Prediction of wastewater N2O emissions using artificial Neural Networks.

Vasilaki V., Mattias T., Angadi V. C., Sousa P., Mousavi A., Katsou E.
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Wastewater treatment design and operation ➡ outdated engineering guidelines from the early 20th century (1)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.


- Need to include GHG emissions/energy consumption in operational strategies ➡ sustainability (5)

- Limited studies (6, 7)
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 \(^{(15)}\)
Aim and objectives

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.

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

**N₂O 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

15-month long N$_2$O monitoring campaign

Aim of analysis
- Input reads
- Data pre-processing
- Spearman’s rank correlation analysis
- Hierarchical k-means clustering
- Principal Components analysis

Preliminary analysis
- Outliers detection
- System Changepoint detection
- Smoothing splines
- SVM Classification

Methods applied
- Neural Network Model for prediction

Aim of analysis
- Anomalies detection
- Changes affecting N$_2$O emissions
- Noise reduction

Methods applied
- Synchronization
- Filtering and aggregation
- Missing data imputation
- Sub-periods selection
- Monotonic relationships identification
- Investigation of data patterns
- Dimensionality reduction
- Investigation of data structure
Results

- Binary segmentation\textsuperscript{(16)} \rightarrow \text{10 sub-periods with different N2O emissions profiles}

N2O emissions profile in the Northern Carrousel reactor

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

- Spearman’s rank correlation\(^{(17)}\)
  - Fluctuation between sub-periods
  - N\(_2\)O correlated with ammonium, nitrate and nitrite
  - Low correlation coefficients can indicate non-monotonic interrelationships

Dependencies differ

Sub-period 2

Sub-period 5

Correlation Matrix
Results

- Hierarchical k-means clustering \(^{(18)}\)
  - Reoccurring patterns and their effect on N2O emissions
  - N\(_2\)O 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

Hierarchical k-means clustering results

<table>
<thead>
<tr>
<th></th>
<th>Cl</th>
<th>N(_2)O C</th>
<th>NH(_4)-N PF</th>
<th>NO(_3)-N PF</th>
<th>Influent</th>
<th>NH(_4)-N C</th>
<th>NO(_3)-N C</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>kg/h</td>
<td>mg/l</td>
<td>mg/l</td>
<td>m(^3)/h</td>
<td>mg/l</td>
<td>mg/l</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>0.87</td>
<td>15.30</td>
<td>2.05</td>
<td>3827</td>
<td>1.51</td>
<td>8.61</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.21</td>
<td>9.13</td>
<td>3.69</td>
<td>3419</td>
<td>0.74</td>
<td>5.28</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.24</td>
<td>12.51</td>
<td>0.81</td>
<td>11132</td>
<td>4.52</td>
<td>5.42</td>
</tr>
</tbody>
</table>

N\(_2\)O emissions profile
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

<table>
<thead>
<tr>
<th>Variable</th>
<th>PC1</th>
<th>PC2</th>
</tr>
</thead>
<tbody>
<tr>
<td>NH(_4)-N PF</td>
<td>-0.28</td>
<td>0.47</td>
</tr>
<tr>
<td>NO(_3)-N PF</td>
<td>0.36</td>
<td>0.21</td>
</tr>
<tr>
<td>Influent</td>
<td>-0.38</td>
<td>-0.31</td>
</tr>
<tr>
<td>NH(_4)-N C</td>
<td>-0.34</td>
<td>0.03</td>
</tr>
<tr>
<td>NO(_3)-N C</td>
<td>-0.04</td>
<td>0.58</td>
</tr>
<tr>
<td>DO1</td>
<td>-0.43</td>
<td>0.06</td>
</tr>
<tr>
<td>DO2</td>
<td>-0.40</td>
<td>0.08</td>
</tr>
<tr>
<td>DO3</td>
<td>-0.37</td>
<td>0.21</td>
</tr>
</tbody>
</table>

**PC2 and N\(_2\)O correlation emissions equal to 0.72**

- PCA biplot and correlation diagram
  - Control strategy of the reactor.
Anomalies detection – DBSCAN clustering \(^{(20)}\)
Identify unexpected patterns in the diurnal profile of the parameters

\[ \text{Results} \]

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).

![Graph showing NO$_3$-N PF levels over time with change points indicated.]

<table>
<thead>
<tr>
<th>Cl</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean</strong></td>
<td>1.9</td>
<td><strong>2.8</strong></td>
<td>0.4</td>
<td>3.1</td>
<td>2.5</td>
<td>5</td>
<td>1.6</td>
</tr>
<tr>
<td><strong>Sd</strong></td>
<td>1.6</td>
<td><strong>2</strong></td>
<td>0.6</td>
<td>1.9</td>
<td>1.3</td>
<td>2.4</td>
<td>1.3</td>
</tr>
<tr>
<td><strong>Median</strong></td>
<td>1.6</td>
<td><strong>2.7</strong></td>
<td>0.3</td>
<td>3</td>
<td>2.7</td>
<td>5</td>
<td>1.4</td>
</tr>
</tbody>
</table>
Sub-period division – NO₃-N PF Changepoint detection

E-divisive: hierarchical divisive estimation of multiple change points

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

- Noise reduction – Smoothing splines

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)

<table>
<thead>
<tr>
<th>Method</th>
<th>Data-base</th>
<th>% wrong period</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method 1</td>
<td>SVM Train</td>
<td>0.6%</td>
</tr>
<tr>
<td></td>
<td>SVM Test</td>
<td>5%</td>
</tr>
</tbody>
</table>

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


References


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.
Thank you for your attention!