ICA2025OSLO
14th IWA Specialised Conference on
Instrumentation, Control and Automation

# Exploration of clustering methods for the Hias Process operation

T.M. Komulainen*, A. Bommo*, K.M. Jansen***, A. Keprate*

*Department of Mechanical, Electrical and Chemical Engineering, Oslo Metropolitan University, Pilestredet 35, 0131 Oslo, Norway tiina.komulainen@oslomet.no , s348854@oslomet.no, arvindke@oslomet.no

**Hias IKS, Sandvikavegen 136, 2312, Ottestad Norway, Katrine.Marsteng.Jansen@hias.no

**Abstract:** In municipal water resource recovery facilities (WRRF), the inlet flowrate and nutrient composition varies with daily, weekly and seasonal patterns. The variation in the flowrate affects the dilution of the nutrients at inlet, and changes the hydraulic retention time in the process, and thereby the biological nutrient removal performance. The aim of process control is to meet the legislative nutrient removal targets during varying influent flowrates and compositions. Clustering and visualization of the current influent condition and the most successful control strategy for each influent condition can provide important information to process operators to adjust the aeration rates.

In this article, we analyse the operational states of the Hias Process using clustering methods. The data set of two years was pre-processed and dynamically time-lagged, and the data set was the labelled according to the phosphorus removal performance. Data exploration and correlation techniques were combined with process knowledge to identify patterns, trends, and relationships within the time-series data. K-means clustering was applied to separate the data into seven distinct clusters based on the influent nutrients, influent flowrate and total aeration rate. Two high nutrient groups were visualized using the phosphorus removal labelling and analysed. While groups of possible process disturbances were spotted, the identification of process states and optimal aeration strategies needs further work. In future work, we will build upon the clusters identified in this work to develop a robust classification model and a decision support tool for operators.

**Keywords**: clustering, machine learning, performance monitoring, wastewater treatment.

## 1. Introduction

In biological wastewater treatment processes, such as the Hias process [1], nutrients in used water are removed, recovered and refined to products, such as fertilizers. At municipal water resource recovery facilities (WRRF), the inlet flowrate and nutrient composition varies with daily, weekly and seasonal patterns, and with precipitation (rain) or snow melting. The variation in the flowrate affects the dilution of the nutrients at inlet, and changes the hydraulic retention time in the process, and thereby the biological nutrient removal performance. The aim of process control is to meet the nutrient removal targets during varying influent flowrates and compositions by optimizing the aeration rates. Visualisation of the current influent condition cluster and the most successful aeration strategy in the data historian for the current condition can provide important information to process operators to manually adjust the aeration rates.

While most of the machine learning applications in the wastewater industry aim to predict the effluent nutrient concentrations, an emerging research topic is clustering and classification of operational states or operating conditions, or operational multi-modes. Unsupervised clustering algorithms have been developed primarily for fault detection in the wastewater industry. Froemelt et al. [2] applied self-organizing maps to identify operational states, and analyze transitions between states to find conditions leading to high N2O emissions in a real large-scale biological process. Chen et al. [3] have compared various clustering methods to detect faults: K-means, connectivity kernel based density peak (CKDP), density-based spatial clustering (DBSCAN) and Density Peak Clustering (DPC) to identify multi-modes and faulty data points, using simulated data from Benchmark Simulation Model No. 1 (BSM1). Li et al. [4] combine self-organizing maps (SOM) and K-means clustering to classify operating conditions prior to data reconciliation using simulated data from Benchmark Simulation Model No. 1 (BSM1). Zhaoqin et al. [5] applied kernel fuzzy c-means clustering for operation optimization in an industrial wastewater treatment facility. SOM and ward linkage clustering was applied by Kern et al. [6] to identify operational patterns and further optimize controller setpoints in simulated data set from Activated Sludge Model No.1

(ASM1). Moon et al. [7] used principal component analysis (PCA) and and K-means clustering to classify the operating states into good and poor performing groups in a large-scale, municipal process. Ebel et al. [8] used fuzzy c-means clustering to identify process state, the state was further used to determine control strategy, tested in Activated Sludge Model No.1 (ASM1).

The Hias Process is a novel, continuous flow, moving bed bioreactor with enhanced biological phosphorus removal and simultaneous nitrification and denitrification (MBBR-EBPR-SND) [1]. In our previous work mechanistic and machine learning models were developed to predict the effluent nutrient concentrations in the Hias Process [9, 10]. We analyzed the effect of the additional instrumentation to the effluent prediction in [11]. We developed virtual sensors to estimate P O4 and COD concentrations at the inlet and zone 3 of the Hias Process [12]. Master students Baqeri and Nermo have developed control strategies for the Hias Process using transfer functions and machine learning methods [13], [13]. In recent master thesis, Bommo has applied K-means, Density-Based Spatial Clustering of Applications with Noise (DBSCAN), Gaussian Mixture Models (GMM), Self- Organizing Maps (SOM) to characterize the operational states in the Hias Process [14].
In this article, we will analyze the Hias process time series dataset by first employing various data exploration and correlation techniques combined with process knowledge to gain a comprehensive understanding of the data. This initial step will include identifying patterns, trends, and relationships within the time series. Following this, we will apply pre-processing methods specifically suited for time series data, including dynamic time-lagging. We use the recommendations of [14] to build clustering of inlet conditions and aeration strategies using K-means algorithm. In subsequent case studies, select the high nutrient clusters, and labeling these according to the achieved nutrient removal performance to study further the effect on the $PO_4$ removal. In future work, we will build upon the clusters identified in this work to develop a robust classification model and a decision support tool for operators.

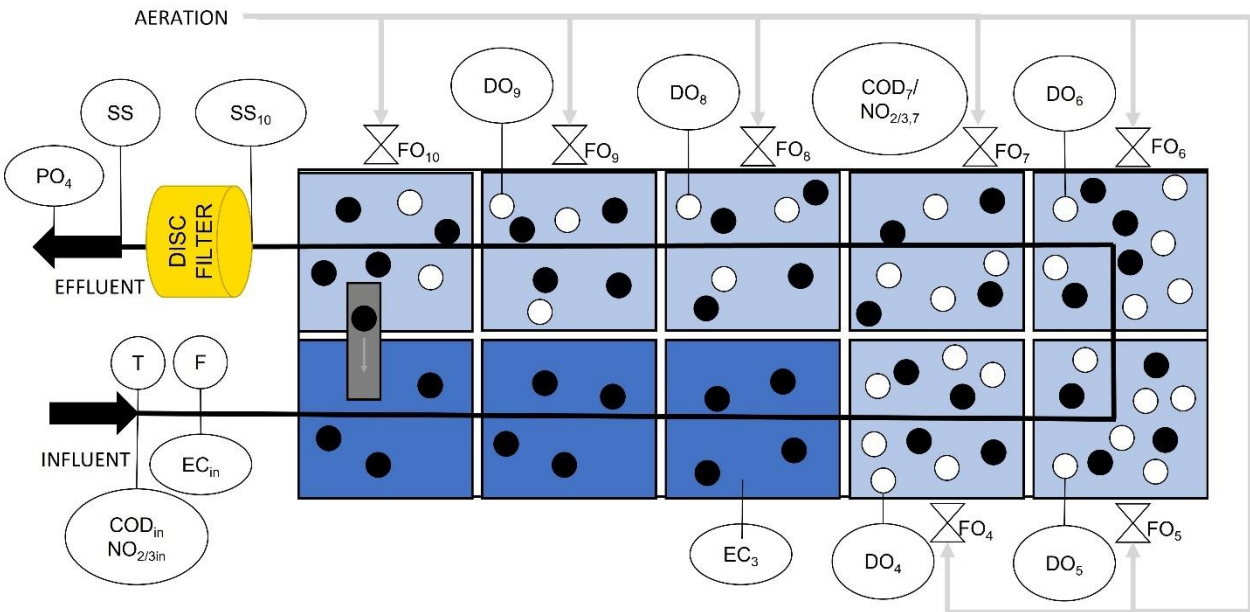


**Figure 1.** The Hias Process with instrumentation.

## 2. Materials and Methods

Hias water resource recovery facility is located in Hamar, Norway and treats used water for municipality and industry around, corresponding to 192 000 PE. The Hias Process is a biological nutrient removal process illustrated in Figure 1. enter the anaerobic zone. The Hias Process consists of three anaerobic and seven aerobic zones, where biomass carriers flow continuously submersed in the used water and remove over 95 % of the phosphorus. The influent water and the biofilm carriers enter the first anaerobic process zone. The three anaerobic zones are mixed to ensure sufficient distribution of biofilm carriers in the water. The following seven zones are aerated to ensure sufficient dissolved oxygen concentrations for aerobic nutrient uptake. Used water and submerged biofilm carriers float through the process with gravity. In the last process zone, the biofilm carriers are transferred to the first process via a conveyor belt. The biomass accumulated in the biological process is removed in a disc filter. The clarified water is disposed to lake Mjøsa, which has nutrient removal requirements of $PO_4$ or under 0.4 mg/L and COD or under 125 mg/L. The Hias Process instrumentation includes continuous online measurements of flowrate, temperature, aeration rates, dissolved oxygen, suspended solids, and nutrient compositions of $PO_4$, COD, $NO_2$ and $NO_3$. Soluble COD, $NO_2$ and $NO_3$ are measured continuously at inlet and zone 7. Suspended solids are measured at zone 10 ($SS_{10}$) and after the disc filter (SS). Effluent $PO_4$ is measured using an online-analyzer. In addition, electrical conductivity sensors are installed at inlet and in zone 3. The Hias Process measurements utilized in this study are listed in Table 1. The online measurement data was collected from July 2023 to January 2025 using sampling time of one minute.

**Table 1.** Online measurements in the Hias Process.

| Symbol | Description | Zone | Unit |
|---|---|---|---|
| F | flowrate | influent | $m^3/h$ |
| T | temperature | influent | ºC |
| $COD_{IN}$ | soluble COD | influent | $g/m^3$ |
| $NO_{2IN}$ | $NO_2$ | influent | $g/m^3$ |
| $NO_{3IN}$ | $NO_3$ | influent | $g/m^3$ |
| $EC_{IN}$ | El.conductivity | influent | mS/cm |
| $PO_{IN}$ | Virtual $PO_4$ | influent | $g/m^3$ |
| $EC_3$ | El.conductivity | 3 | mS/cm |
| $PO_3$ | Virtual $PO_4$ | 3 | $g/m^3$ |
| $COD_7$ | soluble COD | 7 | $g/m^3$ |
| $NO_{27}$ | $NO_2$ | 7 | $g/m^3$ |
| $FO_i$ | Aeration rate | 4,5,6,7,8,9,10 | $m^3/h$ |
| $DO_i$ | Dissolved oxygen | 4,5,6,8,9 | $m^3/h$ |
| SS10 | Suspended solids | 10 | $g/m^3$ |
| $SS_{OUT}$ | Suspended solids | effluent | $g/m^3$ |
| $PO_{OUT}$ | Analyzer $PO_4$ | effluent | $g/m^3$ |

## 3. Results

### 3.1. Data pre-processing

We will begin by assessing data quality, handling missing values, and summarizing the dataset with statistical metrics such as mean, variance, and standard deviation. The dataset consisted of 820519 one-minute samples. First, the missing values were forward-filled with ”previous” values, and then the missing values left at the beginning of the data vectors were backward-filled with ”next” values. The outliers were detected using three standard deviations around the mean value.

The outliers were filled with ”previous” value. Different outlier detection methods were tested, but for example, inter-quartiles method lead into removal of $PO_{OUT}$ values above 0.4, which are interesting in this study. After visual observation of the data, some of the measurements still had clear outliers, which were removed using fast limits. For example for $NO_{3IN}$ under 0, COD7 under 20, $EC_{IN}$ under 400, dissolved oxygen DO under 2.5 and aeration FO under 100. In time period [100000, 420000] and [600000, 800000] the sensor measuring $COD_{IN}$ , $NO_{2IN}$ and $NO_{3IN}$ was defect. Hence, in this study the $EC_{IN}$ sensor was used as proxy for the $COD_{IN}$ sensor. The ECIN sensor gave no measurements in period [115270, 167780] and during this period, the $EC_{IN}$ measurements were replaced with $EC_3$ measurements. However, there was no proxy measurement to replace the $NO_{2IN}$ and $NO_{3IN}$ measurements.
Because the data set was too large for some of the numerical methods, it was down-sampled to 10 minutes sampling time using mean values of 10 samples, in total 82051 samples. The correlation coefficients were calculated between the effluent $PO_{OUT}$ and the main Hias Process variables to study the effect of down-sampling. Correlation coefficients for 1-minute sampling time and 10-minutes sampling time give identical correlation coefficients.

### 3.2. Time delay compensation with dynamic time-lagging

The dynamics in the Hias Process are heavily affected by the influent water flowrate which periodically has high daily variation. The median value of 410 m3/h represents normal daily flowrate, while periodically the nighttime the flowrate sinks down to 200 m3/h. The water flowrate gives a time-variant delay to measurements along the ten process zones. This time-variant delay was modeled with a time-lag which is inversely related to the water flowrate and the water volume between the sensors, as presented in 1. While the median time delay between inlet and outlet is 32 samples (5h20min), the variation is large between minimum of 31 to maximum of 88. It is well to note that the median value is close to the minimum value, which is due to the maximum flowrate restriction to the Hias Process. To accounts for the blending effect of the physical walls between the different zones, additionally, the time-lag could have a moving-averaging filter with window size k, which is proportional to the standard deviation. We explored technique with dynamic time warping, but the results were not successful. In this study, dynamic time-lagging without moving-average filter was developed.

### 3.3. Exploratory Data Analysis and Correlation Analysis

Exploratory Data Analysis for the Hias process time series data set focuses on uncovering the key patterns, trends, relationships, and dependencies critical for effective clustering.
The statistical metrics of the preprocessed data were analysed. Temperature in the process varies between 4.7$^{o}$C to 15.9 $^{o}$C over yearly outside temperature pattern, while the flowrate varies between 140 m3/h in the nighttime without rain and up to 550 m3/h during the daytime with rain. The influent nutrients concentration, measured with $EC_{IN}$ becomes diluted when the influent flowrate is high. The aeration flowrate is highest when the influent nutrient concentration is highest. The aeration flowrate $FO_5$ in zone 5 is approximately 160% higher than $FO_8$ in zone 8, while the dissolved oxygen $DO_5$ in zone 5 is approximately 18% higher than $DO_8$ in zone 8. The kurtosis reveals that while distributions of $EC_{IN}$ , $FO_8$ and $COD_7$ are similar to normal distribution, as the kurtosis value (without bias) is close to zero. However, based on the skewness, most of the variable distributions are asymmetric and do not follow Gaussian distribution. Most extreme of the variables is temperature, which has a camel shaped distribution with two peaks. The non-Gaussian like distributions can indicate that it would not be feasible to apply z-score normalization to scale the variables.
The pair-plots of the main Hias Process variables were employed to understand temporal behaviour and relationships between variables. The pair-plots show high correlation between the aeration rates $FO_5$ and $FO_8$ (0.82) and their corresponding dissolved oxygen measurements $DO_5$ (0.77) and

$DO_8$ (0.70). However, the correlation of the aeration and dissolved oxygen variables to the predicted variable, $PO_{OUT}$ is modest 0.34-0.39. Based on the correlations, a set of main variables is selected, these include influent temperature T , water flowrate F , electrical conductivity ECIN, total aeration FO, aeration rates in zone 5 ($FO_5$) and zone 8 ($FO_8$), dissolved oxygen measurements in zone 5 $DO_5$ and zone 8 $DO_8$, zone 7 soluble COD (COD7) and nitrite ($NO_{27}$). The correlation coefficients between the effluent $PO_{OUT}$ and the main Hias Process variables are presented in Table 3 for the data with 1-minute sampling interval. The influent ECIN (0.39), F O8 (0.39), DO8 (0.38), $COD_7$ (0.42) and $SS_{OUT}$ have modest correlations to $PO_{OUT}$ . Inspecting visually the the pair-plots between $PO_{OUT}$ and the main Hias Process variables, these seem like large blobs. Possible reasons for this are the long and time-varying time-lags between the influent and effluent, nonlinear nutrient removal phenomena and drifting in the $PO_{OUT}$ measurement

The $PO_{OUT}$ measurement shows periodically high spikes which are caused by gradual clogging of the physical filter in the online analyzer. The process operators clean the filter every two weeks or when the $PO_{OUT}$ value gets out of bounds. The $PO_{OUT}$ is highly correlated with the effluent suspended solids ($SS_{OUT}$), and the peak values are aligned in the line plots. More research work is needed to develop robust method that removes the drifting in the $PO_{OUT}$ measurement.

### 3.5 Clustering

In a recent master thesis [14], Bommo has compared K-means, Density-Based Spatial Clustering of Applications with Noise (DBSCAN), Gaussian Mixture Models (GMM), Self-Organizing Maps (SOM) to characterize operational states in the Hias Process. The research work shows best clustering performance using K-means clustering and 8 clusters. We continue Bommos work by testing different combinations of variables for clustering, and develop case studies for two of the highest influent nutrient clusters.

The clustering methodology included normalization of data, selection of variables and number of clusters. The variables and number of clusters is evaluated using Silhouette score and Calinski-Harabasz score. The Silhouette score evaluates similarity of datapoint within the cluster, while the Calinski-Harabasz score evaluates dispersion ration between clusters and within clusters. Hence, the higher the scores, the better the separation of the clusters. First, combinations of the influent variables T,F, $EC_{IN}$ , $NO_{3IN}$ , and total aeration rate $FO_{TOTAL}$ were selected for testing. $FO_{TOTAL}$, $EC_{IN}$ and F exhibit the largest variation, and these form a sound basis for testing. A variable set consisting only of influent variables T,F, $EC_{IN}$ was also tested. In order to test if $NO_{3IN}$ had effect, this was combined with influent variable set T,F, $EC_{IN}$ , $NO_{3IN}$ . Finally a set of all the five variables was tested. Silhouette and Calinski-Harabasz scores were calculated for each combination of variables for a range of 2 to 8 clusters. In figure 3, the highest scores are obtained with variable set $FO_{TOTAL}$, $EC_{IN}$ and F. The curves have peaks at 3 clusters, suggesting three as the optimal number of clusters. However, based on the large number of data points and large variation in the variables, more clusters would be necessary to give a clear indication of the effect of influent and aeration to the phosphorus removal performance. We chose to test 6 and 7 clusters using variable combination $FO_{TOTAL}$, $EC_{IN}$ and F.

Using 7 clusters, the clusters separate between high and low influent water flowrate (F), and three different levels of nutrients ($EC_{IN}$). The mid-nutrient group is further separated in three based on the aeration rate, as illustrated in Figure 4, on top. These clusters applied to effluent variables DO8, $PO_{OUT}$, $FO_{TOTAL}$ shows that while the low $EC_{IN}$ -high flow-low aeration group in blue in Figure 4 (bottom) results in mainly low values of $PO_{OUT}$, the high $EC_{IN}$ -high flow-high aeration group (green) and the high $EC_{IN}$ -low flow-high aeration group (light blue) are much more scattered over all values of $PO_{OUT}$. Generally, low $EC_{IN}$ clusters result in low $PO_{OUT}$, but as they use least aeration, there is a small potential for savings. High $EC_{IN}$ clusters are selected for analysis as these require most aeration, they have widest variation in $PO_{OUT}$, and largest potential for savings in

aeration rates. Figure 4 show no clear separation of clusters in the scatter plot for aeration rate and effluent $PO_{OUT}$. Hence, these clusters are labelled and studied further.

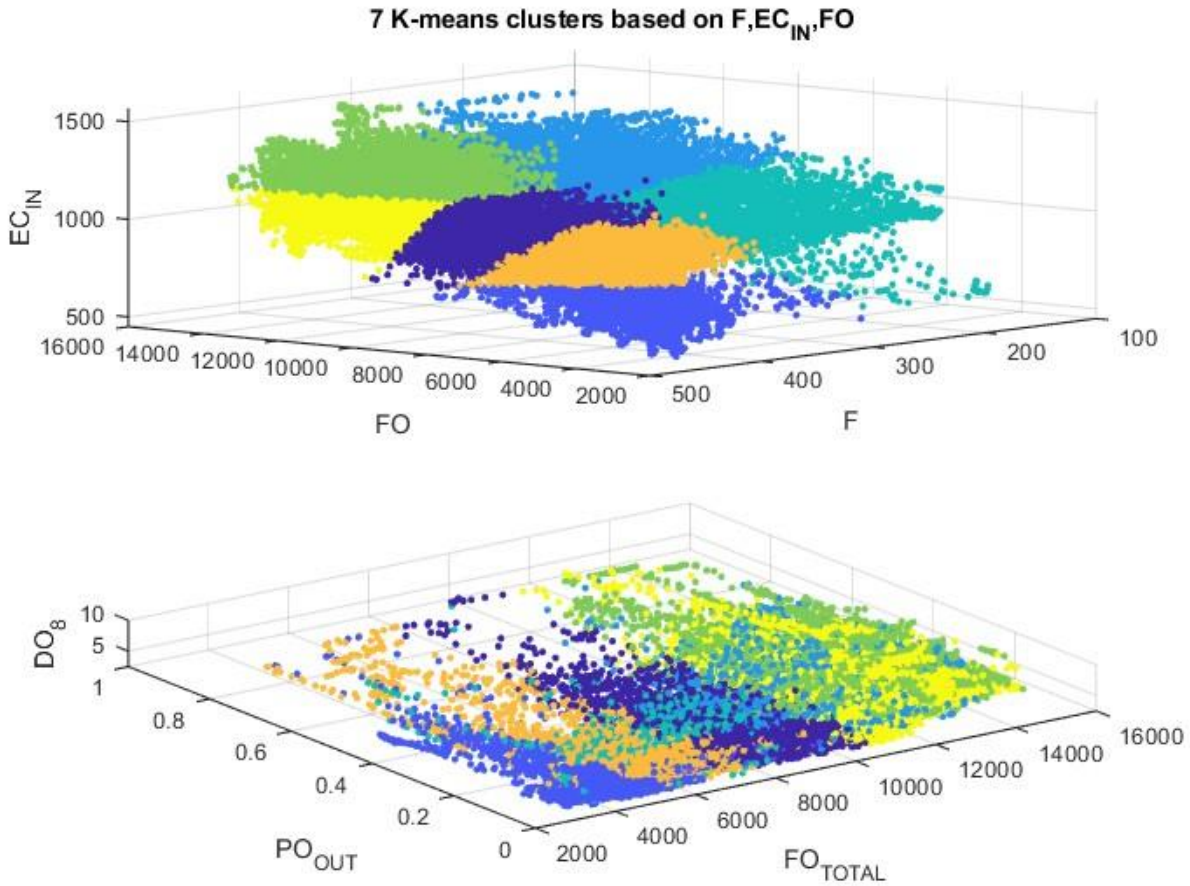


**Figure 4.** K-means 7 clusters. Top: Clusters visualised on variable set $FO_{TOTAL}$, $EC_{IN}$ and F, Bottom: Clusters visualised on variable set: $DO_8$, $PO_{OUT}$ , $FO_{TOTAL}$.

### 3.6 Labelling and case studies

The data was labelled according to the achieved nutrient removal performance to study further the effect on the phosphorus removal. The data points were labelled based on the $PO_{OUT}$ value: Excessive label for $PO_{OUT}$ under 0.3mg/L, Optimal for $PO_{OUT}$ between 0.3 mg/L and 0.4 mg/L, and Insufficient for $PO_{OUT}$ above 0.4 mg/L.

Case study on high nutrient, high flow cluster In the first case study, the high $EC_{IN}$ -high flow-high aeration group (green) is scatter plotted using the $PO_{OUT}$ labels, as shown in Figure 5. The top figure shows that the Excessive group (dark green) have lower aeration rates than the Optimal (light green) and Insufficient (yellow) groups while having the same range of nutrients and water flowrate. The effluent scatter plot of $DO_8$, $COD_7$, $FO_{TOTAL}$ (Figure 5 bottom) illustrates how the optimal group (light green) and the insufficient group (yellow) are at the edge of the excessive group (dark green). The high $DO_8$ combined with high aeration rate can indicate insufficient uptake of dissolved oxygen by the biomass. Further high $COD_7$ can indicate that COD was not sufficiently consumed in the anaerobic stages of the process, and hence the phosphorus removal did not reach it's full potential. For the Excessive aeration group (dark green), the aeration rates could be lowered. Less aeration could have been applied for the Excessive groups.

Case study on high nutrient, low flow cluster In the second case study, the high $EC_{IN}$ -low flow-high aeration group (light blue) is scatter plotted using the $PO_{OUT}$ labels, as shown in Figure 5. The labeling is colored with Excessive aquamarine, Optimal purple, and Insufficient red. The scatter

plot of influent variable set F, $EC_{IN}$ and $FO_{TOTAL}$ (top figure), illustrates how the Optimal and Insufficient groups are also in this cluster packed on the high aeration end. The labeled cluster visualised on the effluent variable set: $DO_8$, $COD_7$, $FO_{TOTAL}$ shows how the Optimal (purple) and Insufficient (red) data points lay on the edge of the data point cloud having high DO8, high COD7. Similarly to the high $EC_{IN}$-high Flow grouping, the high $DO_8$ combined with high aeration rate can indicate insufficient uptake of dissolved oxygen by the biomass. Further high $COD_7$ can indicate that COD was not sufficiently consumed in the anaerobic stages of the process, and hence the phosphorus removal did not reach it's full potential. For the Excessive aeration group (aquamarine), the aeration rates could be lowered, but this requires further investigation.

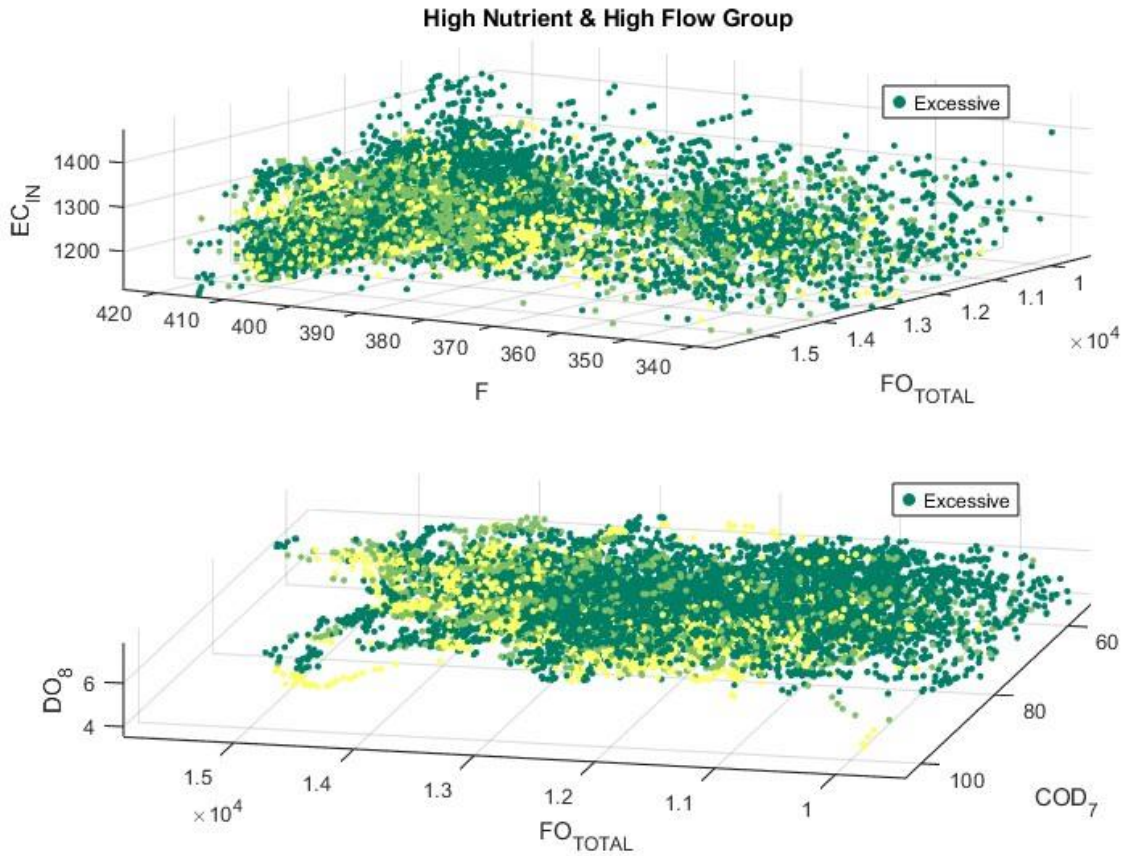


**Figure 5.** Labeled high nutrient, high flow cluster. Top: Labeled cluster visualised on influent variable set F, $EC_{IN}$ and $FO_{TOTAL}$. Bottom: Clusters visualised on effluent variable set: $DO_8$, $COD_7$, $FO_{TOTAL}$. Labeling: Excessive dark green, Optimal light green, Insufficient yellow.

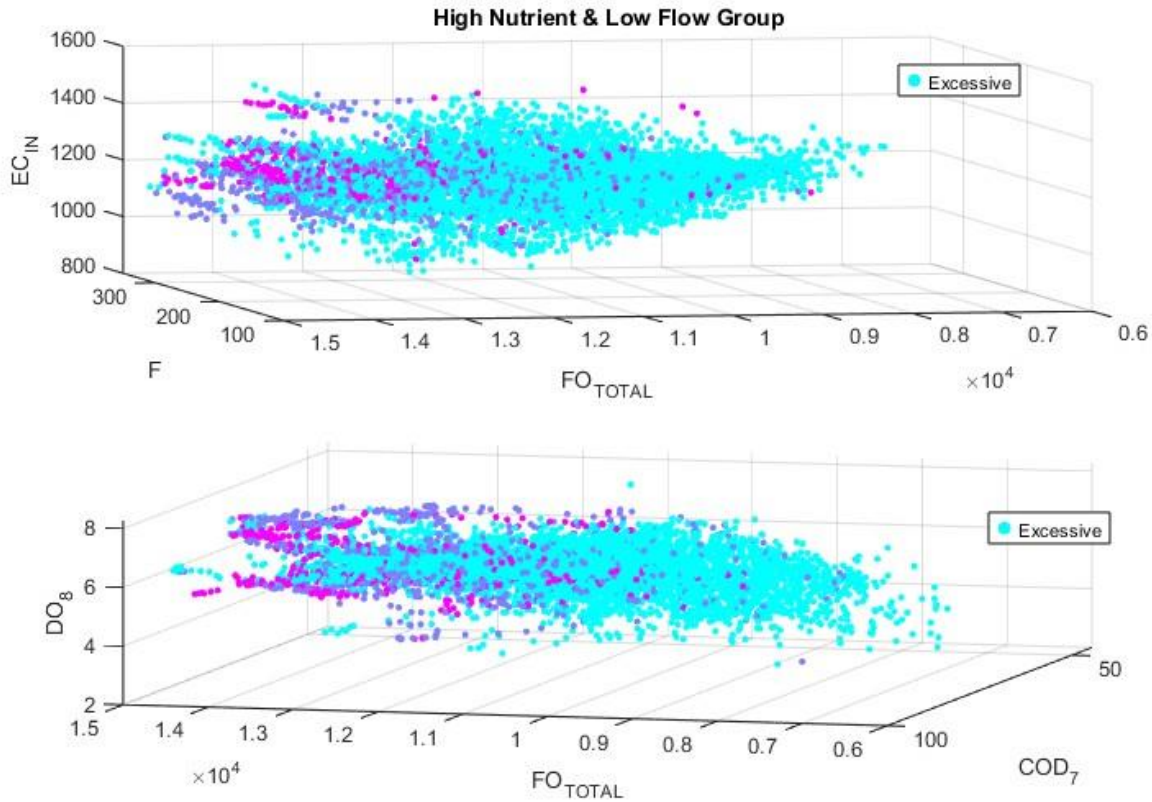


**Figure 6.** Labeled high nutrient, low flow cluster. Top: Labeled cluster visualised on influent variable set F, $EC_{IN}$ and $FO_{TOTAL}$. Bottom: Clusters visualised on effluent variable set: $DO_8$, $COD_7$, $FO_{TOTAL}$. Labeling: Excessive aquamarine, Optimal purple, Insufficient red.

## 4. Discussion

In this work we have developed pre-processing methods, clustering and labelling to visualize and analyse operation states in a biological phosphorus removal process, the Hias Process. The analysis of the high nutrient groups revealed potential to lower the aeration values to avoid excessive phosphorus removal. Based on the labelling of the high nutrient groups, the optimal and insufficient removal groups can be a result of insufficient COD uptake in the anaerobic stages of the process, and hence the phosphorus removal did not reach it's full potential. Further work is suggested in analysing these groups.

We answer our Research Question: Clustering can differentiate between operation states based on the inlet nutrients, flowrate and total aeration rate. The clusters with highest nutrient content however do not provide a clear answer to which operational conditions lead to excessive phosphorus removal. The labelling of optimal and insufficient groups revealed potential disturbances in the anaerobic zones. The clustering could be made more distinct by applying principal component analysis or other statistical methods mapping high dimensional data to lower dimensions. The problem could also be converted to classification problem, and Hierarchical clustering and classification methods could be developed and tested for the Hias Process. Possible sources of uncertainty are the drifting of the effluent phosphorus analyzer measurement. Hence, further work is suggested condition monitoring for the analyzer. Robust algorithms should be developed and tested to detect and remove drifting of analyzer measurements. Further, multi-sensor data fusion with SS and $COD_7$ measurements could be used to replace the removed measurements.

Future work will build upon this work on pre-processing and the clustering, to develop a robust classification model. This classification model can serve as a decision support tool, enabling the optimization of aeration flow rates with greater precision and efficiency.

## REFERENCES

[1] Knut Rudi, Inger Andrea Goa, Torgeir Saltnes, Gjermund Sørensen, Inga Leena Angell, and Sondre Eikås. Microbial ecological processes in mbbr biofilms for biological phosphorus removal from wastewater. Water Science and Technology, 79(8):1467–1473, 2019. doi: https://doi.org/10.2166/wst.2019.149.
[2] Andreas Froemelt, Leon Zueger, Luzia von Kaenel, Daniel Braun, and Wenzel Gruber. Pattern recognition of operational states leading to n2o-emissions in full-scale biological wastewater treatment. Water Research X, 29, 2025. doi: 10.1016/j.wroa.2025.100336.
[3] Wei Chen, Wenjie Guo, and Weijie Mao. A simultaneous multi-mode identification and outlier filtering method for imbalanced nonlinear process monitoring. Chemical Engineering Science, 313, 2025. doi: 10.1016/j.ces.2025.121714.
[4] Wenting Li, Chunhua Yang, Zhenxiang Feng, and Yonggang Li. A multimode data reconciliation method for wastewater treatment processes. Journal of Water Process Engineering, 72, 2025. doi: 10.1016/j.jwpe.2025.107420.
[5] Weng Zhaoqin, Sun Yuzhen, Lin Xiaobin, Yin Erxin, and Peng Daogang. Optimized operation strategy of desulfurization wastewater treatment system based on kfcm. In Proceedings – 2024 6th International Conference on Electrical Engineering and Control Technologies, CEECT 2024, page 104 – 108, 2024. doi: 10.1109/CEECT63656.2024.10898192.
[6] Peter Kern, Christian Wolf, Michael Bongards, Tosin Daniel Oyetoyan, and Seán McLoone. Self-organizing map based operating regime estimation for state based control of wastewater treatment plants. In 2011 International Conference of Soft Computing and Pattern Recognition (SoCPaR), pages 390–395, Dalian, China, 2011. IEEE. doi: 10.1109/SoCPaR.2011.6089275 .
[7] T.S. Moon, Y.J. Kim, J.R. Kim, J.H. Cha, D.H. Kim, and C.W. Kim. Identification of process operating state with operational map in municipal wastewater treatment plant. Journal of Environmental Management, 90(2):772–778, 2009. ISSN 0301-4797. doi: https://doi.org/10.1016/j.jenvman.2008.01.008
[8] Alexander Ebel, Michael Bongards, and Sean McLoone. State based control of wastewater treatment plants - evaluation of the algorithm in a simulation study. In Proceedings – Sixth International Conference on Hybrid Intelligent Systems and Fourth Conference on Neuro- Computing and Evolving Intelligence, HIS-NCEI 2006. IEEE, 2006. doi: 10.1109/HIS.2006. 264922.
[9] Tiina M. Komulainen, A. Malik Baqeri, Einar Nermo, Arvind Keprate, Torgeir Saltnes, Katrine Marsten Jansen, and Olga Korostynska. Estimation of effluent nutrients in municipal mbbr process. In Proceedings of The 64th International Conference of Scandinavian Simulation Society, Västerås, Sverige 26-27.9.2023, 2023. Linköping University Electronic Press. doi: https://doi.org/10.3384/ecp200037 .
[10] Tiina M. Komulainen, A. Malik Baqeri, Arvind Keprate, and Katrine Marsten Jansen. Comparison of ml and asm models for effluent nutrient estimation in the hias process. In Proceedings of the 65th International Conference of Scandinavian Simulation Society SIMS 2024 and the Second SIMS-EUROSIM conference, Oulu, Finland 10-12.9.2024, 2024. Linköping University Electronic Press. doi: https://doi.org/10.3384/ecp212.034 .
[11] Tiina M. Komulainen, Katrine Marsten Jansen, and Arvind Keprate. Analysing the effect of additional instrumentation on prediction of cod removal in the hias process. In Proceedings of the International Conference on Wider-Uptake of Water Resource Recovery from Wastewater Treatment, volume 524 of Lecture Notes in Civil Engineering, pages 400–405, Palermo, Italy 18-21.6.2024, 2024. Springer. doi: https://doi.org/10.1007/978-3-031-63353-9_68.
[12] Tiina M. Komulainen, Malik Baqeri, Katrine M. Jansen, Torgeir Saltnes, Alex T. Bech, and Olga Korostynska. Virtual sensors for the hias process. Water Practice and Technology, page wpt2024176, 07 2024. ISSN 1751-231X. doi: 10.2166/wpt.2024.176 .
[13] Malik Baqeri. Machine learning and control for phosphorous removal. Master thesis, Oslo Metropolitan University, 2024. URL https://hdl.handle.net/11250/3163016.
[14] Anders Bommo. Exploration of clustering and classification of operational states for the Hias Process. Master thesis, Oslo Metropolitan University, 2025. URL underpress.