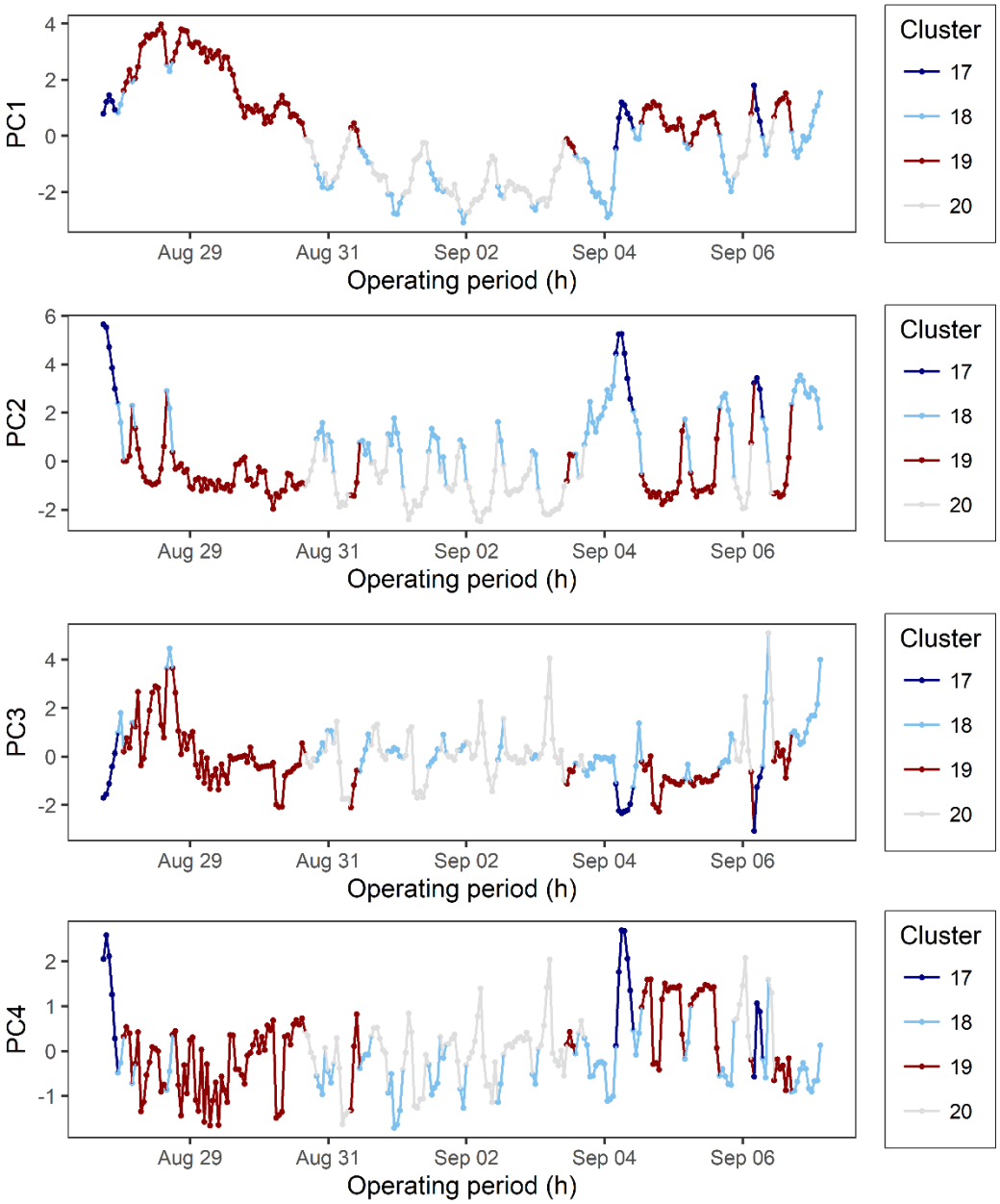


Figure B.18: PC scores for sub-period 7; colored points indicate the respective clusters for sub-period 7.

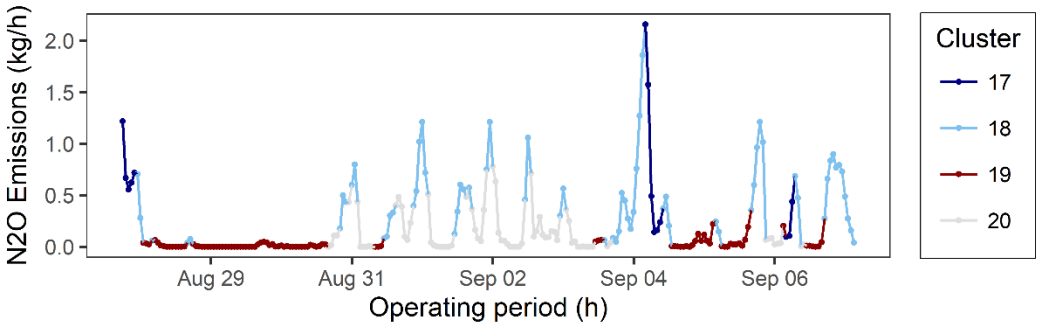


Figure B.19: Profile of N_2O emissions; colored points indicate the respective clusters for sub-period 7

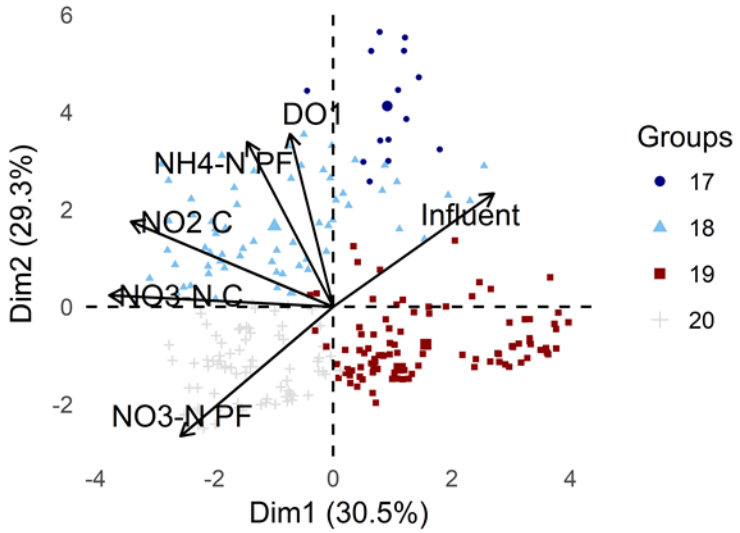


Figure B.20: Score of the first two principal components sub-period 7

- Cluster 19 is characterized by very low emissions (2-3 days in total). The cluster is characterized by diluted streams (very low concentrations of ammonium and nitrates without strong diurnal variability) with relatively high influent flow-rate.
- It is a small sub-period
- In sub-period 7 emissions increase with the increase of NO_3-N C.
- PC2 seems to represent ammonium control via DO instead of PC1 in the other sub-periods. PC2 is a good indication of the N_2O emissions variability and range
- Variation of Nitrates and influent flow-rate is represented by PC1

Sub-period 8

Table B.11: PCA loadings sub-period 8, Carrousel reactor

	PC1	PC2	PC3	PC4
NH_4-N PF	0.14	-0.48	0.13	-0.50
NO_3-N PF	-0.51	0.03	0.03	0.16
Influent	0.51	-0.03	0.10	-0.01
NH_4-N C	0.36	-0.29	0.21	0.13
NO_3-N C	-0.31	-0.38	-0.14	0.11
NO_2-N C	-0.15	-0.47	-0.21	0.09
DO1	0.26	-0.43	-0.12	0.32
DO2	0.24	0.09	-0.61	0.48
DO3	0.02	0.01	-0.69	-0.58
Temperature	-0.30	-0.35	0.03	0.09

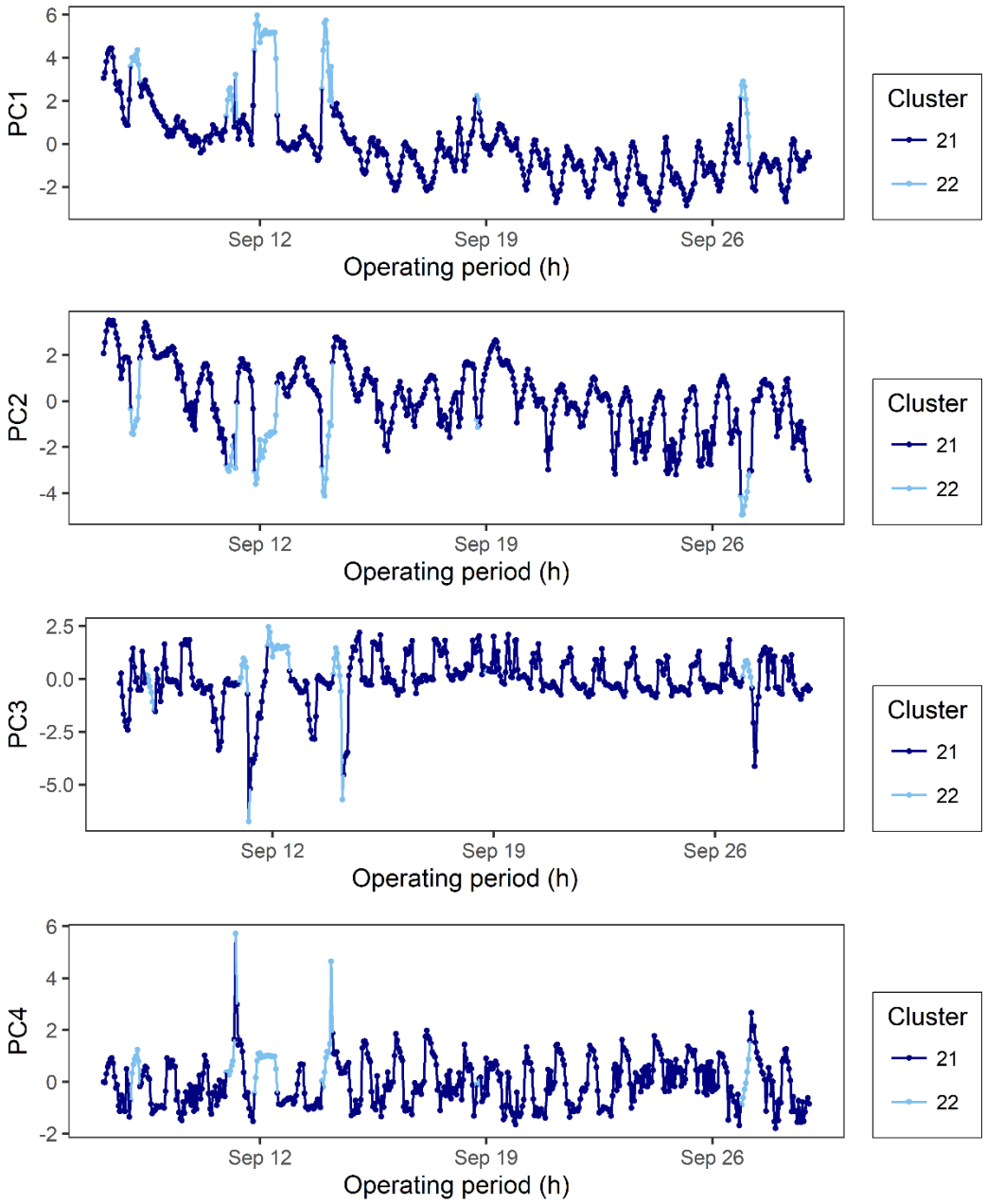


Figure B.21: PC scores for sub-period 8; colored points indicate the respective clusters for sub-period 8.

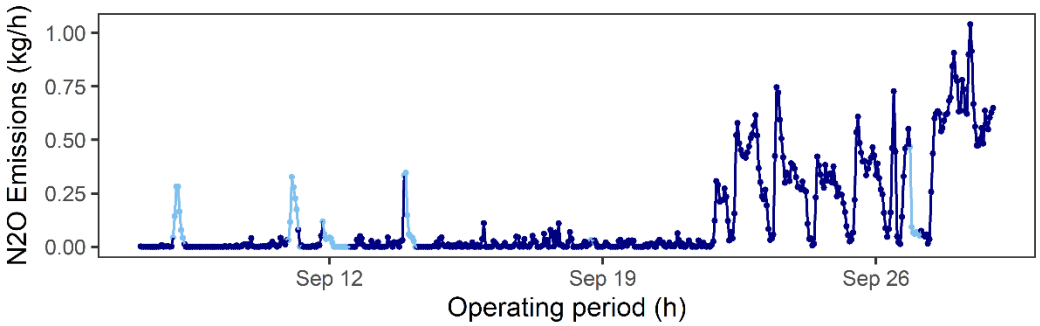


Figure B.22: Profile of N_2O emissions; colored points indicate the respective clusters for sub-period 8

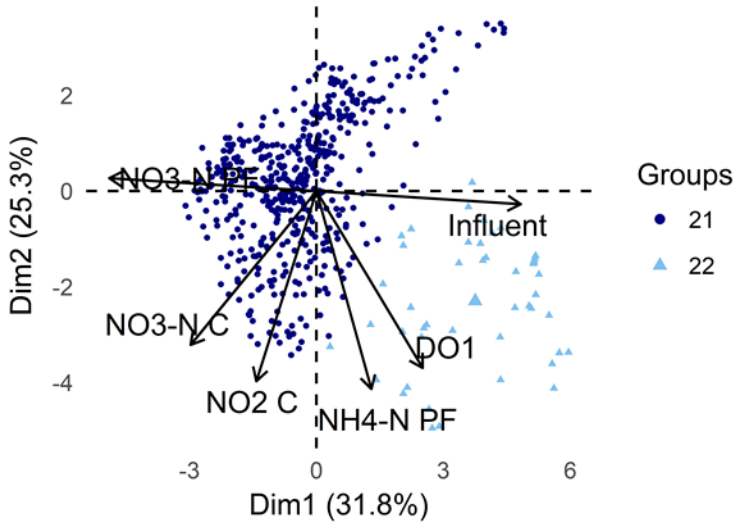


Figure B.23: Score of the first two principal components sub-period 8

- In sub-period 8 emissions increase around 22nd of September without a clear change in the state of the system and the first 3 PCs. Only PC1 and PC2 demonstrate throughout the sub-period a decreasing trend indicating that nitrate, ammonium and nitrite concentration increase in the system
- The correlation between PC2 and N_2O emissions is -0.56

Sub-period 9

Table B.12: PCA loadings sub-period 9, Carrousel reactor

	PC1	PC2	PC3	PC4
NH_4-N PF	-0.10	0.34	-0.58	0.38
NO_3-N PF	0.43	0.08	0.24	-0.32
Influent	-0.38	-0.23	-0.24	-0.06
NH_4-N C	-0.29	0.04	-0.39	-0.23
NO_3-N C	0.03	0.56	0.14	-0.26
NO_2-N C	-0.04	0.55	-0.14	-0.25
DO1	-0.43	0.20	0.25	-0.15
DO2	-0.45	0.16	0.32	-0.05
DO3	-0.07	0.25	0.37	0.73
Temperature	0.41	0.28	-0.23	0.11

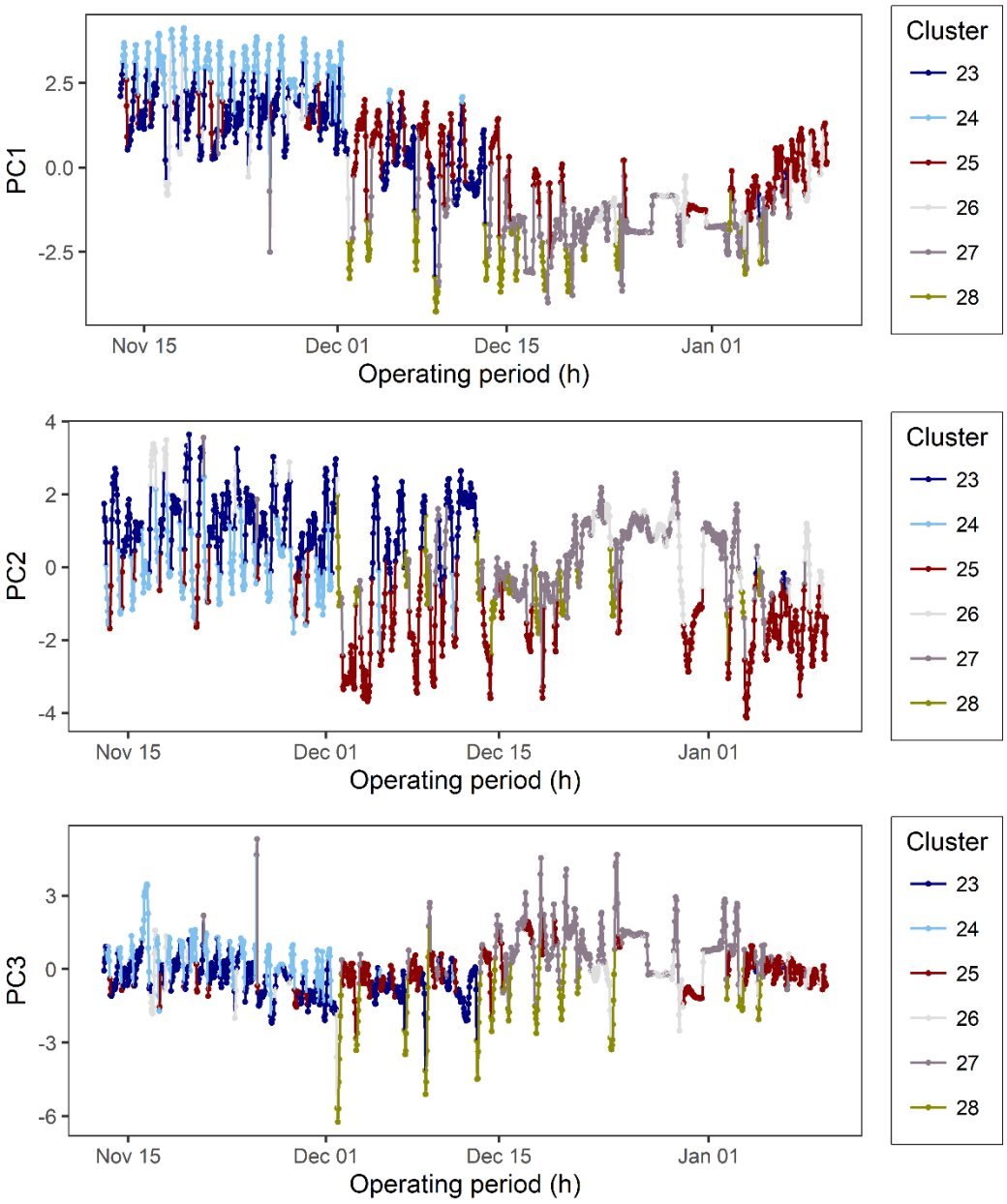


Figure B.24: PC scores for sub-period 9; colored points indicate the respective clusters for sub-period 9

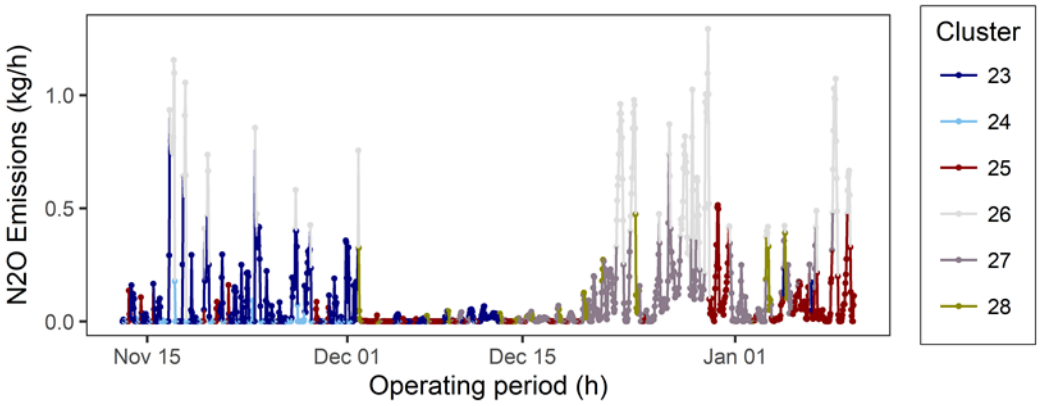


Figure B.25: Profile of N_2O emissions; colored points indicate the respective clusters for sub-period 9

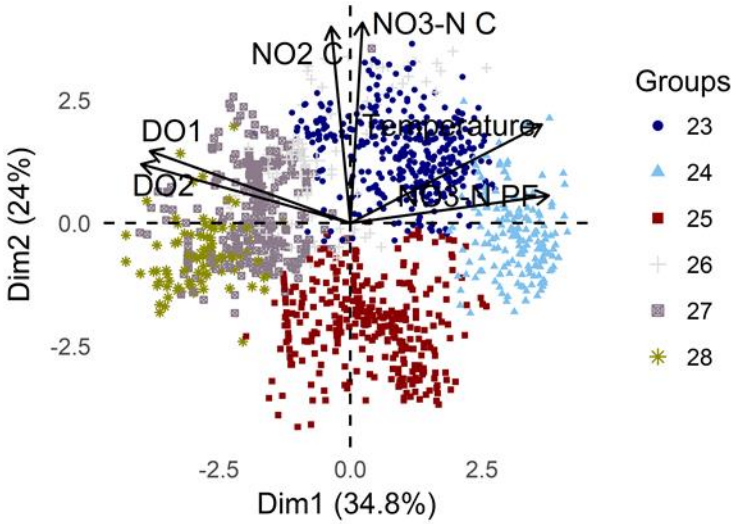


Figure B.26: Score of the first two principal components sub-period 9

- Clusters are not very clear in this subperiod

Sub-period 10

Table B.13: PCA loadings sub-period 10, Carrousel reactor

	PC1	PC2	PC3	PC4	PC5
NH_4-N PF	-0.29	-0.16	0.62	-0.28	0.03
NO_3-N PF	0.34	0.40	-0.27	0.33	-0.19
Influent	-0.27	-0.49	-0.22	0.21	-0.20
NH_4-N C	-0.38	-0.22	0.08	0.49	-0.47
NO_3-N C	-0.37	0.39	0.14	0.13	0.29
DO1	-0.41	0.19	-0.23	0.35	0.26
DO2	-0.45	0.27	-0.16	-0.13	0.19
DO3	-0.28	0.33	-0.24	-0.52	-0.64
Temperature	0.10	0.39	0.58	0.32	-0.31

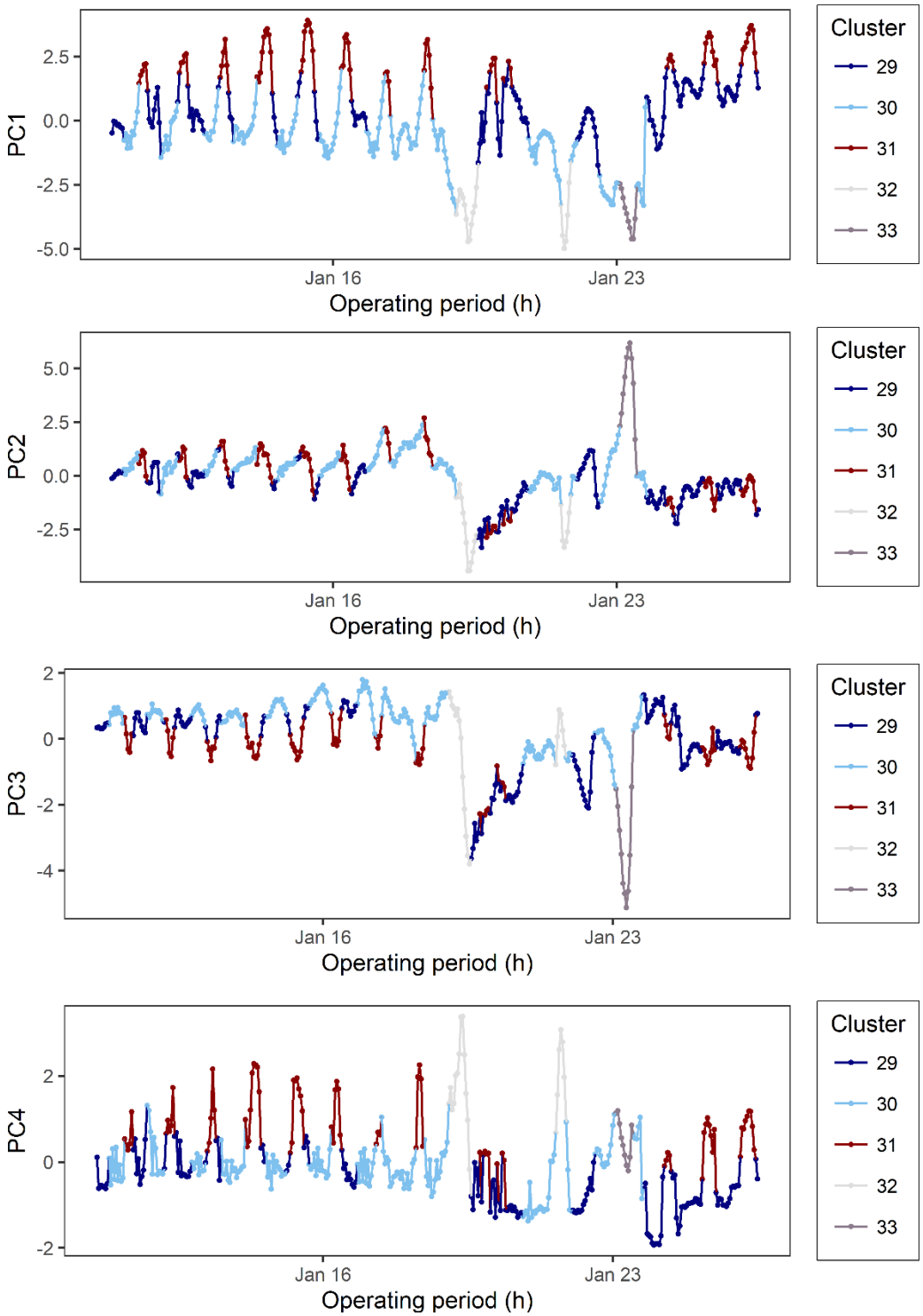


Figure B.27: PC scores for sub-period 10; colored points indicate the respective clusters for sub-period 10

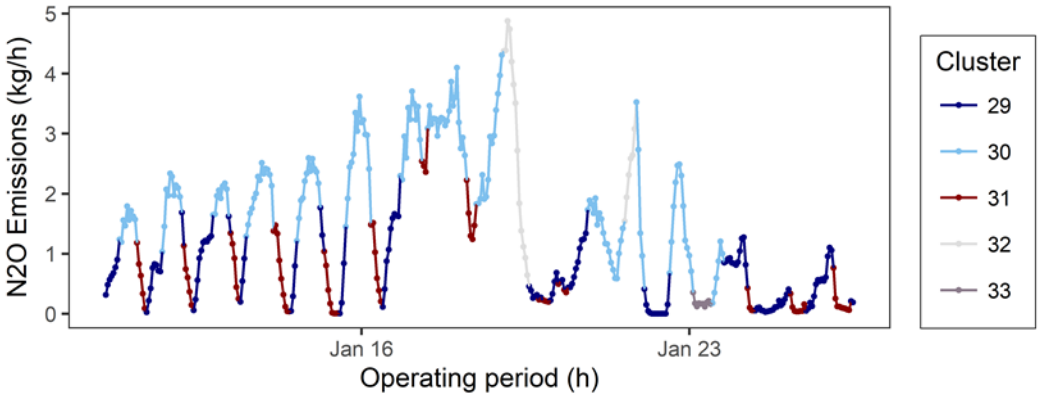


Figure B.28: Profile of N_2O emissions; colored points indicate the respective clusters for sub-period 10

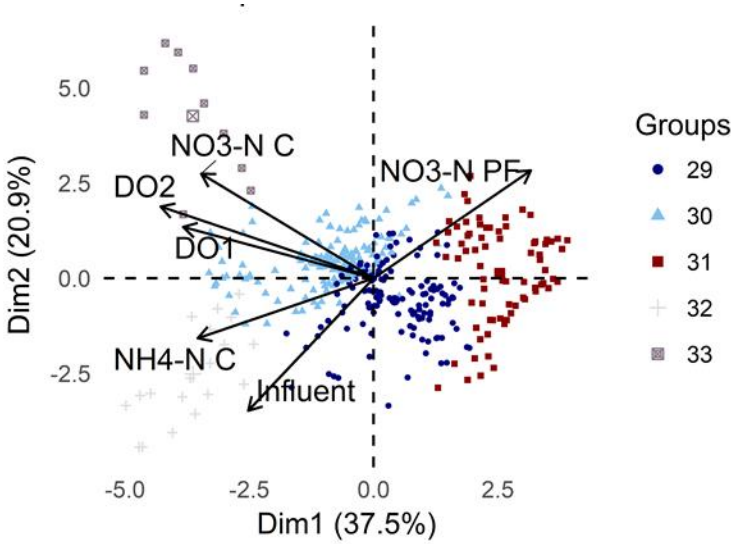


Figure B.29: Score of the first two principal components sub-period 10

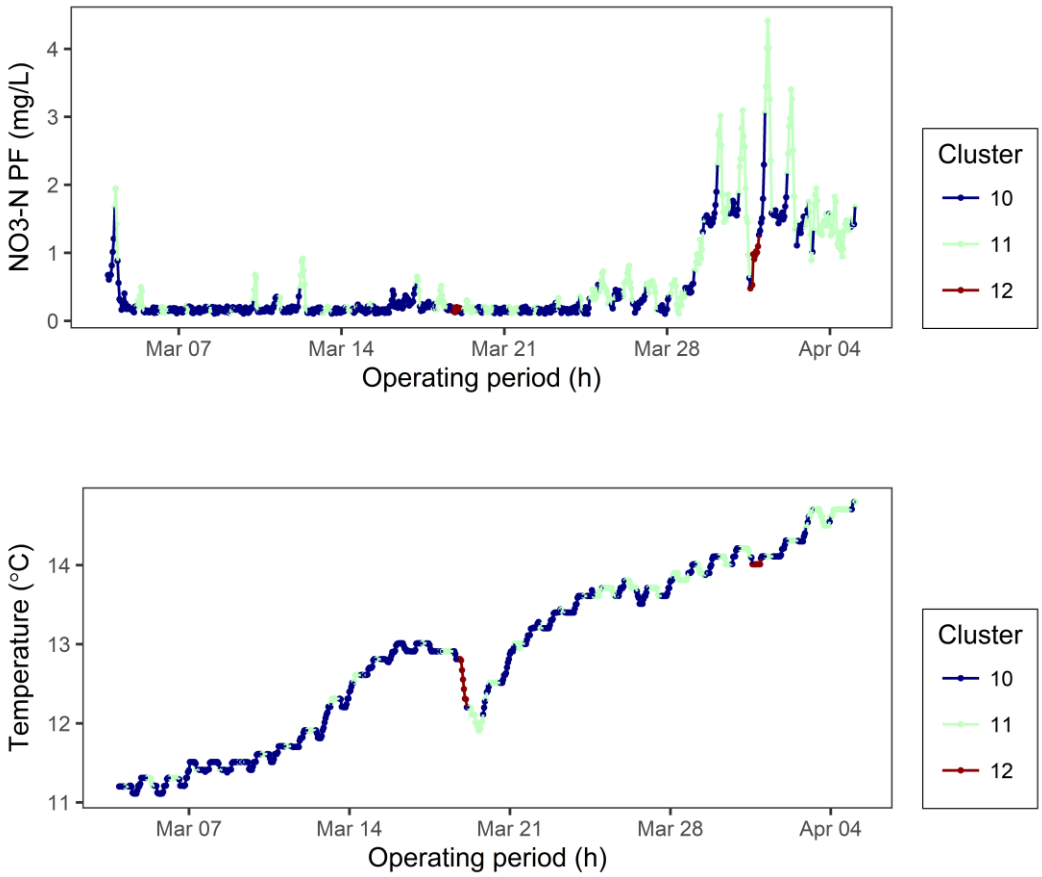


Figure B.30: NO_3-N concentration in the plug-flow reactor and temperature concentration in the Carrousel reactor in sub-period 4

References

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

Kassambara, A., Mundt, F., 2017. factoextra: Extract and Visualize the Results of Multivariate Data Analyses.

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

Killick, R., Haynes, K., Eckley, I., Fearnhead, P., Lee, J., 2016. changepoint: Methods for Changepoint Detection.

Moritz, S., 2017. imputeTS: Time Series Missing Value Imputation.

Ryan, J.A., Ulrich, J.M., Bennett, R., 2017. xts: eXtensible Time Series.

Wei, T., Simko, V., Levy, M., Xie, Y., Jin, Y., Zemla, J., 2017. corrplot: Visualization of a Correlation Matrix.

Wickham, H., 2009. ggplot2: Elegant Graphics for Data Analysis Springer-Verlag. N. Y.

Wickham, H., Chang, W., 2016. ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics.

Appendix C

Table C.1: Features extracted and used in the analysis

Features	Symbol	Variables
Average	μ_i	NH_4-N_{PF}, NO_3-N_{PF}, Influent flow-rate, NH_4-N_C, NO_3-N_C, DO_1, DO_2, DO_3
Standard deviation	σ_i	NH_4-N_{PF}, NO_3-N_{PF}, Influent flow-rate, NH_4-N_C, NO_3-N_C, DO_1, DO_2, DO_3
Max	max_i	NH_4-N_{PF}, NO_3-N_{PF}, Influent flow-rate, NH_4-N_C, NO_3-N_C, DO_1, DO_2, DO_3
Min	min_i	NH_4-N_{PF}, NO_3-N_{PF}, Influent flow-rate, NH_4-N_C, NO_3-N_C, DO_1, DO_2, DO_3
Skewness	sk_i	NH_4-N_{PF}, NO_3-N_{PF}, Influent flow-rate, NH_4-N_C, NO_3-N_C, DO_1, DO_2, DO_3
Kurtosis	k_i	NH_4-N_{PF}, NO_3-N_{PF}, Influent flow-rate, NH_4-N_C, NO_3-N_C, DO_1, DO_2, DO_3

Interquantile range (IQR)	iqr_i	NH_4-N_{PF}, NO_3-N_{PF}, Influent flow-rate, NH_4-N_C, NO_3-N_C, DO_1, DO_2, DO_3
Second-order average	μ_{2i}	NH_4-N_{PF}, NO_3-N_{PF}, Influent flow-rate, NH_4-N_C, NO_3-N_C, DO_1, DO_2, DO_3
Second-order standard deviation	σ_{2i}	NH_4-N_{PF}, NO_3-N_{PF}, Influent flow-rate, NH_4-N_C, NO_3-N_C, DO_1, DO_2, DO_3
*Average concentration of variables for the periods with NH_4-N C is higher than 1.2 mg/L	$\mu_i\ when\ \mathrm{NH4-N_C} > 1.2\ \mathrm{mg/L}$	NH_4-N_{PF}, NO_3-N_{PF}, Influent flow-rate, NO_3-N_C, DO_1, DO_2, DO_3
*Average concentration of variables for the periods with NH_4-N C is lower than 0.9 mg/L	$\mu_i\ when\ \mathrm{NH4-N_C} < 0.9\ \mathrm{mg/L}$	NH_4-N_{PF}, NO_3-N_{PF}, Influent flow-rate, NO_3-N_C, DO_1, DO_2, DO_3
**Average concentration of variables for the periods with NH_4-N PF is lower than 7 mg/L	$\mu_i\ when\ \mathrm{NH4-N_{PF}} < 7\ \mathrm{mg/L}$	NH_4-N_C, NO_3-N_{PF}, Influent flow-rate, NO_3-N_C, DO_1, DO_2, DO_3
**Average concentration of variables for the periods with NH_4-N PF is higher than 15 mg/L	$\mu_i\ when\ \mathrm{NH4-N_{PF}} > 15\ \mathrm{mg/L}$	NH_4-N_C, NO_3-N_{PF}, Influent flow-rate, NO_3-N_C, DO_1, DO_2, DO_3
***(Average NO_3-N C concentration for hours NH_4-N C is decreasing)- (Average of NO_3-N C concentration for hours NH_4-N C is increasing)	$\left(\mu_i when \frac{d}{dt}(\mathrm{NH4-N_C}) < 0\right) - \left(\mu_i when \frac{d}{dt}(\mathrm{NH4-N_C}) > 0\right)$	NO_3-N_C

***(Average DO concentration for hours NH_4-N C is decreasing - Average of DO concentration for hours NH_4-N C is increasing)	$\left(\mu_i when \frac{d}{dt}(\mathrm{NH4-N_C}) < 0\right) - \left(\mu_i when \frac{d}{dt}(\mathrm{NH4-N_C}) > 0\right)$	DO_1, DO_2, DO_3
Local maxima and minima	loc_max_i, loc_min_i	NH_4-N_C, NO_3-N_{PF}, Influent flow-rate, NO_3-N_C
CEEMDAN trend	Figure C.1	NH4-PF, Temperature

*Features based operational control of the system. In the system, aerator 1 operates under on/off pattern (when ammonium is higher than 1.2 mg/l), while aerators 2 and 3 operate always and peak when ammonium is higher than 0.6 and 0.9 mg/l, respectively.

**Features based aiming to investigate the range of operational variables at higher and lower than average NH_4-N concentrations in the plug-flow reactor

***Features based aiming to investigate the response of operational variables when NH_4-N concentration in the Carrousel effluent is increasing and decreasing

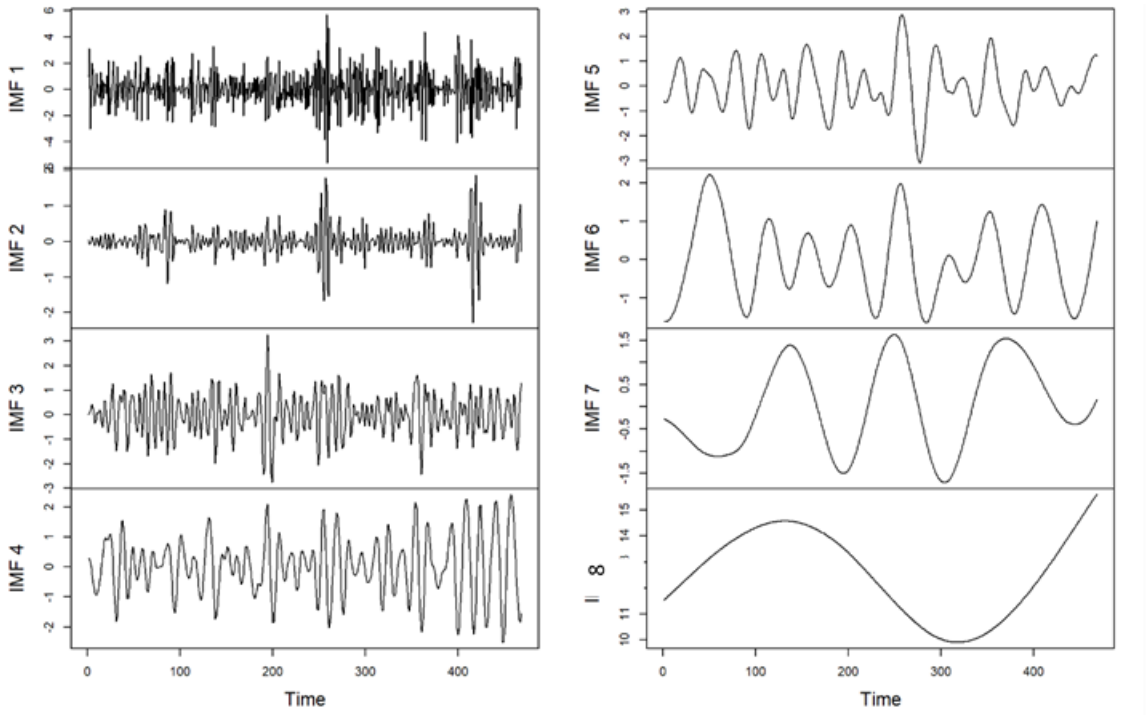


Figure C.1: IMFs from the CEEMDAN analysis and for the NH_4-N concentration in the plug-flow reactor. In order to extract the long-term trend of the NH_4-N concentration in the plug-flow reactor IMFs with oscillatory periods more than 90 days were used (IMF 9-10)

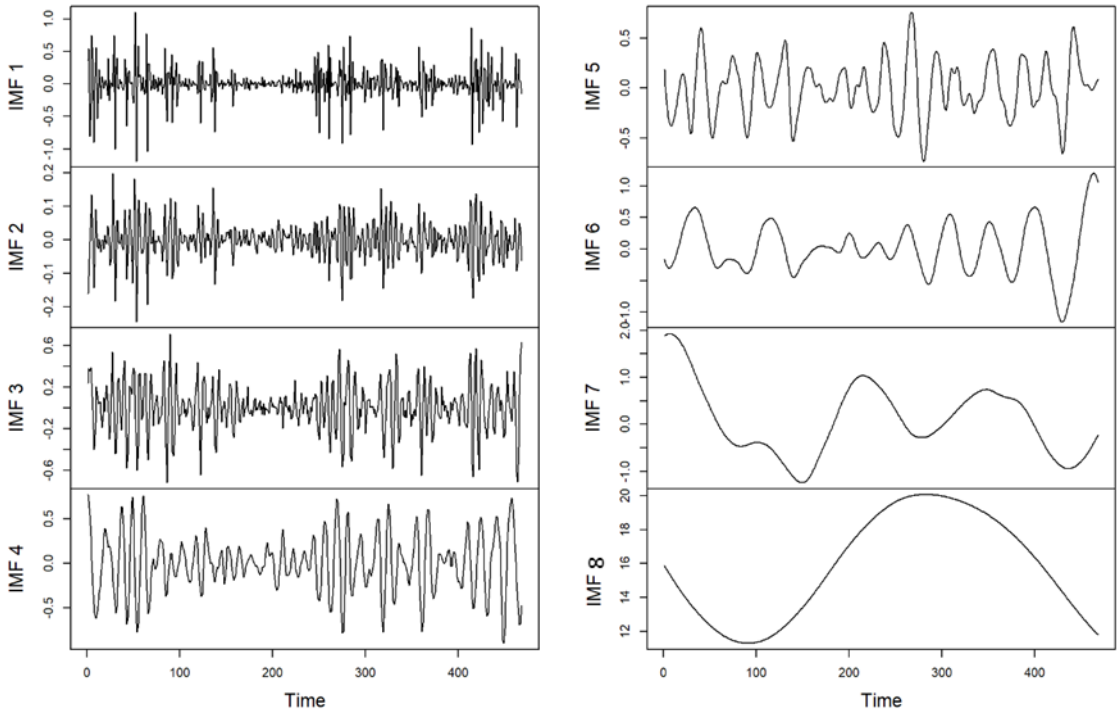


Figure C.2: IMFs from the CEEMDAN analysis and for the Temperature concentration in the plug-flow reactor. In order to extract the long-term trend of the Temperature concentration in the plug-flow reactor IMF 8 was used

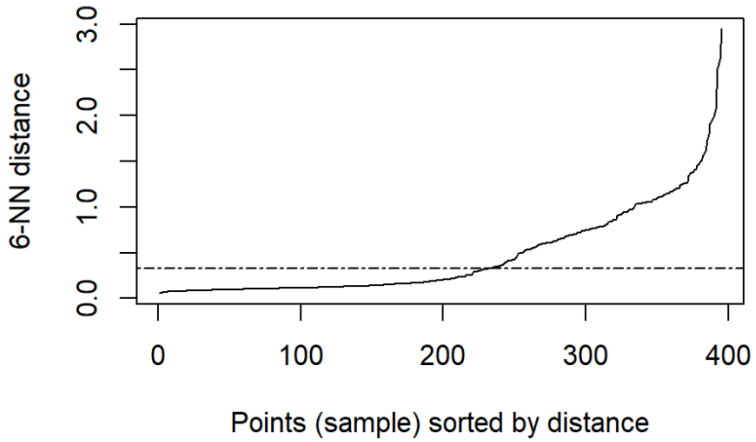


Figure C.3: k-nearest neighbors of the data used in DBSCAN plotted in increasing order. The eps parameter was determined based on the “knee” in the plot (horizontal dotted line)