Brunel University London

NAXOS 2018

Prediction of wastewater N2O emissions using artificial Neural Networks.

Vasilaki V., Mattias T., Angadi V. C., Sousa P., Mousavi A., Katsou E.

15/5/2018

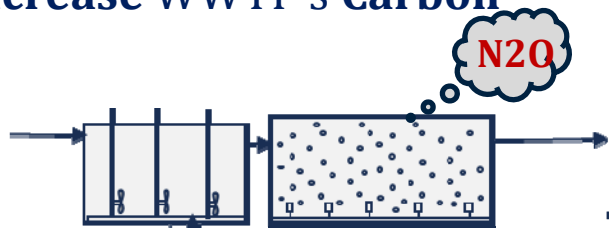
Overview

- Introduction
- Aim and objectives
- Methodology
- Results
  - Changepoint detection results
  - Spearman's rank correlation analysis
  - Hierarchical k-means clustering results
  - Principal Component analysis results
  - Outliers detection - DBSCAN
  - SVM and N2O neural network model results
- Conclusions

Introduction

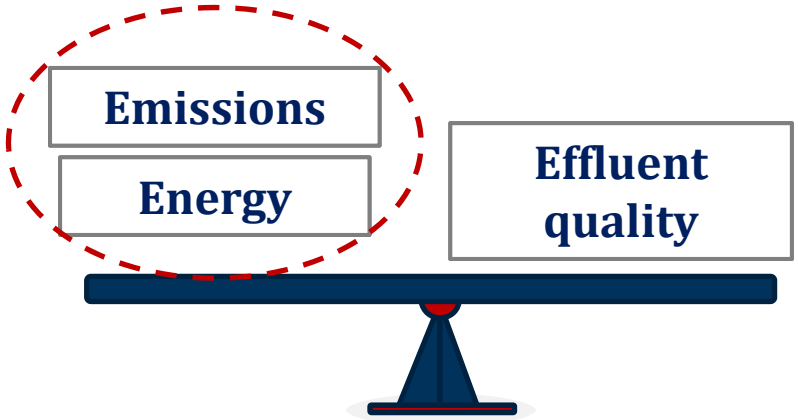
Wastewater treatment **design** and **operation** ➡ **outdated engineering guidelines** from the early 20th century [1]

- **EU 3**% of generated **electricity** ➡ **water industry** [2]
- Solely **N2O** emissions ➡ **60%**[3], or up to **78%**[4] **increase** WWTP's **Carbon Footprint.**



A new sustainable perception of wastewater treatment.

- Need to include GHG emissions/energy consumption in operational strategies ➡ sustainability [5]
- Limited studies [6, 7]



Introduction

Clustering, **artificial neural networks**, **decision trees and classifiers** have been used in WWTPs to:

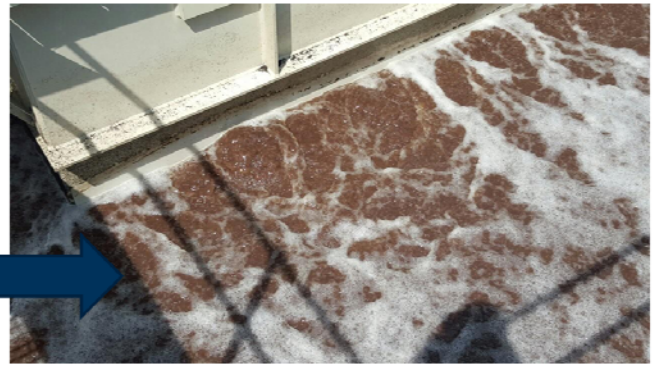
(i) improve process monitoring [8] and provide **insights** [9]

(ii)identify and isolate **process faults** [10] and **sensor faults** [11]

(iii) **predict** significant **operating variables** [12]

However…

- **Few data-driven monitoring** approaches in **full-scale applications**[13]
- **Statistical analysis** is **seldom done** [14]
- **Little guidance** for selection of the **most appropriate** AI **method**[15]



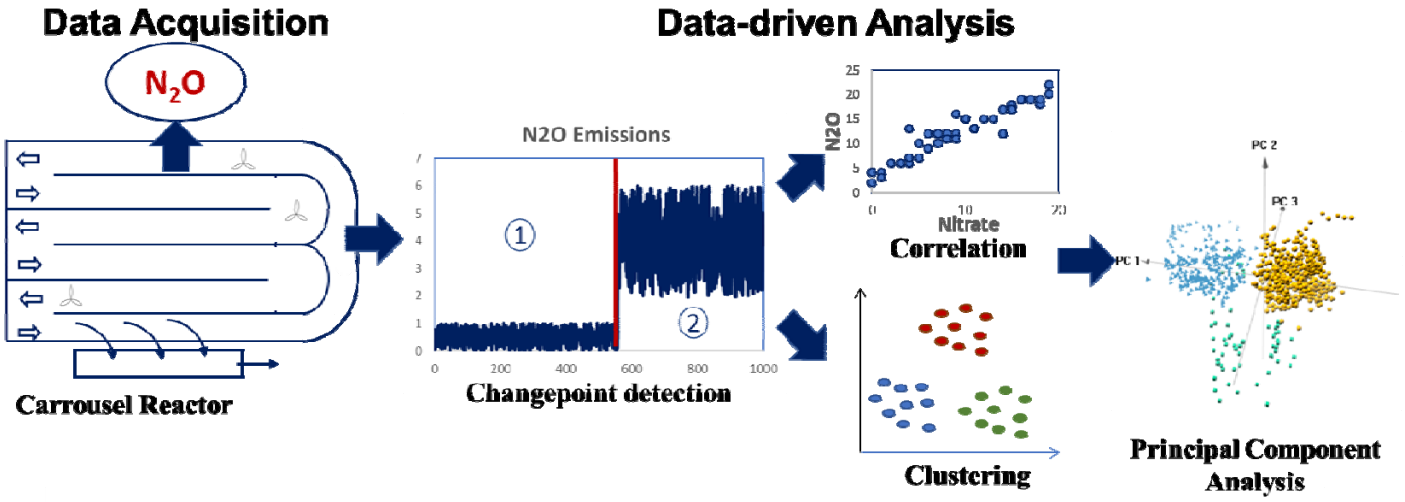
Minworth STW

Aim and objectives

Investigate if **data-driven** methods and **multivariate analysis can** provide **insights** on the **combined effect** of the **operating variables** on **N2O emissions**

WWTP processes are **subject to change**. How can these **changes** be **detected**, and how can they be **considered** in **N2O statistical modelling**?

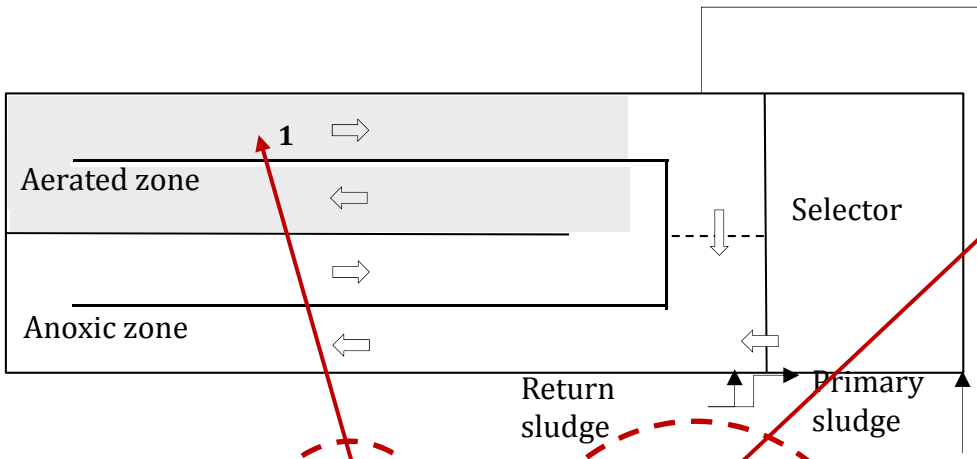
Investigate if **data-driven** methods can be used to predict N2O emissions behavior

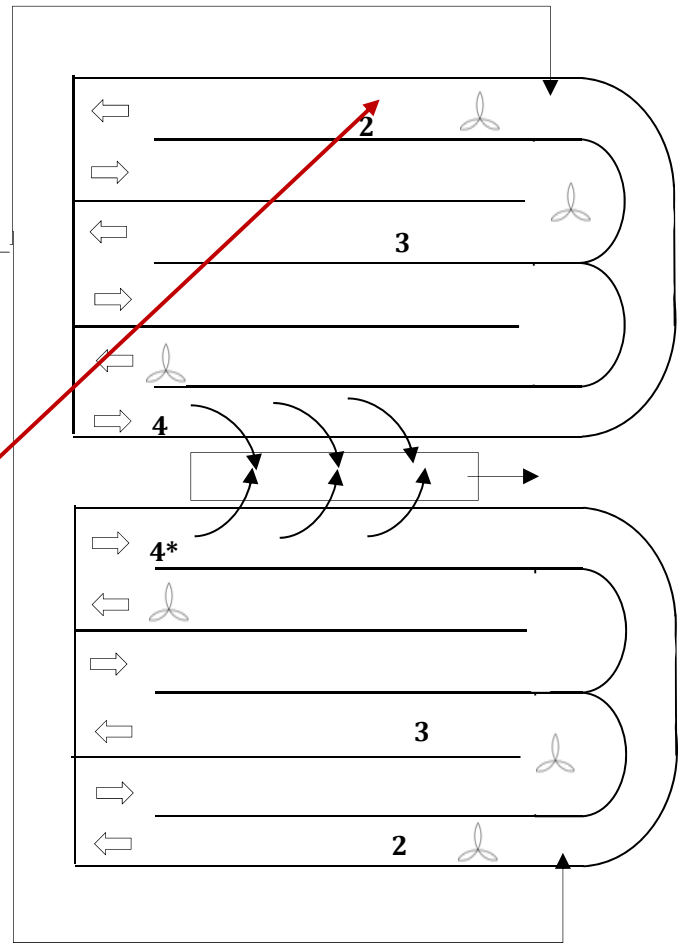


N_2O patterns and dependencies

Methodology

Flow-chart

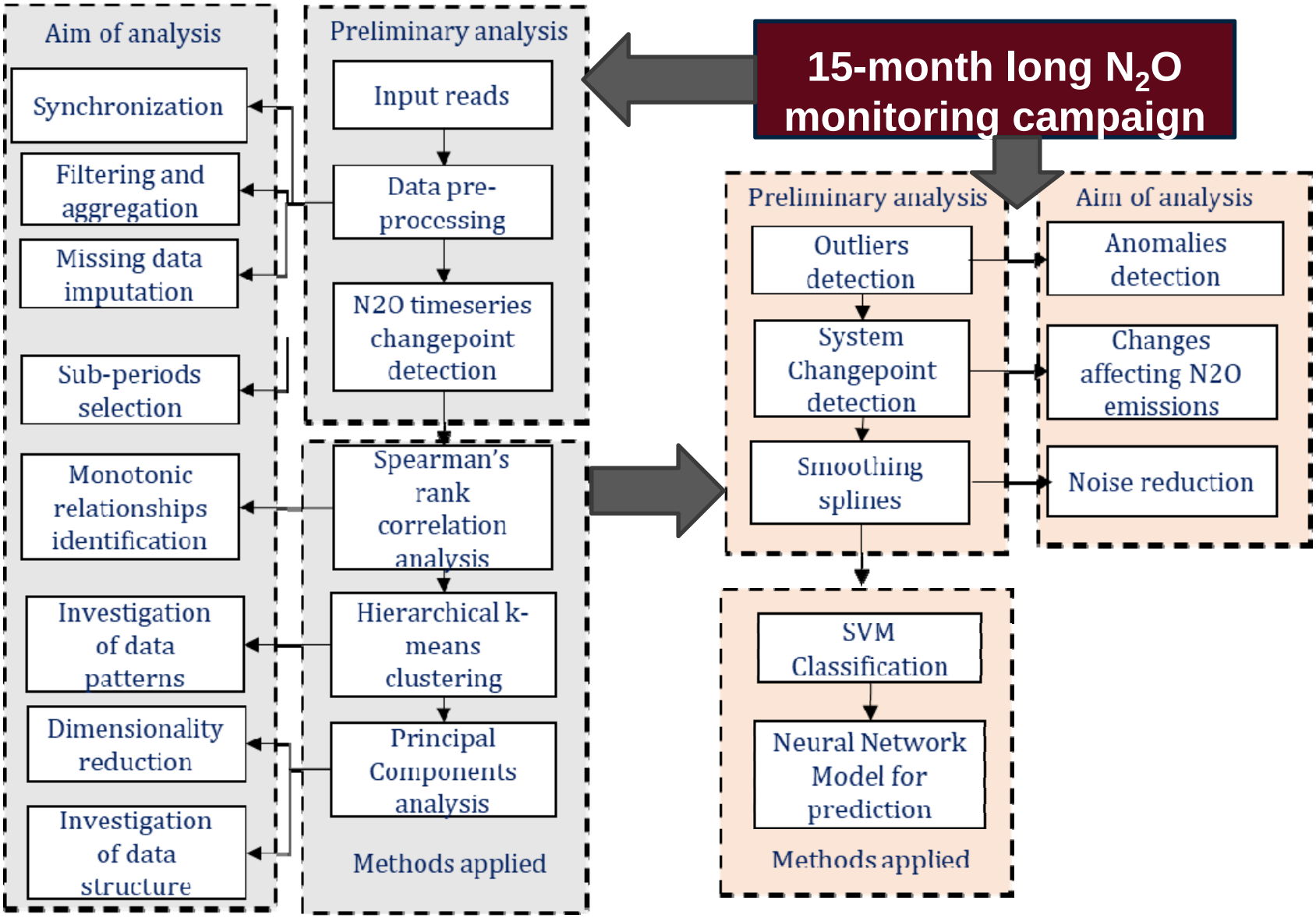




Influent	Plug-flow reactor	Carrousel reactor N
Flow-rate	DO PF	DO1, DO2, DO3
	NH4-N PF	NH4-N C
	NO3-N PF	NO3-N C
	N2O PF	NO2-N C
		N2O C
		Temp C
		TSS C

Kralingseveer WWTP

Methodology

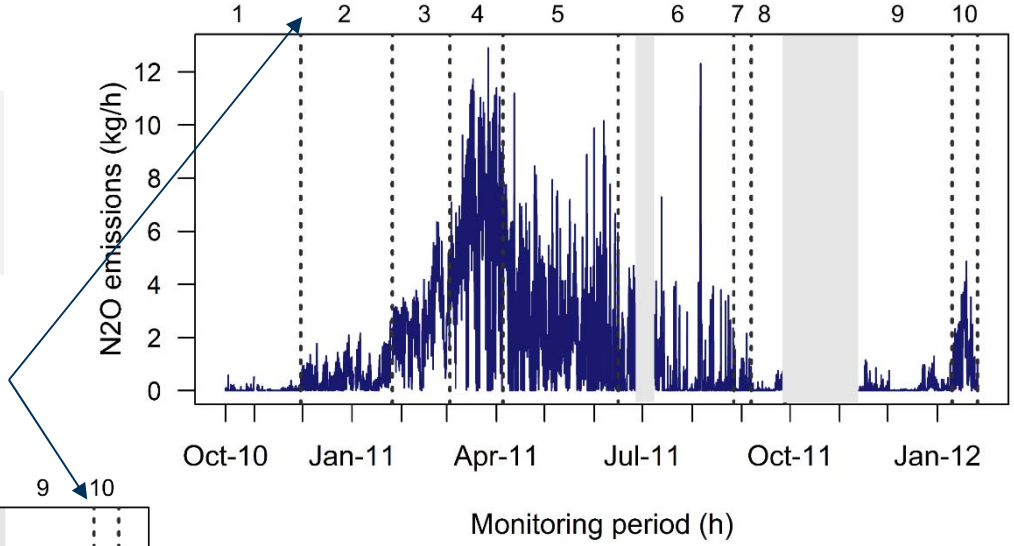


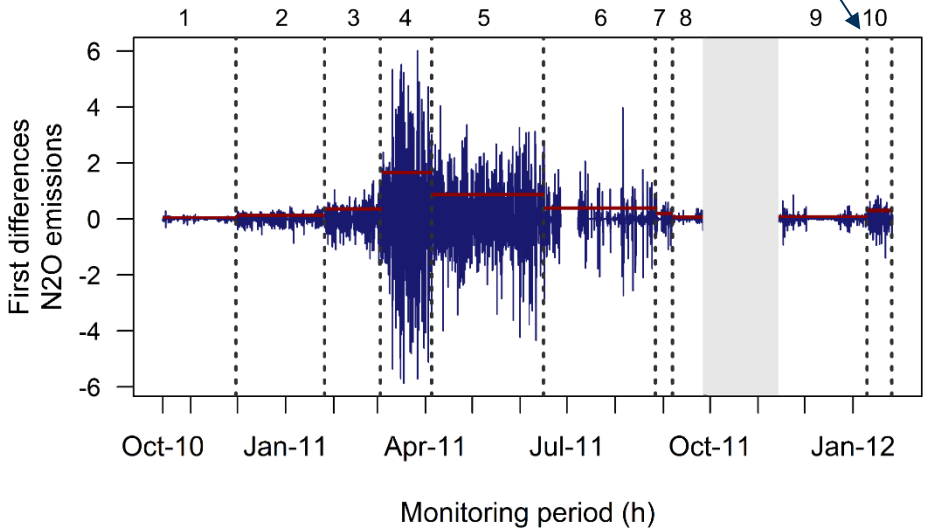
Results

❖Binary segmentation[16] ➡ 10 sub-periods with different N2O emissions profiles

N2O emissions profile in the Northern Carrousel reactor

different sub-periods

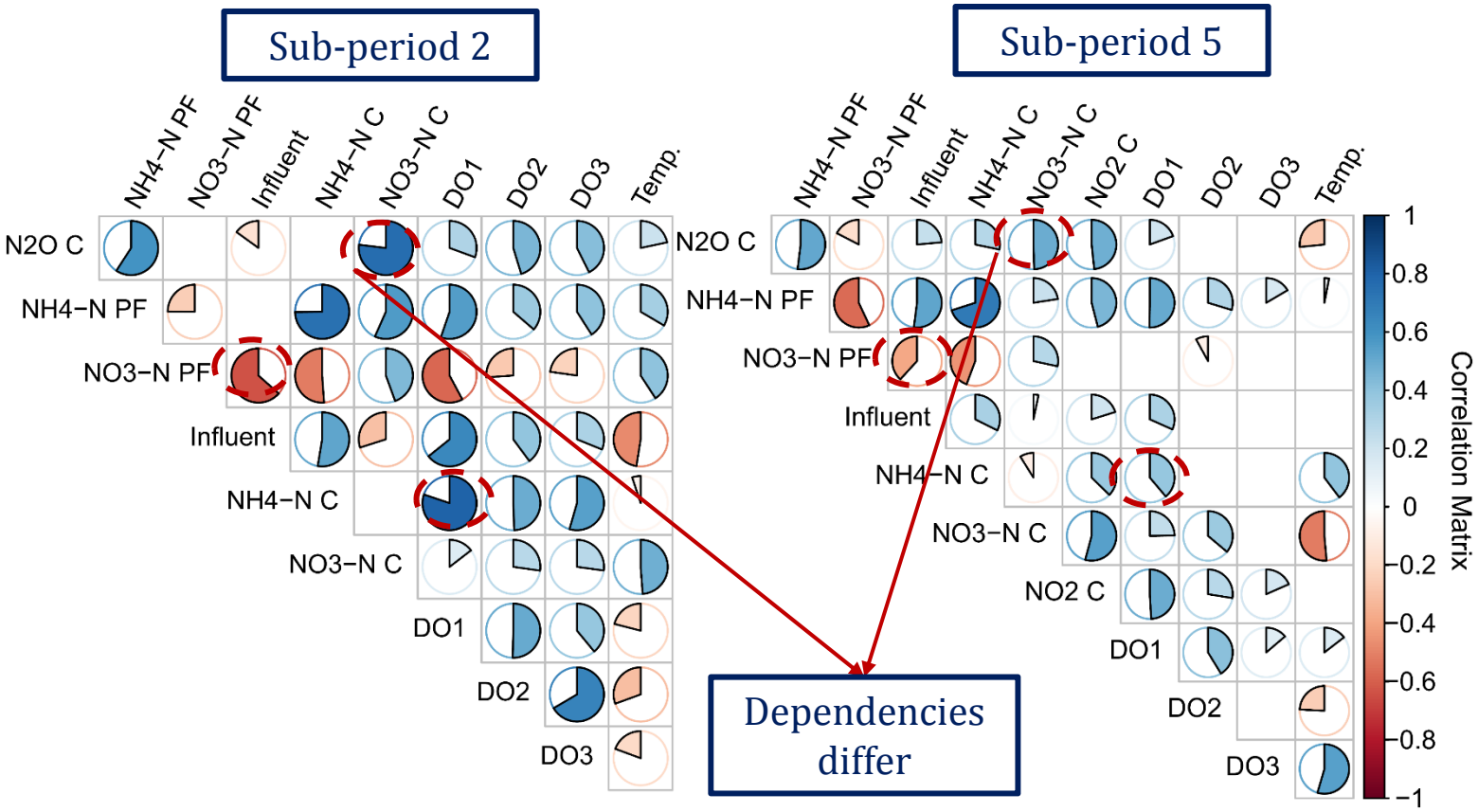




First difference of the N2O emissions timeseries showing the sub-periods

Results

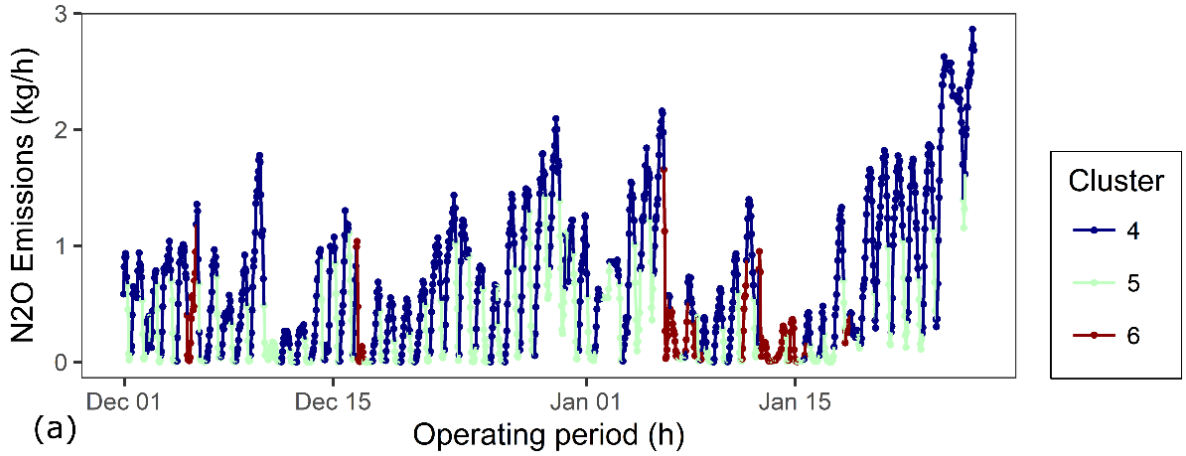
- Spearman's rank correlation [17]
  - Fluctuation between sub-periods
  - N2O correlated with ammonium, nitrate and nitrite
  - Low correlation coefficients can indicate non-monotonic interrelationships



Results

- Hierarchical k-means clustering [18]
  - Reoccurring patterns and their effect on N2O emissions
  - N_2O emission peaks linked with the diurnal behaviour and precipitation events
  - Clusters with NO3-N plug-flow <1 mg/L and Carrousel reactor <4 mg/L
    → N2O fluxes >2 kg/h

N_2O emissions profile



P	Cl	N_2O C	NH_4-N PF	NO_3-N PF	Influent	NH_4-N C	NO_3-N C
		kg/h	mg/l	mg/l	m^3/h	mg/l	mg/l
2	4	0.87	**15.30**	2.05	3827	1.51	8.61
	5	0.21	**9.13**	3.69	3419	0.74	5.28
	6	0.24	12.51	0.81	**11132**	4.52	5.42

Hierarchical k-means clustering results

Results

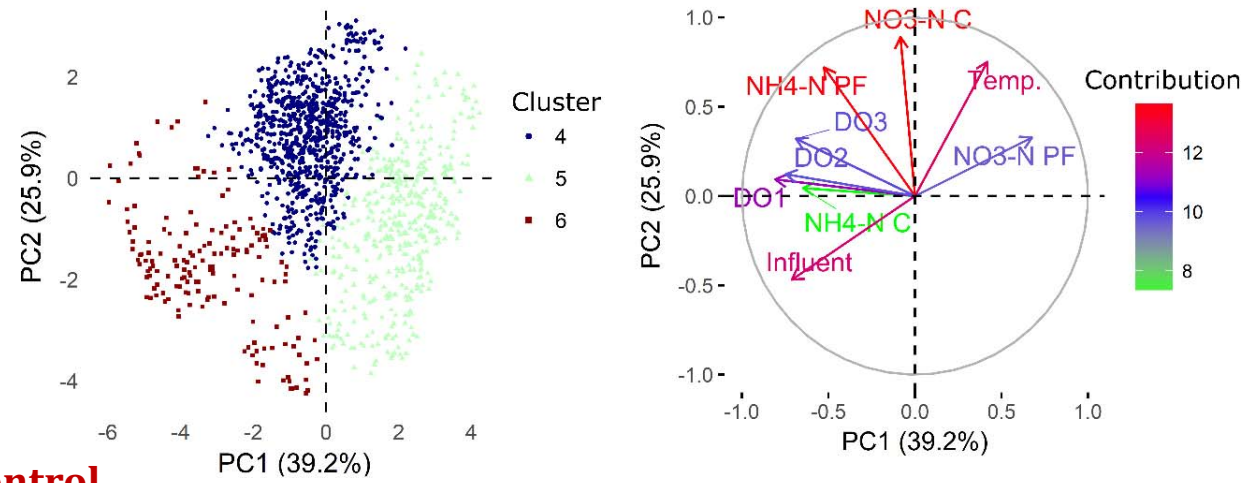
- Principal component analysis[19]
  - Validated the findings from the clustering analysis
  - Ammonium, nitrate, nitrite, influent flow-rate and temperature, explained more than 65% of the variance in the system for the majority of the sub-periods.

PC Loadings

PC2 and N_2O correlation emissions equal to 0.72

Variable	PC1	PC2
NH_4-N PF	-0.28	**0.47**
NO_3-N PF	**0.36**	0.21
Influent	**-0.38**	**-0.31**
NH_4-N C	**-0.34**	0.03
NO_3-N C	-0.04	**0.58**
DO1	**-0.43**	0.06
DO2	**-0.40**	0.08
DO3	**-0.37**	0.21

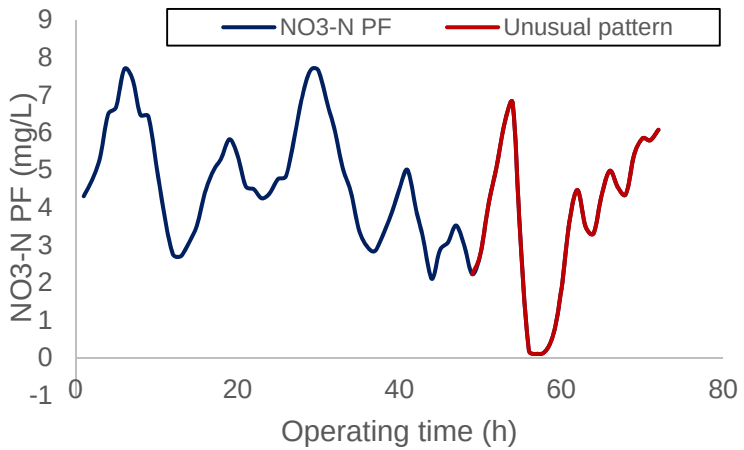
control strategy of the reactor.

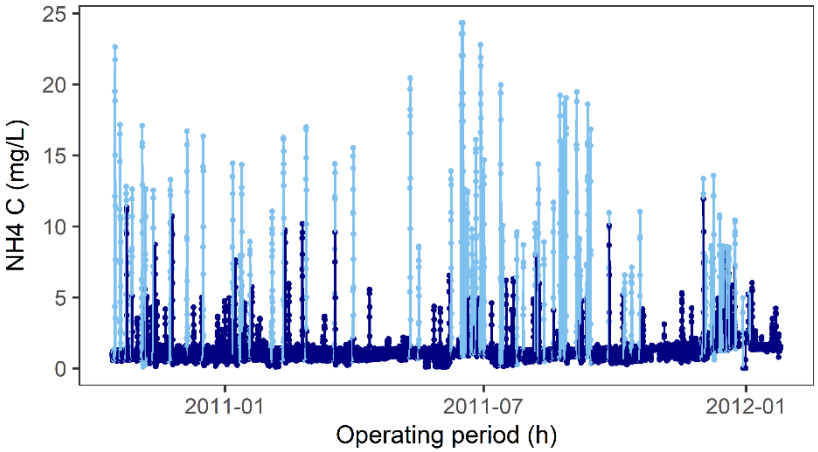


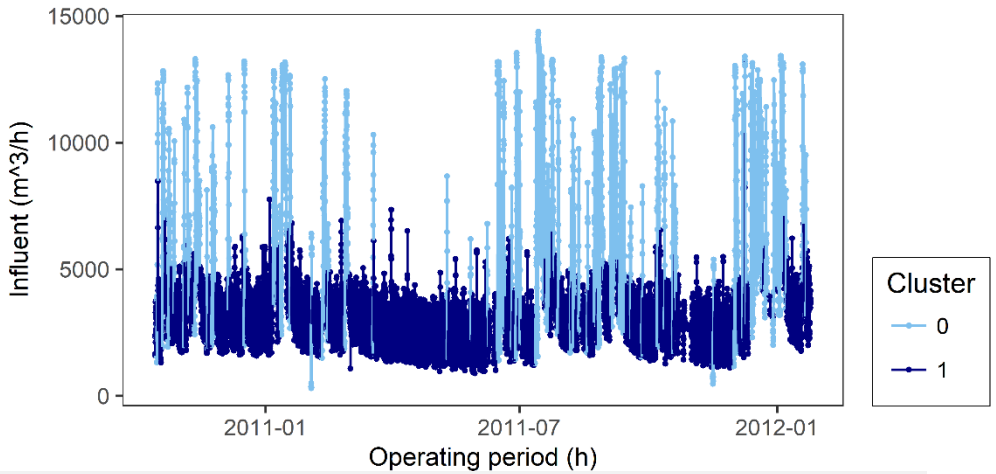
PCA biplot and correlation diagram

Results

- Anomalies detection – DBSCAN clustering [20]
  Identify unexpected patterns in the diurnal profile of the parameters





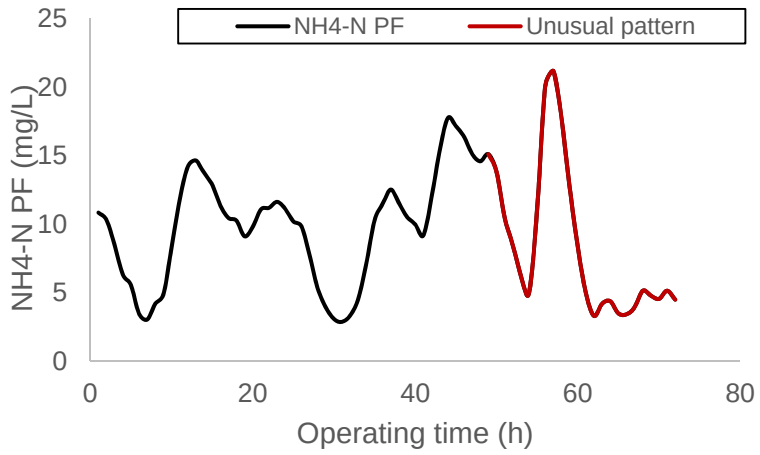


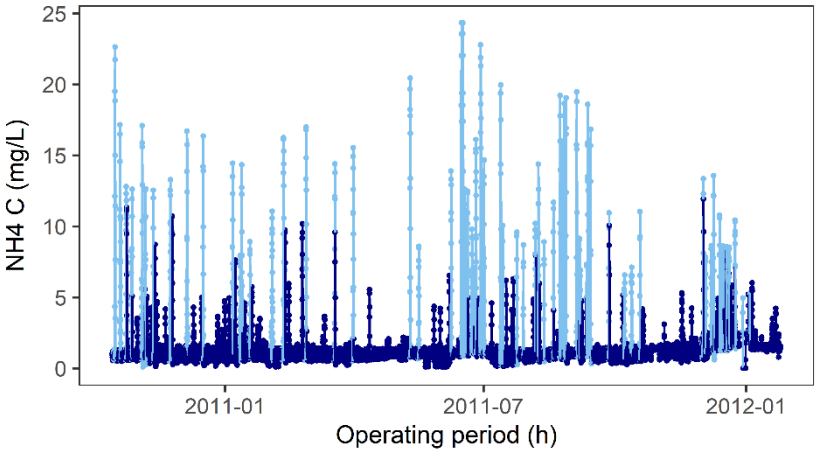
87% common anomalies detected between **NH4-N C** and **Influent** flow-rate

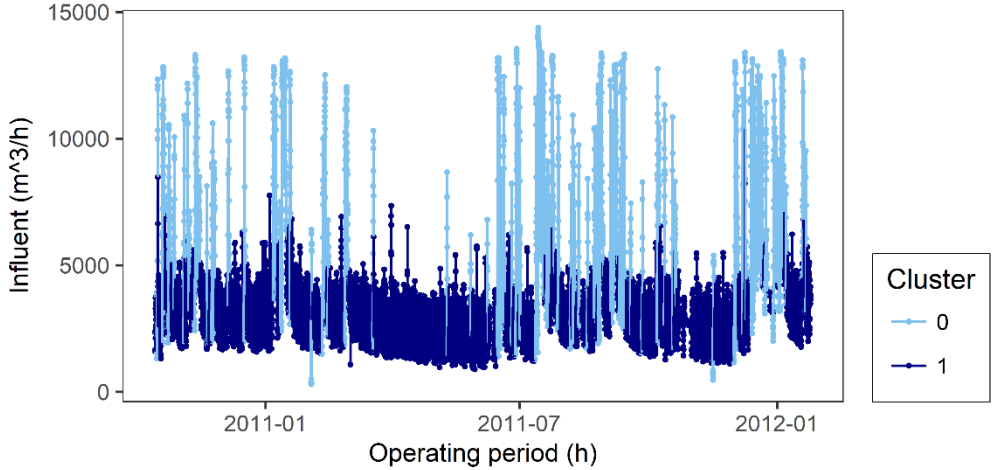
Results

- Anomalies detection – DBSCAN clustering [20]
  Identify unexpected patterns in the diurnal profile of the parameters

  ~80% Common outliers







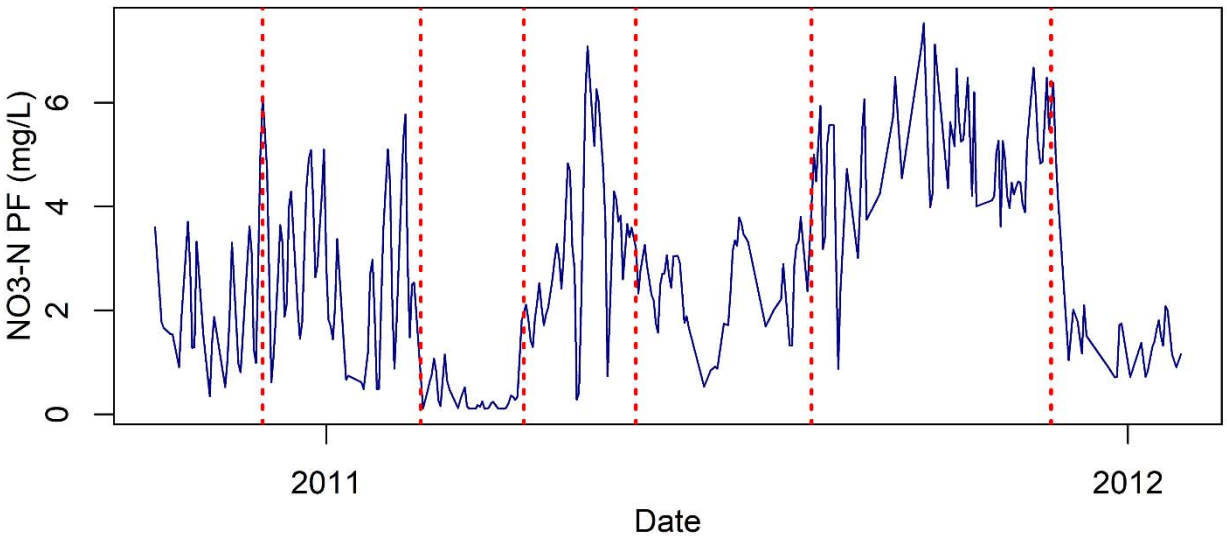
87% common anomalies detected between **NH4-N C** and **Influent** flow-rate

Results

- Sub-period division – NO_3-N PF Changepoint detection

  E-divisive: hierarchical divisive estimation of multiple change points [21]

  Bisection algorithm based on the measurement of divergence between two dataset distributions (nonparametric method).

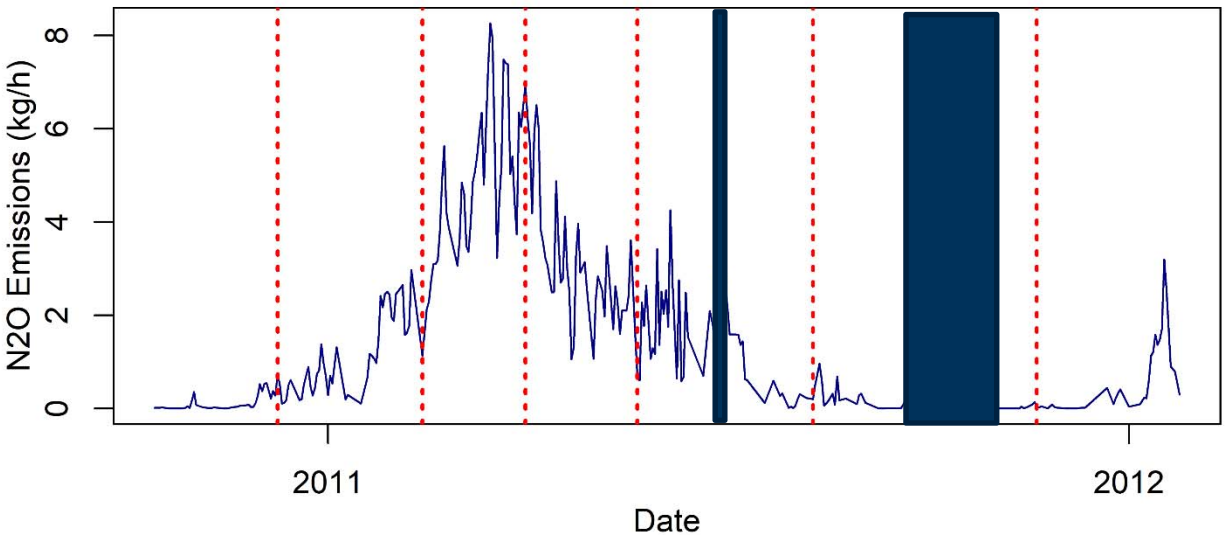


Cl	1	2	3	4	5	6	7
Mean	1.9	**2.8**	0.4	3.1	2.5	5	1.6
Sd	1.6	**2**	0.6	1.9	1.3	2.4	1.3
Median	1.6	**2.7**	0.3	3	2.7	5	1.4

- Sub-period division – NO_3-N PF Changepoint detection

  E-divisive: hierarchical divisive estimation of multiple change points [21]

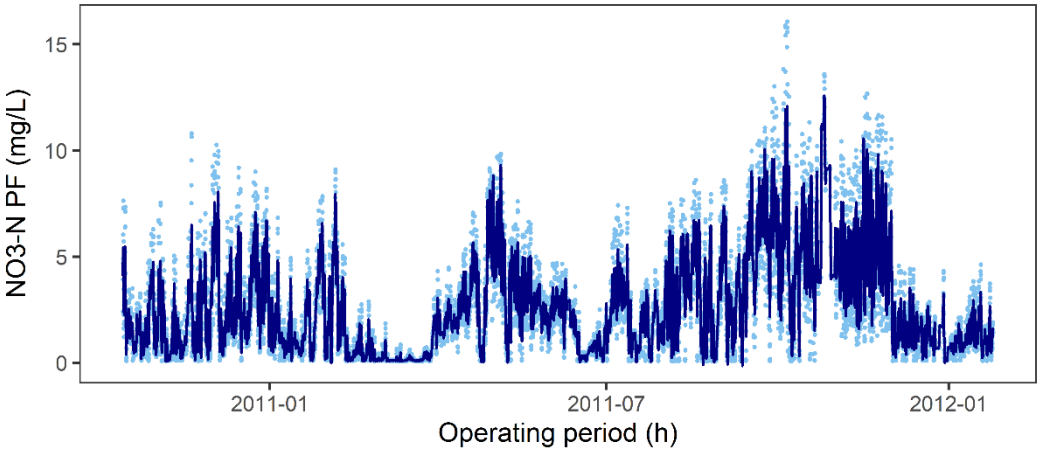
  Bisection algorithm based on the measurement of divergence between two dataset distributions (nonparametric method).



N2O emissions profile sub-periods division

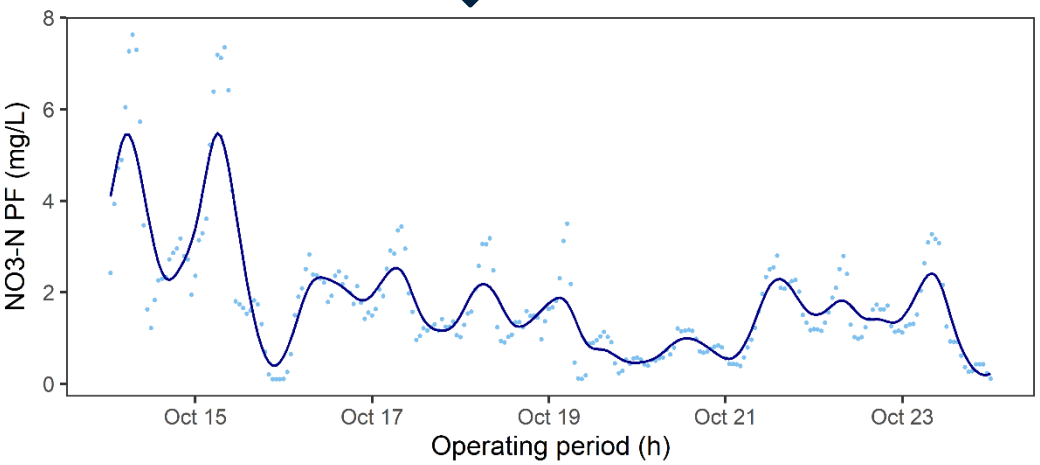
Methodology – Data preprocessing

- Noise reduction – Smoothing splines [22]



The bandwidth of the filtering is as a function of time

Example of NO3-N PF smoothed timeseries



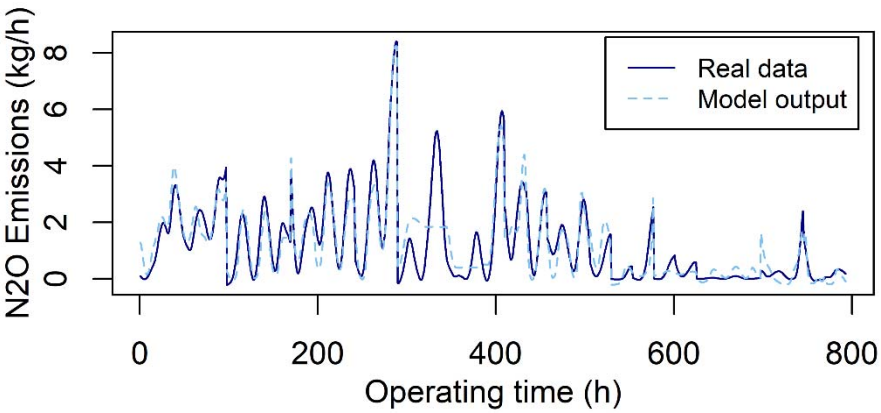
- Data normalization min-max

Results

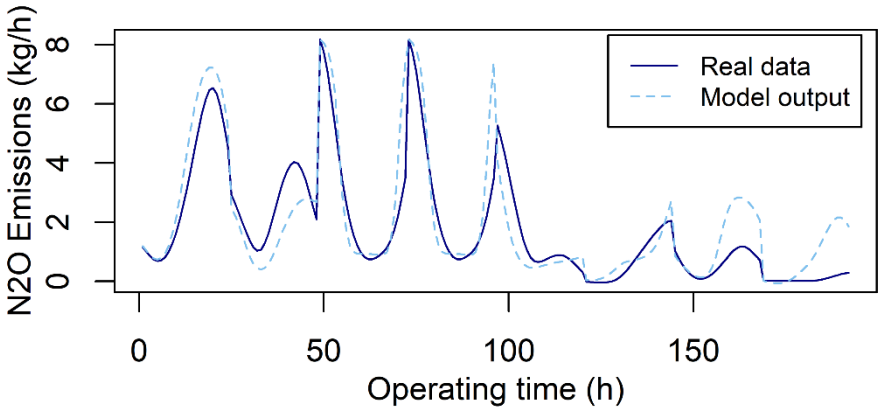
❖ Support Vector machine classification [23]

Method	Data-base	% wrong period
Method 1	SVM Train	0.6%
	SVM Test	5%

❖ Neural Network models [24]



NN model sub-period 5, Train



NN model sub-period 5, Test

Conclusions

- A combination of changepoint detection algorithm, hierarchical k-means clustering and principal component analysis was used to:
  - Detect and visualize disturbances in the system
  - Detect ranges of operating variables that have historically resulted in low or high ranges of N2O emissions
  - Can be used to assist researchers and operators to understand and control the emissions using long term historical data.
- Spearman's rank correlation analysis:
  - showed significant univariate correlations between N2O emissions and ammonium, nitrate and nitrite concentrations.
  - The correlation coefficients fluctuated between the 10 sub-periods.
  - Low values for the correlation coefficients indicated non-monotonic interrelationships that Spearman's rank correlation cannot identify.
- Hierarchical k-means clustering:
  - Provided information on the existence of reoccurring patterns and their effect on N2O emissions.
  - N2O emission peaks were linked with the diurnal behavior of the nutrients' concentrations, with rain events and low nitrate concentrations in the preceding plug flow reactor

Conclusions

- Principal component analysis:
  - validated the findings from the clustering analysis and showed that ammonium, nitrate, nitrite, influent flow-rate and temperature, explained more than 65% of the variance in the system for the majority of the sub-periods.
  - The first principal component corresponded to the control strategy of the reactor.
- DBSCAN :
  - Isolated unusual patterns in the parameters
  - Confirmed that Precipitation events are linked with high NH4-N concentration in the Carrousel effluent
- SVM classification and neural network model:
  - SVM test data classification error ranged between 3-10%.
  - NN model could predict the profile of N2O emissions for sub-periods 1, 2, 3, 4, 5 and 7

References

(1) Daigger, G.T., 2009. Evolving urban water and residuals management paradigms: Water reclamation and reuse, decentralization, and resource recovery. Water Environ. Res. *81*, 809–823.

(2) Brandt, M.J., Middleton, R.A., and Wang, S., 2012. Energy efficiency in the water industry: a compendium of best practices and case studies-global report. (Iwa Publishing).

(3) Rodríguez-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.

(4) Daelman, M. R. J., van Voorthuizen, E. M., Van Dongen, L. G. J. M., Volcke, E. I. P., & Van Loosdrecht, M. C. M., 2013. Methane and nitrous oxide emissions from municipal wastewater treatment–results from a long-term study. *Water Science and Technology*, *67*(10), 2350-2355.

(5) Flores-Alsina, X., Arnell, M., Amerlinck, Y., Corominas, L., Gernaey, K.V., Guo, L., Lindblom, E., Nopens, I., Porro, J., Shaw, A. and Snip, L., 2014. Balancing effluent quality, economic cost and greenhouse gas emissions during the evaluation of (plant-wide) control/operational strategies in WWTPs. *Science of the Total Environment*, *466*, pp.616-624.

(6) Corominas, L., Flores-Alsina, X., Snip, L. and Vanrolleghem, P.A., 2012. Comparison of different modeling approaches to better evaluate greenhouse gas emissions from whole wastewater treatment plants. *Biotechnology and Bioengineering*, *109*(11), pp.2854-2863.

(7) Guo, L., Porro, J., Sharma, K.R., Amerlinck, Y., Benedetti, L., Nopens, I., Shaw, A., Van Hulle, S.W.H., Yuan, Z. and Vanrolleghem, P.A., 2012. Towards a benchmarking tool for minimizing wastewater utility greenhouse gas footprints. *Water Science and Technology*, *66*(11), pp.2483-2495.

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

(9) Moon, T.S., Kim, Y.J., Kim, J.R., Cha, J.H., Kim, D.H., and Kim, C.W., 2009. Identification of process operating state with operational map in municipal wastewater treatment plant. J. Environ. Manage. *90*, 772–778

(10) Haimi, H., Mulas, M., Corona, F., Marsili-Libelli, S., Lindell, P., Heinonen, M., and 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.

(11) Lee, C., Choi, S.W., and Lee, I.-B., 2004. Sensor fault identification based on time-lagged PCA in dynamic processes. Chemom. Intell. Lab. Syst. *70*, 165–178.

References

(12) Rustum, R., Adeloye, A.J., and Scholz, M., 2008. Applying Kohonen Self-Organizing Map as a Software Sensor to Predict Biochemical Oxygen Demand. Water Environ. Res. 80, 32–40.

(13) Haimi, H., Mulas, M., Corona, F. and Vahala, R., 2013. Data-derived soft-sensors for biological wastewater treatment plants: An overview. *Environmental modelling & software*, *47*, pp.88-107.

(14) Olsson, G., Carlsson, B., Comas, J., Copp, J., Gernaey, K. V., Ingildsen, P. & Steyer, J. P. (2014). Instrumentation, control and automation in wastewater–from London 1973 to Narbonne 2013. *Water Science and Technology*, *69*(7), 1373-1385.

(15) Hadjimichael, A., Comas, J. and 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 Communications*, *29*(6), pp.747-756.

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

(17) Spearman, C., 1904. General Intelligence, Objectively Determined and Measured. Am. J. Psychol. *15*, 201–292.

(18) Arai, K., and Barakbah, A.R., 2007. Hierarchical K-means: an algorithm for centroids initialization for K-means. Rep. Fac. Sci. Eng. *36*, 25–31.7

(19) Jolliffe, I.T., 2002. Principal component analysis and factor analysis. *Principal component analysis*, pp.150-166. Principal Component Analysis. Springer Series in Statistics. Springer, New York, NY

(20) Ester, M., Kriegel, H.P., Sander, J. and Xu, X., 1996, August. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Kdd* (Vol. 96, No. 34, pp. 226-231).

(21) Matteson, D.S. and James, N.A., 2014. A nonparametric approach for multiple change point analysis of multivariate data. Journal of the American Statistical Association, 109(505), pp.334-345.

(22) Craven, P.; Wahba, G. (1978). Smoothing noisy data with spline functions. Numer. Math., 31, 377–403.

(23) Cortes, C., and Vapnik, V., 1995. Support-vector networks. Mach. Learn. *20*, 273–297

(24) Bishop, C.M., 1995. Neural networks for pattern recognition (Oxford university press).

Acknowledgements

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

Brunel University London

Thank you for your attention!