3-31-2022 10:00 AM

Wastewater Aeration Process Dynamic Modelling: Combined Mechanistic and Machine Learning Approach

Yuehe Pan, *The University of Western Ontario*

Supervisor: Dagnew, Martha, *The University of Western Ontario*

A thesis submitted in partial fulfillment of the requirements for the Master of Engineering Science degree in Civil and Environmental Engineering

Abstract

The aeration process is the largest energy consumer in wastewater treatment plants (WWTPs), and the optimization of the process based on computational models can offer significant savings for the plant. Recent theoretical developments have revealed that many of the parameters commonly assumed as constants in aeration modelling, in fact, have a dynamic nature; however, there still lacks a universal way to model these factors in an easy, accurate and timely manner. This work proposed a machine learning-based modelling approach to offer real-time estimations of the oxygen transfer rate, airflow demand, and energy consumption.

Utilizing the field data collected from Adelaide WWTP (London, Ontario, Canada), the study developed and screened a combination of modelling approaches and input parameters for optimum predictive power under different data availability conditions.

The results demonstrated that the machine learning models provided significantly higher predictive power than the traditional mechanism-based models. These models can provide informative predictions of the aeration parameters with only operational parameters and limited knowledge about the underlying mechanisms of the system. When integrated with the theoretical equations, the models still produce reasonable estimations without losing interpretability. The present finding confirms using the machine learning modelling approach on dynamic factors involved in the aeration process to be feasible and effective. It calls for further investigation into such methods to explore more in the field of wastewater modelling.

Keywords: wastewater, aeration model, energy model, Dynamic model, Machine learning, alpha factor.

Summary for Lay Audience

Just like all dogs go to doggy heaven, all drains in our city lead to a wastewater treatment plant, where the microorganisms break down the pollutants and turn the wastewater back into clean water for us to use. The aeration process is the part in which we pump air into the water to let the microbes breathe and do their jobs. The more air they get, the happier they are and often the cleaner the resulting water can be. However, pumping air is a costly process, so our plants want to minimize the input air without harming the outflow water quality. One potential approach is to adjust the airflow pumps based on need. With the water flow rate and amount of pollutants changing, it is difficult to determine the need manually, but this can be quickly done if we have a mathematical model for estimations. In this study, we present a new way to make accurate and timely estimates of the many factors involved in the aeration process we need to achieve this goal.

While most previous researchers focused on describing mathematically how aeration works, an alternative could be to utilize machine learning methods for predictions. Machine learning models train themselves automatically through experience; however, we cannot easily explain the underlying reasoning of their predictions. To make up for this, we combined the two approaches and tested their performance under different scenarios.

We found that the machine learning models doubtlessly work better most of the time, but the combined models are also generally acceptable and easier for human to understand. We found some interesting relationships between parameters that we never thought of, and we are hoping to have them studied further by other inspired researchers.

To sum up, our study offered a new way to increase the accuracy of our various predictions in the aeration process, to have better control over the energy use in the pumping systems, and hopefully to save a fortune in treating our waste soon.

Co-Authorship Statement

This MSc thesis contains "submitted" materials or in preparation for submission in peer-reviewed journals, as listed below:

Chapter 3: "A new approach to estimate dynamic alpha factor in aeration systems using automatic machine learning models."

Submitted to *Environmental modelling and software.*

Authors: Yuehe Pan and Martha Dagnew. The primary author of this chapter is Yuehe Pan, under Dr. Dagnew's supervision. Yuehe collected, cleaned data, developed and validated models, interpreted the results, and drafted the manuscript with Dr. Dagnew's guidance.

Chapter 4: "Dynamic aeration and energy modelling."

In preparation for submission to *Water Environment Research*

Authors: Yuehe Pan, Kyle Murray, Rajev Goel and Martha Dagnew. The primary author of this chapter is Yuehe Pan, under Dr. Dagnew's supervision. Yuehe collected, cleaned data, developed and validated models, interpreted the results, and drafted the manuscript with Dr. Dagnew. Kyle Murray provided the data and helped check the quality of the data, review the manuscript. Dr. Rajev Goel assisted with the mechanism-based aeration and energy model assessment and manuscript review.

Acknowledgments

One big thank you goes to my supervisor, Dr. Martha Dagnew, for providing guidance, support, and encouragement through my studies. Thanks also to Reenste Filler and Kyle Murray for their contribution to data collection and their willingness to impart their knowledge. Further, I would like to thank all other members of our lab, especially Lin Sun, Wenjuan Mu and Henry Zhang, for their selfless help and encouragement. I would particularly like to thank my partner, whose support and patients allowed my studies to go the extra mile.

Finally, I would like to thank all the microorganisms we raised in our lab, in the aeration tank of Adelaide, and all other WWTPs across the world and throughout history. They take such a vital part in the wastewater treatment process and yet their roles are so commonly overlooked. They have made the ultimate sacrifice to recycle our waste, for which we, as humans, should always be grateful.

Table of Contents

List of Tables

List of Figures

List of Appendices

List of Abbreviations and Symbol

AutoML	Automatic Machine Learning
ASM	activated sludge model
AWS	Amazon Web Services
BOD	biochemical oxygen demand
C	The equilibrium concentration of oxygen in test liquid
CF	conversion factor
C_{in}	influent concentration
C_{out}	effluent concentration
C^s	Saturation concentration of oxygen in test liquid in equilibrium with exit gas
C_{20}^S	Saturation concentration of oxygen in test liquid in equilibrium with exit gas at 20 ºC, 1 atm, and zero salinity
COD	chemical oxygen demand
Cp	heat capacity of air at constant pressure
d	Index of Agreement
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
DO	dissolved oxygen
DP	delivered power
eps	radius in DBSCAN
E	coefficient of efficiency
f	Fouling factor of diffuser
HRT	hydraulic retention time
IWA	international water association
$k_L a$	Overall oxygen mass transfer coefficient of liquid

KNN	k-nearest neighbour algorithm
minPts	the minimum number of neighbours
ML	machine learning
MLSS	mixed liquor suspended solids
MW_{O2}	Molecular weight of oxygen
NH_3-N	Ammonia nitrogen
NSE	Nash-Sutcliffe efficiency
OTE	Oxygen Transfer Efficiency
OUR	oxygen uptake rate
PCP	pollution control plant
P_a	Absolute pressure upstream (inlet) of blower
P_d	Absolute pressure downstream (outlet) of blower
PE	primary effluent
PO_4-P	orthophosphate-phosphorus
Q_{in}	influent flow rate
Q_{out}	effluent flow rate
q_i	Total gas volume flow rates of inlet gas
q_o	Total gas volume flow rates of outlet gas
r	rate of reaction
R	Ideal gas constant
ρ	Density of oxygen at temperature and pressure at which gas flow is expressed
RMSE	Root-mean-square-error
SCASA	Supervisory Control and Data Acquisition
sCOD	soluble chemical oxygen demand

SOTE	standard oxygen transfer efficiency
SVM	Support vector machine
t	time
T	Water temperature
TSS	total suspended solids
θ	Theta factor
V	liquid volume or volume of aeration tank
VSS	volatile suspended solids
w	Molecular mass flow rate of air
WWTP	wastewater treatment plant
Y_R	Volumetric fractions of oxygen gas in inlet gas
Y_{og}	Volumetric fractions of oxygen gas in outlet gas

Chapter 1

1 Introduction

1.1 Rationale

As a process commonly adapted in modern wastewater treatment, research on aeration processes has a long tradition. Aeration is the process of delivering oxygen or air into wastewater treatment process units to allow aerobic biodegradation of the organic matter and nutrients present in the water. The blower and air support systems are major energy consumers in wastewater treatment plants (WWTPs) in the aeration process. On average, the aeration process accounts for 50% of the total energy consumption of the plants. While decreasing the energy use of the aeration process by replacing more efficient motors could lead to an annual saving of 20,000 Dollars (Government of Ontario, 2016), a much more significant energy saving can also be achieved by aeration process optimization without sacrificing water quality.

Microorganisms in the aeration process rely on the dissolved oxygen (DO) level to perform biodegradation; therefore, sufficient aeration is critical for adequate wastewater treatment. In theory, the ideal DO concentration in the aeration process would be around 2 mg/L, representing equilibrium: the oxygen supplied is the same as the oxygen demand of biodegradation (Meng et al., 2017). However, due to the complex nature of the wastewater, many factors often contribute to variation in DO concentration. With the complex cross-effect of diffuser type, wastewater loading, temperature, time of day, and many other

factors, estimation of the DO outcome made upon the level of airflow is often inaccurate and lagging (Drewnowski et al., 2019). To make up for such a lack of information, many WWTPs tend to over-aerate the process hoping to ensure the effluent water achieves effluent quality standards. For such an energy-intensive process, even aerating at 1 or 2 mg/L over the DO setpoint may lead to a significant increase in cost (S. Bolles, 2006). Considering the above, we can naturally conclude that there is much room for improvements in the energy efficiency ratio in the aeration process.

Over the past decades, considerable efforts have been made to design and modify the aeration process to reduce energy costs. The majority of the modern WWTPs utilize feedback control loops to control the DO. However, feedback control is often vulnerable to time delay and disturbance, especially at larger plants where data collected at one point often is not representative of the whole tank (Åmand et al., 2013). Alternatively, the concept of dynamic aeration modelling can further be utilized for aeration process control and optimization (Martin & Vanrolleghem, 2014). In this case, the dynamic aeration models were used to build feedforward control loops to predict DO or required airflow based on real-time or forecasted influent data (Åmand et al., 2013). Martin and Vanrolleghem (2014) demonstrated that by modifying the control system in WWTPs, the total energy cost of aeration could be saved by 27% on average. However, one of the standing research questions is the accuracy of the aeration models to capture dynamic aeration and energy demand within the process.

One of the parameters that were considered to impact dynamic aeration modelling accuracy is the alpha factor. To date, numerous studies have investigated the correlation between the

alpha factor and an extensive range of operational parameters, but the existing empirical equations often lack precision in predicting the airflow needed for a single WWTP under dynamic operational conditions. One approach to address this problem is developing other alpha modelling concepts and using alternative data-based modelling approaches such as machine learning (ML) algorithms. Several studies have shown that ML is particularly suitable for estimating values within the environmental domain for its ability to model complex, nonlinear relationships among numerous variables (Guo et al., 2015).

1.2 Objective

Although many models aimed to describe the aeration process exist, most of those are not efficient in dynamic modelling, especially in high sampling frequency time scale. Few studies have shown methods designed for airflow estimation based on real-time water quality measurements, but the existing empirical equations often lack precision in predicting the airflow needed for a single WWTP under dynamic operational conditions. To fill this knowledge gap, this study proposes a new framework for building a combination of mechanism-based equations and machine learning methods to model the aeration process of WWTPs. The objectives of this study were: 1). To evaluate the dynamics of oxygen transfer factors and develop protocols representing dynamic alpha-factor; 2). To establish models to estimate real-time airflow demand, and 3). To assess the dynamic blower power consumption in the aeration process.

1.3 Thesis format and organization

This thesis is structured as follows: Chapter 1, *Introduction*, has provided a brief overview of the background, defined the research aim, and established the key objectives of this study. Chapter 2, *Literature Review*, served as an appraisal of previous studies within the field of aeration modelling. It covered the known mechanisms of the aeration process and provided a comprehensive overview of the existing modelling approaches. Chapter 3, *Dynamic Alpha Modelling*, presents our investigation on the oxygen transfer factors and the model established for real-time estimation. Chapter 4, *Dynamic Airflow and Energy Modelling* followed a similar research structure but focused on modelling the airflow demand and the energy consumption in the aeration process. This completes the second and third objectives of the study. Finally, Chapter 5, *Conclusions and recommendation*, summarized the main findings in this study and suggested possible directions for future work.

Chapter 2

2 Literature Review

2.1 Wastewater Treatment Modelling

2.1.1 Traditional Modelling

2.1.1.1 Modelling background

WWTP models are used in various applications, including plant design, operation, training, optimization, and diagnosis. They are physical models that consider the mass balance of the treatment system and biochemical kinetic processes to simulate the operation of a virtual or full-scale treatment plant. The models operate by simultaneously solving a series of mass balance and kinetic relationships to predict the behaviour of the wastewater as it is processed through the plant under a given set of operating conditions. The output of these models, which describe the wastewater characteristics over both space (location within the treatment train) and time, can then be used for the purposes mentioned above.

2.1.1.2 Kinetics and mass balance

Kinetic relationships describe the transformation of materials such as organic matter and nutrients in the water into bacterial cells, thereby removing contaminants from a dissolved, unsettle-able state (organics dissolved in the wastewater) into a particulate state (contained within the cells) that can be removed via settling or filtering. Mass balances are applied to

the solid (sludge, suspended solids) and liquid flows of the wastewater. By solving the kinetic and mass balance equations simultaneously for a given time step, the temporal changes in the wastewater characteristics throughout the treatment process are predicted.

Conservation of mass must be maintained throughout the modelled system. In its simplest form, this describes the flow distribution in the treatment plant; any liquids and solids entering the treatment plant must be expelled either in the effluent or through some removal mechanism along the treatment path. Mass balances included in wastewater models contains the followings: solid and liquid flows; physical reactions such as gas-liquid mass transfer, diffusion, adsorption; solids separation such as sedimentation and filtration; and reactor hydraulics. Equation 2.1 shows an example mass balance for a reactor with one influent and one effluent for a single contaminant concentration.

$$Q_{in}C_{in} - Q_{out}C_{out} + r = \frac{d(C \cdot V)}{dt} \quad \text{(Equation 2.1)}$$

Where

Q_{in}	Influent flow rate	[m^3/d]
Q_{out}	Effluent flow rate	[m^3/d]
C_{in}	Influent concentration	[mg/L]
C_{out}	Effluent concentration	[mg/L]
r	Rate of reaction	[g/d]
V	Liquid volume	[m^3]
t	Time	[d]

2.1.1.3 Activated Sludge Models

Biological treatment is the principal treatment method most wastewater facilities implement to remove organics and nutrients. Activated sludge is used as a treatment mechanism within which active biomass is allowed to grow under a preferential set of conditions to achieve specific treatment outcomes. The treatment is achieved by assimilating dissolved organic and nutrient components in the wastewater into bacterial cells, which can then be removed from the water through solids separation such as settling or filtration. The most commonly employed models used to describe this process include; variations of the activated sludge model (ASM) such as ASM1 (Henze et al., 1987), ASM2 (Gujer et al., 1995) ASM2d (Henze et al., 1999) and ASM3 ((Henze et al., 1999); the Dold's General Model (Barker & Dold, 1997); the Mantis2 Model (Hydromantis Environmental Software Services Inc., 2015); and the Metabolic Models (Roles, 1983). Differences between the models are primarily based on the number of processes included. The first widely accepted activated sludge model was the Activated Sludge Model No. 1 (ASM1), developed by the international water association (IWA) task group (Henze et al., 1987). The model incorporates the kinetic and physical processes of hydrolysis, carbon oxidation, nitrification, and denitrification. This model was later modified and expanded upon to include fermentation, biological phosphorous removal, carbon storage (Phosphorus removal in ASM2 and Bio-P in ASM2d), and biological phosphorous removal with denitrification (ASM2d only) in ASM2; and general carbon storage in ASM3.

ASM1 has the most fundamental structure of the above-mentioned models; a system of 8 equations describing the aerobic growth of heterotrophs, anoxic growth of heterotrophs, aerobic growth of autotrophs, decay of heterotrophs, decay of autotrophs, ammonification

of soluble organic nitrogen, hydrolysis of entrapped organics, and hydrolysis of entrapped organic nitrogen.

2.1.2 Machine Learning Models

Other than the theory-driven traditional modelling methods, machine learning is a data-driven approach that can rapidly process datasets and uncover the underlying patterns that wastewater researchers might not have yet discovered. Machine learning comprises a group of computational algorithms that can perform model fitting, pattern recognition and prediction based on existing data that are suspected to be related to the desired outcome. On the other hand, as the machine learning models are created by the algorithms directly from the data, these models are often "black-boxes," meaning that they are often uninterpretable, making it hard for a human to understand how the input variables are linked with each other to reach the final prediction.

The ability of machine learning models to find patterns in complicated systems provides it with the advantage of solving water engineering problems (Olden & Jackson, 2002). This is especially true within the field of wastewater process modelling, as this is a complex, multivariate, highly nonlinear system where limited knowledge is available. The use of machine learning models also allows for fast and easy development and handling, with or without the necessity of conducting web-lab experiments. These advantages make the machine learning model an effective and accurate pathway for monitoring, optimizing and

predicting the behaviour of the activated sludge process in WWTPs without the need to install additional sensors or new control systems.

Machine learning models have been proved to be a useful tool and accurate predicting method in hydrological modelling, rainfall-runoff modelling, flow modelling, and many other fields (Kişi, 2004). Newhart *et al*. summarized the various data-driven models used in the field of activated sludge modelling (Newhart et al., 2019). Most of them are either deterministic first-principal models, secondary clarifier models or a combination of the two. To the best of our knowledge, machine learning methods have not yet been widely adopted within the area of activated sludge modelling.

2.1.3 Data Collection for Modelling

2.1.3.1 Data Collection

For a dynamic model to be considered for application, it must be accurate to within an acceptable threshold determined by the model's intended use. As a result, the model inputs that are required to produce accurate simulations and/or predictions must be available in terms of both sufficient quality (measurement accuracy and reliability) and quantity (spatial and temporal intervals between measurements) (Rieger et al., 2010).

The data required is governed principally by the intended application of the model. However, generally, some form of influent characterization is required (either to the process of interest or the system as a whole), with the most critical being the most influential constituents (such as COD) and most actively varying (such as Flow). Flow

measurements are among the most critical because all concentrations must be multiplied by the respective flow to determine the overall loading to the plant (Rieger et al., 2010). Characteristics that can be easily estimated and/or remain relatively constant may not need to be measured as frequently or at all.

As mentioned, the input interval for a dynamic model can vary in a range of minutes to weeks, depending on the application and data availability. It is infeasible to manually collect, analyze, and report such a large quantity of samples for a prolonged period, based on the labour and lab time required. Therefore, a treatment plant needs to be outfitted with appropriate online sensors.

2.1.3.2 Data Resolution

The complexity of a model can also be increased by discretizing the modelled period (timescale) into smaller time steps, where the model input and operation parameters are changed during each timestep. This is also termed static and dynamic modelling. A completely static model would have one set of input variables (such as influent characterization) classified at the beginning of the simulation, remaining constant throughout the simulation. An example of this would be a steady-state model. A dynamic model would have a set of input variables altered at given time intervals based on the available data. The complexity increases as the resolution of the dynamic data increases, for example, from model inputs that are varied daily or every 15 minutes.

Similar to the model dimensions and processes, the data resolution required for a model is dependent on the application. Figure 2-1 shows examples of common model applications.

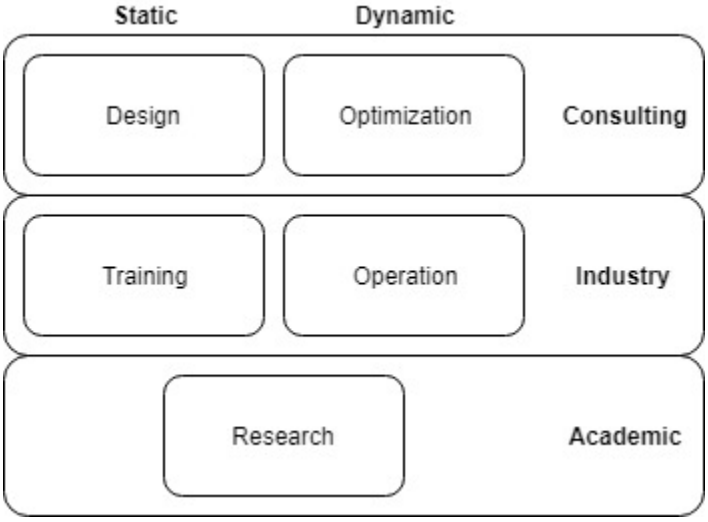


Figure 2-1 Common model applications

Models used for Design and Training purposes often rely on static datasets to reproduce steady-state scenarios which provide information on the long-term behaviour of the plant. Models for optimization and operation require dynamic datasets to predict the behaviour of the plant under short-term variable conditions (Rieger et al., 2010). The resolution of these dynamic datasets is dependent on data availability and application and can vary between weekly, daily, hourly, and 15-minute data sets.

The resolution of the model input limits the resolution of the model output. Therefore applications that require model outputs on the scale of hours to minutes should implement equally or further discretized input datasets. The availability of accurate high-resolution data has been the limiting factor in creating influent data sets for dynamic modelling. With

the increased availability of online sensors, high-resolution data has become more readily available. However, when datasets become too discretized and the temporal resolution becomes very small (for example 15-minute data), input data can become excessively "jagged" in which case implementing moving averages (for example 2-hour moving average) can result in a smoother dataset (Martin & Vanrolleghem, 2014). There is currently no standard for determining the optimal data resolution that should be used for dynamic modelling. There is also little literature comparing the accuracy of models as the data resolution increases.

2.1.3.3 Data Preprocessing

The result of data-driven research depends highly on the quality of the input datasets. However, raw datasets collected directly from WWTPs and labs are often corrupted with inconsistencies, noise, and missing values. Without proper treatments, these datasets could introduce bias and/or cause under-coverage to the results and conclusions.

Data preprocessing is a set of techniques used to clean, transform, and encode raw datasets to reflect the actual condition of the inspected variables and can be easily parsed by the algorithms. Generally, it consists of four main stages: data cleaning, data integration, data transformation and data reduction. In wastewater research, the input data are limited and are often collected from a single source. Therefore, integration, transformation and data reduction are usually less concerned, leaving data cleaning the primary focus in data preprocessing procedures.

Data cleaning includes filling missing values, eliminating the noise in data, and treating inconsistent data based on external references and knowledge of the typical data range for each parameter collected. The missing value issue can be solved by either ignoring the tuples containing missing values or filling the gap in various ways. Common methods include filling the values manually, replacing them with data attributes like average values, and predicting the missing values with regression, maximum likelihood or other statistical methods (Kang, 2013). The noise in a measured variable is the abnormal entries caused by random errors or variance. Besides removing the erroneous data manually, it can be achieved through techniques like binning, regression model fitting and clustering. After these steps, the resulting dataset should have enough accuracy and reliability in terms of information and be ready to be fed into the models.

2.1.4 Model Evolution

After the parameters are chosen, and the appropriate model is applied, we unavoidably come to the question: how reliable are the predictions of our model? This is essential for several purposes, including model validation, evaluation, comparisons and out-of-sample comparisons. One common evaluation of models is the goodness-of-fit measurement, which summarizes the discrepancy between the observed and the predicted values generated by the model. As questions in the field of wastewater modelling often fall in the scope of regression analysis, a list of potential goodness-of-fit measures is summarized by Ahnert et al. (Ahnert et al., 2007). This list includes the R-squared measure, Root-mean-

square-error (RMSE) and coefficient of efficiency (E) and is continuingly growing as new applications emerge.

2.2 Wastewater Aeration Oxygen Transfer Efficiency

A widely used practice is activated sludge treatment, a process during which a bulk liquid consisting of new wastewater and recycled sludge from the secondary clarifiers is aerated using air. This process maintains a high concentration of biomass in the mixed liquor. The biomass uses contaminants in the wastewater as substrate and dissolved oxygen as an electron acceptor. The resulting biomass growth removed contaminants from the bulk liquid, which is separated using settling or filtration. The process must provide sufficient oxygen to facilitate an optimal amount of biomass growth; blowers and diffusers are implemented to continuously re-oxygenate the bulk liquid.

2.2.1 Aeration Theory and Modelling

Bubble aeration systems in activated sludge processes serve two main functions: meet process oxygen demand and provide mixing. The required oxygen volume for mixing is significantly less than the amount required for biomass growth; therefore, oxygen requirement is calculated based on optimizing biomass growth. Estimating the oxygen demand of the system and modelling the aeration process is useful for design, reducing operating costs, energy footprints, and/or carbon emissions and research.

Conventionally the desired DO concentration for aerobic conditions is 2 mg/L (Amand et al., 2013), below which the growth rate of organics may become limited by oxygen availability. Ideally, this condition will provide sufficient oxygen required by organisms, keep solids in suspension, and create sufficient residual dissolved oxygen levels in the tank effluent to satisfy downstream processes. When the oxygen concentration is exceeded, it may cause lower denitrification and higher energy demand. However, if the aeration system does not meet the demands, it may cause poor downstream and effluent water quality, filamentous bulking, poor settling, and growth inhibition (Amand et al., 2013). The amount of DO to meet treatment requirements can be determined using process models and/or steady-state calculations. Following this, an appropriate aeration method needs to be selected to deliver the required DO.

Several kinds of aeration methods are implemented in wastewater treatment facilities, including passive aeration (eg. surface aeration) and active aeration (e.g. paddle aeration, fine/coarse bubble aeration). The most commonly used method is fine bubble aeration, where bubbles are formed by pushing pressurized air through porous materials. The amount of oxygen transferred from the gas phase to the liquid phase in clean water is described by Equation 2.2 (Pittoors et al., 2014)

$$\frac{dC}{dt} = K_L a \cdot (C^s - C) \quad \text{(Equation 2.2)}$$

Where

$\frac{dC}{dt}$	Change in concentration over time	[mg/L·h]
$k_L a$	Overall oxygen mass transfer coefficient of liquid	[h^{-1}]
C^s	Saturation concentration of oxygen in test liquid under equilibrium	[mg/L]
C	The concentration of dissolved oxygen in test liquid	[mg/L]

Under 1 standard atmosphere condition, Saturation concentration of oxygen (C^S) can be calculated based on Equation 2.3 (VLMP, 2014)

$$C^s = 14.189e^{-0.022T} \quad \text{(Equation 2.3)}$$

Where

T	Temperature of test liquid	[℃]

In clean water, $k_L a$ can be estimated using a mass transfer model by de-oxygenating and measuring the dissolved oxygen while re-oxygenating the water (American Society of Civil Engineers, 2007). Compared to clean water transfers, the amount of oxygen transferred in wastewater treatment processes is lower as it is impacted by various factors, including the aeration equipment, operating conditions and process design (Pittoors et al., 2014).

2.2.2 Factors Affecting Oxygen Transfer Efficiency (OTE)

Oxygen Transfer Efficiency (OTE) is a percentage describing how much of the oxygen bubbled through the tank transfers from the bubbled gas phase to the bulk liquid. Figure 2-2 and Equation 2.4 show the gas phase mass balance and corresponding derivation of OTE, respectively (Boyle et al., 1989).

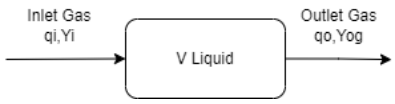


Figure 2-2 Gas Phase Mass Balance

$$OTE = \frac{Mass\ Oxygen\ Transferred}{Mass\ Oxygen\ Supplied} = \frac{Mass\ Oxygen\ In - Mass\ Oxygen\ Out}{Mass\ Oxygen\ In}$$

$$OTE = \frac{\rho(q_i Y_R - q_O Y_{OG})}{\rho q_i Y_R} \qquad \text{(Equation 2.4)}$$

Using mass oxygen transferred (Equation 2.2), OTE can also be expressed as Equation 2.4.

$$OTE = \frac{\alpha(k_L a)_{ww} \cdot (C^S - C)V}{\rho q_i Y_R} \qquad \text{(Equation 2.5)}$$

where

OTE	Oxygen transfer efficiency	[-]
ρ	Density of oxygen at temperature and pressure at which gas flow is expressed	[m^3/d]
q_i	Total gas volume flow rates of inlet gases	[m^3/h]

q_o	Total gas volume flow rates of outlet gases	[m^3/h]
Y_R	Volumetric fractions of oxygen gas in inlet gases	[-]
Y_{OG}	Volumetric fractions of oxygen gas in outlet gases	[-]
α	Alpha factor	[-]
$(K_L a)_{ww}$	Overall oxygen mass transfer coefficient of wastewater	[h^{-1}]
C^S	Saturation concentration of oxygen in test liquid under equilibrium	[mg/L]
C	The concentration of dissolved oxygen in test liquid	[mg/L]
V	The volume of aeration tank	[m^3]

OTE can also be expressed in terms of standard oxygen transfer efficiency (SOTE), which is the oxygen transfer rate in clean water conditions, and various correction factors. Equation 2.6 shows the oxygen Transfer Efficiency in Wastewater (Yoon, 2016).

$$OTE = \alpha \cdot \theta^{(T-20)} \cdot F \cdot SOTE \cdot \left(\frac{\beta \cdot C^S - C}{C_{20}^S}\right) \quad \text{(Equation 2.6)}$$

where

OTE	Oxygen transfer efficiency under field condition	[-]
α	Alpha factor	[-]
θ	Theta factor	[-]
T	Temperature of aeration tanks	[℃]
F	Fouling factor	[-]

SOTE	Oxygen transfer efficiency under standard condition	[-]
β	Beta factor	[-]
C^S	Saturation concentration of oxygen in test liquid under equilibrium	[mg/L]
C	The concentration of dissolved oxygen in test liquid	[mg/L]
C_{20}^S	Saturation concentration of oxygen in test liquid in equilibrium with exit gas at 20 ºC, 1 atm, and zero salinity	[mg/L]

The correction factors account for discrepancies between the SOTE in clean water and the OTE in wastewater. OTE is a highly variable property dependent on the treatment process's physical, environmental, and operational state. Physical properties such as diffuser type and depth (Table 2-1), distribution, fouling, airflow rate, and reactor properties such as depth, volume, and type of reactor affect SOTE and alpha (Tchobanoglous et al., 2003). The impact of environmental factors such as temperature and pressure are directly referenced using the beta and theta factors. Operating conditions such as SRT, temperature, turbulence, and wastewater composition affect the alpha factor as well as the saturated oxygen concentrations (Pittoors et al., 2014).

Table 2-1 Diffuser Standard Oxygen Efficiency Source (Yoon, 2016)

Diffuser Type	SOTE at 4.5m submergence
ceramic discs	0.26-0.33

ceramic domes	0.27-0.29
porous plastic discs	0.28-0.32
perforated membrane tubes	0.22-0.29
coarse bubble diffusers	0.09-0.13

Environmental Factors are not under the control of the operator or designer of the treatment plant. Higher water temperature negatively impacts the solubility of oxygen, thereby decreasing C^s and negatively impacting OTE (Jenkins, 2013). Barometric pressure and salinity also influence the OTE, as C^s decreases when the atmospheric pressure below 1atm or the salinity is high (Baquero-Rodriguez et al., 2018). Process Conditions of the system may or may not be under the control of the operator or designer. The presence of surfactants in the water can accumulate at the bubble surface and negatively influence the OTE (Chern et al., 2001; Rosso et al., 2008; Baquero-Rodriguez et al., 2018). While the operator cannot directly control the presence of these surfactants, introducing a higher SRT enhances the degradation of more complex substances, including surfactants. Increasing SRT results in a higher MLSS concentration. This may also influence the OTE; at high MLSS concentrations, solids accumulating on the bubble surface reduce the permeability, and OTE is reduced (Henkel et al., 2011; Krampe & Kreauth, 2003; Germain et al., 2007). Combing the results of several studies has shown a general trend in which up to a threshold of approximately 4 g/L an increase in MLSS increases OTE, while concentrations in excess of 4 g/L subsequently decrease the OTE. Finally, bubble size, impacting the specific

surface area, is dependent on the airflow rate (Baquero-Rodriguez et al., 2018). An increase in airflow makes the bubble size bigger, negatively impacting the OTE (EPA, 1989). Finer bubbles create a larger surface area and longer residency time, thereby improving the oxygen transfer. It is of note that bubble size and shape, and consequently interfacial surface area, are impacted by the presence of surfactants. Finally, the age and cleaning frequency of the diffusers impacts the fouling of the diffusers. As fouling decreases the number of available pores for diffusion, the pressure increases results in larger bubbles, decreasing OTE.

2.2.2.1 Correction Factors

Several factors are used to account for the observed discrepancy between clean water and wastewater OTE, including alpha, beta, gamma, and omega. These correction factors are introduced to increase the accuracy of the calculated OTR in wastewater. Alpha accounts for tank and wastewater and sludge characteristics, beta for salinity and surface tension, gamma for temperature and omega for pressure.

Beta, gamma, and omega can all be determined based on a singular, easily measurable property (salinity, temperature, and pressure respectively) and remain relatively constant within a typical municipal plant (Pittoors et al., 2014). The presence of dissolved solids such as NaCl in industrial wastewater may require the use of a beta factor if industry provides a significant amount of the inflow to a municipal plant, however, generally, the use of a constant factor of 0.99–1 is applicable (Tchobanoglous et al., 2003). Gamma can be calculated based on the wastewater temperature, which may vary slightly seasonally. Pressure is based on the ambient barometric pressure and remains constant at any given location. Table 2-2 summarizes the typical ranges of the correction factors in wastewater. For typical wastewater beta, gamma, and omega only vary slightly compared to alpha. Therefore, the impact of these factors is often ignored, and emphasis is placed on the alpha factor.

Table 2-2: Correction factors Sources: (Pittoors et al., 2014), (Yoon, 2016)

Correction Factor	Accounts For	Range	Typical Value for diffused aeration in WWTP
Alpha	Tank and wastewater characteristics	0.3-1.2	0.4-0.8
Beta	Salinity and surface tension	0.7-1.0	0.99-1.00
Gamma	Temperature	1.015-1.04	1.024
Omega	Pressure	0.82-1.0	0.99-1.00
Fouling	Buildup on diffuser	0.5-1.0	0.7-0.9

Alpha accounts for variations in both tank and wastewater characteristics. Both properties encompass a wide range of individual measures, each of which influences the factor to varying degrees. As a result, the type of direct calculation applied to the other correction factors is not viable for the alpha factor because the properties it describes cannot be simplified into one singular measurement

2.2.3 Alpha Factor

In wastewater applications, alpha is cited to vary between 0.3 for fine bubbles, up to 0.85 (Baquero-Rodriguez et al., 2018) and from 0.6 (conventional activated sludge) to 0.5 (membrane bioreactor) (Henkel et al., 2011). Alpha represents the reduction in oxygen transfer efficiency in the mixed liquor compared to clean water and can be defined as shown in (Pittoors et al., 2014).

$$\alpha = \frac{k_L a_{wastewater}}{k_L a_{clean\ water}} \qquad \text{(Equation 2.7)}$$

Where

α	Alpha factor	[-]
$k_L a_{wastewater}$	Overall oxygen mass transfer coefficient of wastewater	[h^{-1}]
$k_L a_{clean\ water}$	Overall oxygen mass transfer coefficient of clean water	[h^{-1}]

As discussed in section 2.2.1, alpha is a correction factor that accounts for the discrepancies between the OTE in clean versus process water, the most cited in the literature as being dynamic.

Conventionally in modelling applications, alpha is assumed to be a constant value, which is calibrated to match model outputs to plant performance. However, the alpha factors vary temporally (or spatially) in pilot and full-scale treatment plants. Literature has mentioned a temporally varying alpha for over 30 years (Boyle et al., 1989) however, research of a variable alpha in modelling has only been conducted in the last two decades (Gunder, 2001;

Krampe & Kreauth, 2003; Henkel et al., 2011; Amerlink et al., 2016; Jiang et al., 2017; Ahmed et al., 2021). These studies attempt to determine and implement an empirical relationship between a process condition and the alpha factor to improve the modelled dissolved oxygen (DO) accuracy. Ahmed *et al.* (2021) confirmed the existence of the COD-alpha relationship and further investigated the correlation between the alpha factor and the real-time sCOD in the bioreactor level using data obtained from sequencing batch reactors.

Table 2-3 summarizes the proposed empirical alpha factor correlations published between 2003 – 2021. It should be noted that the reasoning for developing an empirical relationship between alpha and a process condition in favour of directly measuring alpha is the intensive process required in obtaining an alpha measurement. The use of an off-gas hood is considered the most accurate method for determining alpha (Boyle et al., 1989). This equipment presents a high overhead cost as well as additional maintenance requirements for a treatment plant.

All seven publications agree that wastewater characteristics affect the alpha factor. However, there is no consensus regarding which characteristic is predominantly responsible for affecting alpha, nor regarding the nature of the relationship (Figure SD 1). Krampe and Kreauth (2003) conclude that there is an inverse relationship between MLSS and alpha, while Henkel *et al.* (2011) determines that the spread is too large to associate a relationship and MLVSS is a better indicator. Moreover, Baquero-Rodriguez *et al.* (2018) concludes that a double exponential equation best represents the relationship between MLSS and alpha. Ahmad et al. (2021), Rosso *et al.* (2005) and Henkel *et al.* (2011) all compared alpha to the SRT; however, while Rosso argues for an exponential relationship, Henkel determines a linear relationship is more representative using a collection of data sets, including Rosso's. Finally, Jiang *et al.* (2017) argues that COD can be used to estimate alpha using an inverse relationship. Ahmed *et al.* (2021) confirmed the existence of the COD-alpha relationship and further investigated the correlation between the alpha factor and the real-time sCOD in the bioreactor level using data obtained from sequencing batch reactors.

Table 2-3: Empirical relationships used to estimate alpha based on dynamic conditions (MLSS: concentration of mixed liquor suspended solids, SRT: solids retention time, Q_{air}: airflow flow rate, COD: chemical oxygen demand)

Empirical Relationship	Definitions	Source
$\alpha = e^{-0.0771 \cdot MLSS}$	MLSS [0-30g/L]	(Gunder, 2001)
$\alpha = e^{-0.08788 \cdot MLSS}$	MLSS [0-30g/L]	(Krampe & Kreauth, 2003)
$\alpha = 0.505 - 0.062 \cdot MLVSS + 0.019 \cdot SRT$	MLVSS [1-12 g/L] SRT [1-30 days]	(Henkel et al., 2011)
$\alpha = 0.172 \cdot \log \frac{SRT}{Q_{air}} - 0.131$	Q_{air} [m^3/s] SRT [day]	(Rosso et al., 2005)
$\alpha = e^{(-1.82 \cdot 10-3 \cdot COD - 0.213)}$	COD [mg/L]	(Jiang et al., 2017)
$\alpha = \left(u/(u-v)\right) \cdot (e^{-v \cdot MLSS} - e^{-u \cdot MLSS})$	MLSS [0-30 g/L] u = 0.507248767 v = 0.1043568988	(Baquero-Rodriguez et al., 2018)
$\alpha = 4.275 \cdot sCOD^{-0.557}$	sCOD[mg/L]	(Ahmed et al., 2021)

In some studies, portions of the relationship curve determined are predominantly influenced by a single dataset. For example, in Figure SD 1c), Baquero-Rodriguez bases the 0-4 CODg/L portion of the curve on data from a single study which does not overlap with any data from the other studies used. In Figure SD1g) the same phenomenon is observed in which Jiang's data is segregated into two distinct sources which do not overlap, one set of data determining the 200-600 CODmg/L portion of the curve and the other study determining the 700-1700 CODmg/L portion of the curve. This illustrates the need for additional data that spans the proposed relationship's entire range to ensure consistency.

2.3 Energy Consumption by Aeration Process

2.3.1 Aeration Energy Consumption

Energy usage represents a significant percentage of treatment plant operation costs. The energy consumption of treatment plants is rising due to more rigorous effluent quality requirements demanding additional treatment, enhanced treatment of biosolids to reduce landfill volume, ageing collection systems allowing inflow and infiltration and leading to higher volumes of raw wastewater and increasing electricity rates (EPA - Office of Wastewater Management, 2010). In 1999 the annual energy costs of a treatment plant were reported to contribute to 5-10% of the overall operating budget (Novak, 1999). Based on a study in 2016, the average energy costs of a treatment plant represent 26% of the total operating budget, as seen in Table 2-4 (Lorenzo-Toja, 2016). This means that energy requirements have more than doubled in the last 17 years.

Moreover, of the energy demand, the majority is consumed by the aeration process. The aeration process typically takes 50-56% of WWTPs' operational energy (Government of Ontario, 2016). This energy-intensive process requires mechanical energy to provide head to wastewater or air. The typical distribution of energy consumption in an activated sludge treatment plant in Ontario is shown in Table 2-5.

Table 2-4: Average Operation Cost Breakdown based on 22 European Wastewater Treatment Plants Source: (Lorenzo-Toja, 2016)

Category	% of Total Cost
Materials	6.19%
Chemicals	2.18%
Energy	26.16%
Personnel	27.67%
Waste	5.43%
Fees	13.34%
Maintenance	18.29%
Lab Analysis	0.73%

Table 2-5: Typical Energy Consumption Distribution within WWTPs Source: (Government of Ontario, 2016; Panepinto et al., 2016; Henriques & Catarino, 2017)

Process	Typical Energy Consumption
Screening & Grit Removal	1.5%
Clarifiers	1.2-3.3%
Aeration	50-56%
Solids Dewatering	0.1-15%
Solids Digestion	14%
Tertiary Treatment	0.3-0.5%
Lighting & Buildings	8%
Pumping & Conveyors	9-15%
Nutrient Removal	8%

2.3.2 Aeration Energy Optimization

Aeration provides DO for aerobic organisms performing BOD removal and nitrification. In this process, oxygen availability is a rate-limiting step; therefore, sufficient DO must be present in the mixed liquor to maximize the treatment capabilities of the organisms. Providing oxygen to the system is energy intensive. Therefore, DO should be held at a

minimum in which it is no longer rate-limiting so that increasing oxygen concentration would have little to no impact on plant performance. It was reported that between the concentration of 4-8 mg/L of DO, COD, and NH_3 removals reach their highest level, that is, 93% and 83% (Meng et al., 2017). However, if energy cost is considered, the range of DO that provides the best performance-to-price ratio drops to 2-4 mg/L (Meng et al., 2017).

Traditionally, controllers are implemented to maintain a desired DO concentration or treatment level using feedback and/or feedforward control philosophy (Table 2-6). However, such control systems often involve continuously changing the airflow input based on the control parameter. It is also difficult to estimate the airflow demand for a longer timeframe. These deficiencies can all be resolved if we can accurately estimate the required airflow input based on a pre-set desired DO standard. In this way, we can easily adjust the airflow input based on the current operational condition of the plant and the DO needed, cutting out unnecessary energy waste during the aeration process.

Table 2-6: Aeration Controller Mechanisms, listed in order of lowest to highest complexity/control (Amand et al., 2013; Rieger et al., 2014)

	Control Mechanism	Equipment Required
1)	Manual Aeration Control	None
2)	DO Cascade Control	DO sensor
3)	Ammonium Feedback Control *	NH4 and DO Sensor
4)	Ammonium Feedforward Control *	Model and NH4 and DO Sensor
5)	Advanced Controllers	Advanced Model and Applicable Sensors

2.4 Machine Learning (ML)

2.4.1 What is ML?

As defined by the pioneer of artificial intelligence research, Arthur Samuel, “machine learning (ML)” refers broadly to the many methods that give the computer the ability to learn without being explicitly programmed (Vestby, 2020).

Nowadays, ML is being increasingly integrated into all research areas and industries because of its irreplaceable ability to automate manual labor and to discover patterns hidden in huge amounts of data.

2.4.2 What can ML do?

The machine learning approach of model building is especially preferable when the relationships among the various features of the source data are too complex or are unclear, as ML can be used to build black boxes without the necessity to have prior knowledge of the relationships described mathematically in detail.

Relationships in the wastewater treatment processes often possess these properties: the physio-chemical interactions processes of accumulation, transportation and degradation of the pollutants are usually unique for each wastewater treatment plant and are not usually well-understood.

The data-driven ML models can replace the labor-intensive manual calibration process of the kinetic models traditionally used in the field with a more customized and time-efficient pipeline.

2.4.3 How does ML Work?

The establishment of a ML model can be divided into three processes: training, validation and testing. Similar to the learning process of humans, ML also starts with gaining information from the input training data and finding patterns with it. During this process, the parameters inside the algorithm are automatically adjusted to improve its performance in estimating the correct outcome. This initial process is known as the learning process. Usually there is also a separate collection of data called the validation dataset that is used

to monitor but not influence the training process to continuously evaluate the performance of the algorithm. Finally, a testing dataset is fed into the ML algorithm to test whether the algorithm works correctly. If the predictions made during this process and the pre-defined answers do not match, the algorithm would be continuously re-trained until the desired outcome is obtained. Otherwise, the ML model developed in this process would be reported as a satisfactory solution and can be used with more data to make predictions.

2.4.4 Key Concepts in ML

To further understand the techniques behind ML, we must first introduce a few key concepts within this field.

2.4.4.1 Hyperparameters

Hyperparameters are parameters whose value is used to control the learning process. The values of the hyperparameters do not change during the learning process, but they can be adjusted manually for each model before the training begins.

Some common examples of hyperparameters include the number of threads used in building the models (how much computational power can be used by the model training process) and learning rate for gradient descent models (how quickly the model is adapted to the problem).

2.4.4.2 2 Major Approaches in ML: Supervised and Unsupervised Learning

There are two major approaches in ML, supervised and unsupervised learning. The two approaches differ in whether a set “correct answers” are there for the ML algorithms to learn from.

Supervised: Supervised learning is characterized by the process of fitting a model to data that has been labelled, I.e., data that is already tagged with the correct answer. Most of the wastewater modelling problems fall into this category. Examples include water effluent quality predictions (Guo et al., 2015) and wastewater treatment process efficiency (Harpaz et al., 2022). In both cases, the raw data is already tagged with plant measurements or efficiency factors that are derived ultimately from observations.

Unsupervised: On the contrary, unsupervised models work on their own to discover information, often from unlabeled data. Unsupervised learning can help researchers to find features which can be useful for categorizations, but it is generally more computationally complex and is less accurate and trustworthy than supervised learning. Unsupervised learning can be used to detect pH anomalies in WWTPs (Gigante et al., 2021) and to select the most efficient aeration strategies based on weather (Borzooei et al., 2020).

2.4.4.3 3 Major Types of ML Problems: Classification, Regression and Clustering Problems

When judging from the type of output variable and the type of problem that needs to be addressed, we can roughly divide the ML algorithms into three categories: classification, regression and clustering.

Classification is a supervised technique that predicts the category to which a new data belongs based on its experience gained from existing data that are labelled. Classification can be used to estimate categorical variables, for example the state of the activate sludges wastewater treatment processes (Khan et al., 2018).

Regression technique predicts a single, continuous output value using the training data. This technique can be used to predict a variety of wastewater quality indicators such as the biochemical oxygen demand (BOD), chemical oxygen demand (COD), and total suspended solids (TSS) (Granata et al., 2017). Due to its ability to predict parameters that gives a measure of the main pollutants in the wastewater and therefore directly reflecting the performance of the treatment process, regression models are vastly studied and used in the field of wastewater modelling.

Clustering is the major unsupervised machine learning algorithm type that we would discuss here. It explores input data by identifying clusters of data points with shared similarities. Implications of the clustering algorithm were discussed in the unsupervised ML section and therefore will not be repeated here.

2.4.4.4 2 Sources of Poor Predictions: Overfitting and Underfitting

Poor ML model performance is usually caused by either overfitting or underfitting. The goal of supervised ML models is to approximate a target function that maps input variables to an output variable.

Overfitting happens when a model is too adapted to the training data, learning noise to the extent that it negatively impacts the performance of the model on other datasets. This is often caused by excessive flexibility (e.g., too many parameters) used in the model.

On the contrary, underfitting happens when the model fails to capture the relationship between the input examples and the target values. It is usually a problem that can be easily detected given a good performance metric. Underfitting can usually be solved by loosening the time requirements of the training process, adding more parameters, or switching to a more suitable model type for the data.

2.4.5 Different Types of ML Models

We will finish this section on ML by briefly discussing several key ML methods, with an emphasis on their strengths, weaknesses and potential use cases in wastewater treatment studies.

Support vector machine (SVM) is a supervised ML model that solves classification problems. It transforms the raw data into an n-dimensional (n is number of features) space where each feature in the input dataset is tied to a particular set of coordinates. It then separates the data with a gap that is made as wide as possible. Compared to newer

algorithms like neural networks, SVM is generally faster and may perform better when there is a limited amount of data to analyze. SVM has been widely used in wastewater research. Many WWTP operational parameters, such as nitrate, nitrite, BOD and TSS, have aroused great interest in real-time predictions due to the potential benefits they could bring to the wastewater control systems. This requirement meets just the strengths of SVM and it has been successfully applied (Yang et al., 2006; Ribeiro et al., 2013).

Decision tree is one of the most popular ML algorithms in use nowadays. It is a supervised classification algorithm that works well with both categorical and continuous target variables. It follows a flowchart-like structure by continuously splitting the input data into many homogeneous sets based on their most significant attributes. Decision tree models are white-box models, that is, the results are usually simple to understand and interpret. It also handles nonlinear relationships between parameters well. However, decision trees are sensitive to small variations in the data and often require methods such as bagging or boosting to reach higher accuracy. To our knowledge, pure decision trees are not as widely used as the more complex models as follows, however, studies have shown that these models are able to make accurate predictions on operational conditions of the plant such as bulking sludge (Atanasova & Kompare, 2002; Deepnarain et al., 2019).

Gradient boosting is an ensemble learning algorithm that combines the predictive power of several weak base estimators (typically decision trees) to improve robustness. It can be

used with regression and classification tasks, among others. Gradient boosting models are usually blessed with higher accuracy; they typically outperform random forests when decision trees are used as the weak learners. However, its advantages come with the sacrifices of intelligibility and interpretability. Gradient boosting models usually have a higher computational demand as well. The gradient boosting method has been reported to achieve great accuracy in estimating the efficiency in anaerobic bio-waste and co-digestion facilities (De Clercq et al., 2019).

The k-nearest neighbour algorithm (KNN) is a supervised learning method used for both classification and regression, although it is used more frequently to solve classification problems in the industry. KNN works by storing all available cases and classifying new cases by taking a majority vote of its k neighbours. KNN models adapt instance-based learning. It, therefore, does not explicitly require a training step and can constantly evolve. However, KNN requires all variables to be normalized and pre-processed to avoid bias in the result. It is also computationally expensive. KNN has been used to predict the flow rate and quality of water flow in WWTP (Kim et al., 2016; Wang et al., 2022). Wang et al. Has reported that kNN and gradient boosting decision tree have outperformed many other algorithms including SVR and decision tree in the effluent water quality prediction models (Wang et al., 2022).

2.5 Research Gap

Although the specific type of fine bubble aerators used throughout the studies is not consistent, Rosso noted that the type of fine bubble diffuser used (ceramic dome, disk, plate, or membrane tube, disk or panel) had no impact on alpha (Rosso et al., 2005). Additionally, except for studies that consisted of consolidation of results and the method for measuring alpha was not explicitly mentioned, all studies conducted after (and including) 2005 report using the off-gas method for determining alpha. Finally, none of the studies that propose an empirical relationship are limited to one dataset; they all involve the evaluation of more than one treatment plant.

There is a distinct lack of confirmation through additional studies for any of the proposed relationships (Amerlink et al., 2016) that implement the relationship determined by (Rosso et al., 2005) and conclude that direct measurements of alpha are significantly superior in improving model accuracy. Henkel, Cornel and Wagner (2011) and Baquero-Rpdriguez, et al. (2018) also conclude that there is a need for validation and the procurement of additional data sets.

A complete and easy-to-use framework for dynamic aeration modelling has not yet been widely adopted in the practical application of the wastewater treatment processes. Though empirical equations describing the influence of potential characteristics of the wastewater on the various factors involved in the aeration process exist, optimization of the energy use in the plants requires an accurate and robust prediction of the alpha-factor and an estimate of the airflow needed to meet effluent wastewater standards. The purpose of this study is

to establish such a framework and apply it to estimate the values of the dynamic alpha-factor and the aeration airflow with operational data through a case study in the Adelaide WWTP.

Chapter 3

3 Dynamic Alpha Modelling

3.1 Introduction

Aeration accounts for 50 to 70% of the energy cost in activated sludge processes treating municipal wastewater (Henderson, 2002; Schierholz et al., 2006; Zuluaga-Bedoya et al., 2018). While the primary function of aeration is to supply oxygen to satisfy the aerobic bacteria demand, it also provides the desired complete mixing in activated sludge processes. Typically for activated sludge processes, aeration systems are designed to maintain at least a dissolved oxygen (DO) concentration of 2 mg/L (Meng et al., 2017). Combined with subsequent processes, contaminants such as organics and nutrients can be removed from the wastewater. However, an over-aeration approach vastly practiced in wastewater operations increases energy demand and negatively impacts the overall energy performance of the plants (Drewnowski et al., 2019). Over the past decades, considerable efforts have been made to design and modify the aeration process to reduce energy costs. The majority of the modern WWTPs utilize feedback control loops to control the DO. However, feedback control is often vulnerable to time delay and disturbance, especially at larger plants where data collected at one point often is not representative of the whole tank (Åmand et al., 2013). Alternatively, the concept of dynamic aeration modelling can further be utilized for aeration process control and optimization (Martin & Vanrolleghem, 2014). In this case, the dynamic aeration models were used to build feedforward control loops to

predict DO or required airflow based on real-time or forecasted influent data (Åmand et al., 2013). Martin and Vanrolleghem (2014) demonstrated that by modifying the control system in WWTPs, the total energy cost of aeration could be saved by 27% on average. However, one of the standing research questions is the accuracy of the aeration models to capture dynamic aeration demand within the process.

The most accepted aeration models are based on the oxygen mass transfer model (Equation 3.1). The transfer resistance at the gas film is generally considered negligible due to the low oxygen solubility in water (Garcia-Ochoa & Gomez, 2009). Therefore, the driving force for oxygen transfer in this model solely depends on the concentration gradient between the oxygen saturation concentration (C*) and the actual oxygen concentration (C) that is typically reported as DO in routine measurements. The overall oxygen mass transfer coefficient (K_La) partially reflects the difference in oxygen transfer rate in wastewater and clean water. The K_La in wastewater can be expressed with equation 3.2 by introducing several correction factors (Mueller, Boyle & Popel, 2002). Table 2-2 (in chapter 2) lists these correction factors and their typical values in municipal wastewater.

$$OTR = K_L a(C^S - C) \quad (Equation\ 3.1)$$

$$(K_L a)_{ww} = \alpha\beta\gamma(K_L a)_{20°C} \quad (Equation\ 3.2)$$

Where

OTR	Oxygen transfer rate	[g/h]
$K_L a$	Overall oxygen mass transfer coefficient	[h^{-1}]
C^S	Saturation concentration of oxygen in test liquid under equilibrium	[mg/L]
C	The concentration of dissolved oxygen in test liquid	[mg/L]
$(K_L a)_{ww}$	Overall oxygen mass transfer coefficient of wastewater	[h^{-1}]
α	Alpha factor	[-]
β	Beta factor	[-]
γ	Gamma factor	[-]
$(K_L a)_{20°C}$	Overall oxygen mass transfer coefficient of clean water at standard condition	[h^{-1}]

The alpha, beta, and gamma factors are directly involved in calculating the overall oxygen mass transfer rate; these factors can be adjusted to improve the aeration models (equation 3.2) in supporting document). The alpha factor (α) is defined as the volumetric mass transfer ratio of oxygen between wastewater and clean water. The beta factor (β), also known as the salinity-surface tension correction factor, is the ratio of saturation DO concentration between wastewater and clean water. The beta factor represents the impact

of constituents, such as salts and suspended solids, on the solubility of dissolved oxygen (Pittoors et al., 2014). It is well-studied that beta only influences 0.05% or less of total mass transfer so that it can be ignored in most cases (Rodríguez et al., 2012). The gamma factor (γ) represents the influence of temperature (T) on the oxygen, $\gamma = \theta^{20°C-T}$. The theta factor lies 1.016-1.024 (Water Environment Federation, 2010). Therefore, the gamma factor only depends on the water temperature T within this temperature range. The only and most uncertain correction factor left is the alpha factor, with all other factors eliminated. Alpha factor represents the effect of diffuser design, tank geometry and water constituents on the overall oxygen transfer performance in wastewater (K. Al-Ahmady, 2011). While the alpha factor was traditionally considered a constant value, recent studies suggest that alpha factors fluctuate daily and seasonally and should be considered a dynamic factor (Leu et al., 2009). However, there is still no universally valid approach directly to measure the alpha factor. In practice, the overall oxygen mass transfer coefficient can be indirectly determined through the off-gas method using equation 3.3 (Mueller et al., 2002). This equation is established based on the gas mass balance at a steady state, assuming that both CO_2 and nitrogen are conservative. With this equation, KLa may be estimated based on the measurement values of the total gas volume flow rate of inlet and outlet gases, volumetric fractions of oxygen in them, and DO. The estimation of oxygen mass transfer based on this theory has been referred to as the off-gas method. The off-gas method has proven to offer a practical and accurate estimation under process conditions (Iranpour et al., 2000).Combining equations 3.2 and 3.3, we can calculate the alpha factor indirectly through equation 3.4.

$$\frac{\rho}{V}(q_i Y_R - q_o Y_{OG}) = (K_L a)_{ww}(C^S - C) \quad (Equation\ 3.3)$$

$$\alpha = \frac{\rho(q_i Y_R - q_0 Y_{OG})}{V\theta^{(20℃-T)}(\beta C^S - C)(K_L a)_{20℃}} \quad (Equation\ 3.4)$$

Where

ρ	Density of oxygen at temperature and pressure at which gas flow is expressed	(Constant)	[g/L]
q_i	Total gas volume flow rates of inlet gases	(Dynamic measured)	[m^3/h]
q_o	Total gas volume flow rates of outlet gases	(Dynamic measured)	[m^3/h]
Y_R	Volumetric fractions of oxygen gas in inlet gases	(Constant)	[-]
Y_{OG}	Volumetric fractions of oxygen gas in outlet gases	(Measured or modelled)	[-]
$(K_L a)_{ww}$	Overall oxygen mass transfer coefficient of wastewater	(Calculated)	[h^{-1}]
C^S	Saturation concentration of oxygen in test liquid under equilibrium	(Constant measured)	[mg/L]
C	The concentration of dissolved oxygen in test liquid	(Dynamic measured)	[mg/L]
α	Alpha factor	(Calculated or modelled)	[-]
V	The volume of aeration tank	(Constant)	[m^3]

θ	Theta factor	(Constant)	[-]
T	Temperature of aeration tanks	(Dynamic measured)	[°C]
$(K_L a)_{20°C}$	Overall oxygen mass transfer coefficient of clean water at standard condition	(Constant)	[h^{-1}]

Although entirely accurate, this method is expensive to adapt during operation as it requires additional sensors to measure the volumetric fraction of the outlet gas fraction (y_i). Besides, the alpha factor measured with this method can only represent the water column of the sampled niche, which may not be representative for other regions in the aeration tanks. Therefore, numerous attempts have been made to predict the alpha factor's change with more accessible and cheaper parameters to measure under field conditions (Baquero‐Rodríguez et al., 2018; L. Jiang et al., 2017; Amerlinck et al., 2016). It has generally been agreed that the impact of surfactants is responsible for the dynamic nature of the alpha factor (Rosso & Stenstrom, 2006b; Schierholz et al., 2006). Therefore, many of these empirical (regression) models utilized routine measurements that reflect the level of surfactants such as mixed liquor suspended solids (MLSS) and chemical oxygen demand (COD), as input parameters (Amerlinck et al., 2016; Baquero‐Rodríguez et al., 2018; Günder, 2001; L. Jiang et al., 2017; Krampe & Krauth, 2003). However, the resulting predictions are often too inaccurate due to many assumptions and over-simplification that must be made while utilizing these empirical equations.

Like the previous regression models, machine learning (ML) allows engineers to develop data-based analytical models. ML models make predictions based on real-time relationships between input and output data instead of formulated mechanisms. However, the ML models make significantly fewer assumptions and are proved to be more robust, comprehensive, and suitable for representing wastewater process units or components where detailed mechanisms remain unclear (Guo et al., 2015). In such a rapidly developing field, however, numerous ML models exist that each features unique characteristics and has various hyper-parameters to be set for optimal performance. It is increasingly unrealistic for wastewater engineers to incorporate all the best practices in ML model development into their models. The Automatic Machine Learning (AutoML) framework offers an easy-to-use toolset for general ML model developers to flatten the learning curve. AutoML automatically determines the best model and the corresponding hyper-parameters to yield the best performance based on the input dataset. The model development process can be easily replicated or transferred to a completely different WWTP using site-specific input parameters without substantial loss of predictability.

To date, numerous studies have investigated the correlation between the alpha factor and an extensive range of operational parameters, but the existing empirical equations often lack precision in predicting the airflow needed for a single WWTP under dynamic operational conditions. The objectives of this paper were to (i) develop accurate models and alternative approaches to predict dynamic alpha factor in aeration systems; (ii) assess wastewater parameters that influence dynamic alpha factor values; and (iii) assess the optimal timescale for dynamic alpha modelling.

3.2 Methods

3.2.1 Research site description

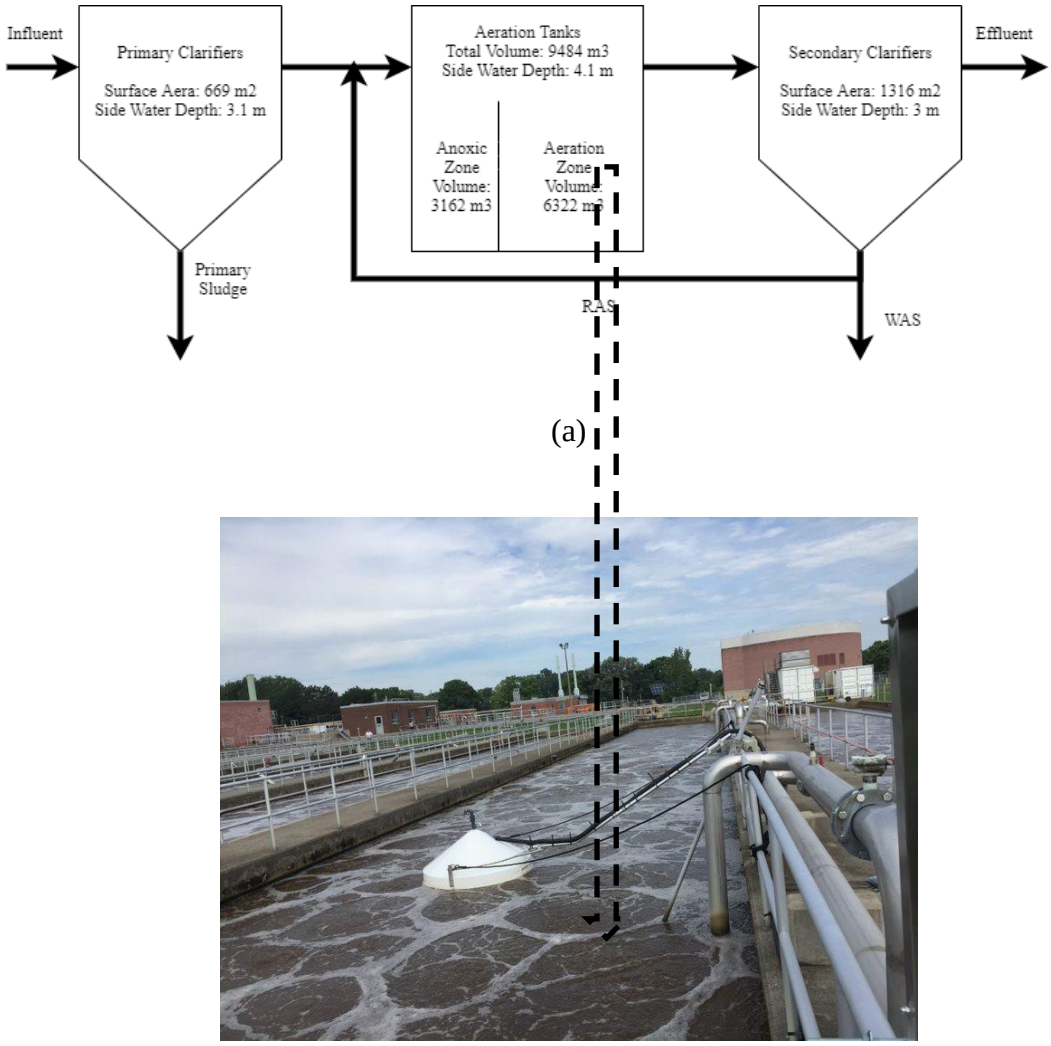
This research was conducted at the Adelaide pollution control plant (PCP) (London, Ontario, Canada). The Adelaide PCP is an activated sludge process treatment plant that was designed for a maximum daily flow capacity of 29,600 cubic meters per day (m^3/d) with an approved peak flow rate of 59,200 m^3/d. In 2017, the plant was operating at 90.8% capacity with an average flow of 26,882 m^3/d (City of London, 2018). The coldest and warmest month in 2017 was March and September, with an average stream temperature of 12.4 °C and 20.7 °C, respectively. The geometric properties of the unit processes of Adelaide PCP are summarized in Figure 3.1a. The wastewater treatment process related to this study consists of primary clarifiers, aeration chambers and secondary clarifiers.

3.2.2 Data Collection

The Adelaide PCP is equipped with DO, ammonia, TSS, phosphorus and density sensors integrated into the Supervisory Control and Data Acquisition (SCASA) System. The data used for the modelling were obtained from two sources: the existing online sensors and a field sampling campaign. The sampling frequency, sampling location and analysis/sensor types are summarized in Table 3-1. The online sensor data were collected every 15 minutes from the raw wastewater channels, primary clarifier effluent channels, aeration tanks, and

plant effluent channels. The average measurement results corresponding to the field campaign period are presented in Table 3-1.

The 50-day field campaign collected dynamic COD, NH_3-N, off-gas oxygen fraction, TSS and VSS data. An autosampler was installed during the campaign to collect hourly samples daily from the primary influent and aeration tanks. Each hourly sample was collected from 4 sub-samples that were grabbed by the autosampler every 15 minutes. The daily-24 field samples were immediately transported to the lab and analyzed for COD (HACH number 8000), NH_3-N (HACH number 10031) and TSS/VSS (US EPA Method 1684). During the field campaign, an INVENT ALPHAMETER® (32013_REV 2018) was also deployed at one of the aerations tanks near the DO sensors to sample the off-gas oxygen concentration released from the aeration tank (Figure 3.1b). The alphameter has a net surface intake of $1m^2$ and was used to take measurements at 3/4 part of the tank (Figure 3.1b). The alphameter was connected to the measurement array on the walkway via PVC tubing and was equipped with a dedicated data acquisition system. The constant saturation DO (C^S) used in this study was also measured at 3/4 part of the tank.



(a)

(b)

Figure 3-1 Adelaide pollution control plant (a) Layout and (b) Off-gas measuring alpha-meter in place in Tank 5 of Adelaide.

Table 3-1 Adelaide WWTP Sampling Frequency, Location and Analysis Methods.

Location	Parameters	Unit	Interval	Source	Method	Avg	Stdev	median	max	min
Plant influent	Liquid flow rate	m^3/d	15 minutes	Adelaide WWTP	SCADA online	287.0	49.7	284.3	712.5	162.0
Post-primary clarifier/Aeration influent	Chemical oxygen demand	mg/ L	1 hour	Field Campaign	HACH Method (2125915-CA)	277.6	44.4	274.0	406.0	126.0
	Phosphate	mg P/L	15 minutes	Adelaide WWTP	SCADA online measurements	3.5	0.7	3.4	8.4	0.1
	Ammonia	mg N/L	15 minutes	Adelaide WWTP	SCADA online	13.7	3.9	14.3	22.7	1.7
	Total suspended solids	mg/ L	15 minutes	Adelaide WWTP	SCADA online	145.0	160.7	141.6	7708.4	59.9
Aeration tank	Mixed liquor suspended solids	mg/ L	15 minutes	Adelaide WWTP	SCADA online measurements	1380.4	61.9	1380.8	2339.6	7.7
	Ammonia	mg N/L	15 minutes	Adelaide WWTP	SCADA online	3.7	2.5	3.4	12.1	0.5
	Temperature	°C	15 minutes	Adelaide WWTP	SCADA online	20.5	0.6	20.5	21.6	19.1
	Airflow	m^3/d	15 minutes	Adelaide WWTP	SCADA online	12565.8	2158.2	13211.4	16366.6	5652.2
Other	Off-gas oxygen fraction	%	15 minutes	Field Campaign	(InventTM Alpha-Meter)	19.2	0.4	19.3	19.9	17.9

3.2.3 Outlier Removal

All the collected data were first preprocessed by the density-based cluster analysis "Density-Based Spatial Clustering of Applications with Noise (DBSCAN)" filter for outlier removal (Ester et al., 1996). As a density-based clustering algorithm, DBSCAN classifies a set of points based on their distance to each other. Points that lie alone in low-density regions are marked as outliers. DBSCAN requires the user to define two factors: eps and minPts. The minPts, the minimum number of neighbours within a defined radius "eps," were set to be 5 for all datasets treated. The values of eps were chosen by finding the first strong bend on a k-distance graph for each dataset. This paper performed DBSCAN with the Python package scikit-learn (Pedregosa et al., 2011). Missing data were treated with the mean substitution method, as suggested by Kang (2013).

3.2.4 Parameter Calculation

The alpha factor (α), hydraulic retention time (HRT) and soluble chemical oxygen demand (sCOD) were calculated. The α factor was calculated as the overall transfer coefficient (K_La) ratio between wastewater and clean water, based on the off-gas oxygen fraction obtained during the field campaign and other plant operational data (Equation 3.4). HRT was calculated as (volume of aeration tank)/(influent flow). Due to the lack of soluble chemical oxygen demand (sCOD) sensor on-site, our raw dataset was complemented with sCOD estimations made by an aeration tank model generated with GPS-X version 7.0 (Hydromantis, 2017). The GPS-X model was built following the procedures and specification listed in the Adelaide Operational Report provided by

City of London (Hydromantis, 2013). It was calibrated with the daily operational measurements retrieved from Adelaide WWTP from September 2018 to October 2018 to reflect the condition of the plant during the studied period, the sampling frequency, sampling location and data analysis are summarized in Table 3-2. After visual validation, raw dataset obtained in the sampling campaigns were fed into the GPS-X model to generate the sCOD estimations. A summary of the estimated sCOD dataset is presented in Table 3-2. The GPS-X model export operational parameters listed in appendix figure SD4 (a) to SD4(p).

Table 3-2 Adelaide WWTP monthly operational report dataset

Location	Parameters	Unit	Interval	Avg	Stdev	Median	Max	Min
Plant influent	Liquid flow rate	m3/d	1d	25360.2	2814.8	25016	40823	19884
	Biological oxygen demand 5	mg/L	Twice per week	258.5	59.6	267.0	365.0	132.0
	Total suspended solids	mg/L	Twice per week	309.2	98.9	313.0	472.0	148.0
	Total kjeldahl nitrogen	mg/L	Once per week	50.6	8.3	48.2	64.0	42.1
	Ammonia	mg/L	Twice per week	29.6	6.7	26.6	43.9	23.3
	Phosphate	mg/L	Twice per week	8.4	1.3	8.2	10.6	6.4
	pH	-	Twice per week	7.5	0.2	7.5	7.8	7.2
Post-primary clarifier/Aeration influent	Biological oxygen demand	mg/L	Twice per week	168.4	74.0	146.0	342.0	98.4
	Total suspended solids	mg/L	Twice per week	114.8	41.2	103.0	228.0	59.0
Aeration tank	Airflow	m3/d	1d	13240.0	1625.1	13153.0	17561.0	9678.0
	Temperature	°C	1d	19.7	0.7	19.6	21.4	18.0
	Mixed liquor suspended solids	mg/L	Twice per week	1427.1	132.8	1480.0	1600.0	1200.0

	Mixed liquor volatile suspended solids	mg/L	Twice per week	1074.8	111.0	1110.1	1212.8	867.6
	Return activated sludge flow	m3/d	1d	20560.8	1386.4	20465.0	25886.0	16705.0
	Return activated sludge concentration	mg/L	Twice per week	2795.3	447.8	2820.0	3760.0	2180.0
	Return activated volatile sludge concentration	mg/L	Twice per week	2126.9	354.9	2102.4	2861.4	1584.9
	Waste sludge flow	m3/d	1d	1158.6	370.0	1263.0	1777.0	505.0
	Dissolved oxygen	mg/L	1d	4.7	0.6	4.6	6.5	3.8
Plant effluent	Biological oxygen demand 5	mg/L	Once per week	1.1	0.2	1.0	2.0	1.0
	Total suspended solids	mg/L	Once per week	4.9	5.7	1.5	22.0	1.5
	Total kjeldahl nitrogen	mg/L	Once per week	1.7	0.4	1.5	2.5	1.3
	Ammonia	mg/L	Once per week	0.4	0.4	0.2	1.2	0.1
	Nitrite	mg/L	Once per week	0.4	0.2	0.4	0.8	0.2
	Nitrate	mg/L	Once per week	10.9	2.1	9.7	14.4	8.5
	Phosphate	mg/L	Once per week	8.4	1.3	8.2	10.6	6.4
	pH	-	Once per week	7.3	0.1	7.3	7.5	7.0

Table 3-3 Exported sCOD data summary from GPS-X model

Location	Unit	Interval	Avg	Stdev	median	max	min
Post-primary	mg/L	1h	142.7	23.6	142	205	64.4
Aeration Tank	mg/L	1h	50.4	9.7	49.4	83.1	20.9

3.2.5 Optimal Sampling Frequency Determination

The optimal timescale of the input dataset for model development was assessed by averaging and transforming the original data recorded every 15 minutes intervals to 1 hour, 2 hours, 3 hours, 4 hours, 6 hours, and 8 hours. The resulting new datasets were used to generate pioneer AutoML models, and ten replicate models were built for each timescale tested. The accuracy of each model was calculated to quantify the effect of sampling frequency on the model's performance.

3.2.6 Auto-Machine Learning Processes

The AutoML models were developed with the AutoGluon-Tabular package, an efficient and easy-to-use Python framework developed by Amazon Web Services (AWS) (Van Rossum & Drake Jr., 1995; Erickson et al., 2020). The input parameters used to develop the model, and their value range

is summarized in Table 3-1. The performance of the models is evaluated by the root mean squared error (RMSE) during training. The train-test-split configuration was set as a training: testing ratio of 8:2. As the dataset was initially arranged by time, the order in which data was fed per epoch (iteration) could potentially influence the model, increasing the training and testing error. Each split was randomly shuffled before each epoch to avoid bias during training and testing datasets.

3.2.7 Regression Model Development

In addition to the AutoML models, existing exponential alpha models were explored per the models developed by Krampe and Krauth (2003) and Jiang *et al.* (2017) using the current dataset. The models were expressed as:

$$\alpha = A * e^{-K_a * [Parameter]} \quad (Equation\ 3.5)$$

Whereby A and K_a are constants adjusted by residuals minimization.

α	Alpha factor	(Calculated)	[-]
A	Constant adjusted by residuals minimization	(Constant)	[-]
K_a	Constant adjusted by residuals minimization	(Constant)	[-]
$Parameter$	Operational parameter of wastewater treatment process	(Dynamic measured)	[mg/L]

Whereby A and K_a are constants adjusted by residuals minimization.

The current study used sludge and wastewater characteristics, including MLSS and COD, as input parameters to calibrate existing regression models. The considered regression models include MLSS based by Baquero-Rodríguez et al. (2018), Günder (2001) and Krampe and Krauth (2003) and COD based by Jiang *et al.* (2017) (Table 3.2).

3.2.8 Parameter Sensitivity Analysis

The permutation-shuffle importance score built in the AutoGluon package was used to determine the sensitivity of the AutoML models to each available input parameter (Van Rossum & Drake Jr., 1995; Erickson et al., 2020). The permutation importance score is defined as the decrease in model performance (as in RMSE in this study) when a single feature value is randomly shuffled (Breiman, 2001). All permutation importance scores were calculated on the validation set in this study. The modelling approach is summarized in Figure 3.2.

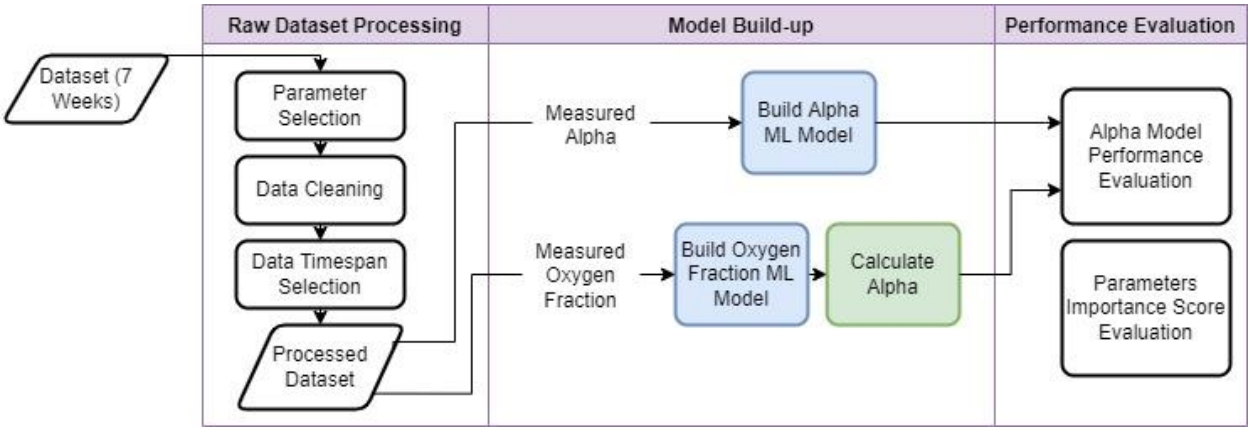


Figure 3-2 Alpha modelling process flow chart

3.2.9 Model Performance Evaluation

To assess the performance of each model, the researchers followed the procedures for goodness-of-fit measures in dynamic wastewater modelling, suggested by Ahnert *et al.* (2007). After visual evaluation and tests for fitting the time-dependent behaviour, the three criteria: Root Mean Square Error (RMSE), Nash-Sutcliffe efficiency (NSE), and Index of Agreement (d) were applied (N.ash & Sutcliffe, 1970; Willmott, 1981):

$$RMSE = \sqrt{\frac{\sum_{t=1}^{T}(x_m^t - x_o^t)^2}{T}} \quad (Equation\ 3.6)$$

$$NSE = 1 - \frac{\sum_{t=1}^{T}(x_m^t - x_o^t)^2}{\sum_{t=1}^{T}(x_o^t - \overline{x_o})^2} \quad (Equation\ 3.7)$$

$$d = 1 - \frac{\sum_{t=1}^{T}(x_O^t - x_m^t)^2}{\sum_{t=1}^{T}(|x_m^t - \overline{x_o}| + |x_o^t - \overline{x_o}|)^2} \quad (Equation\ 3.8)$$

where

x_m^t	modelled parameter at time t	[-]
x_o^t	observed parameter at time t	[-]

3.3 Results and Discussion

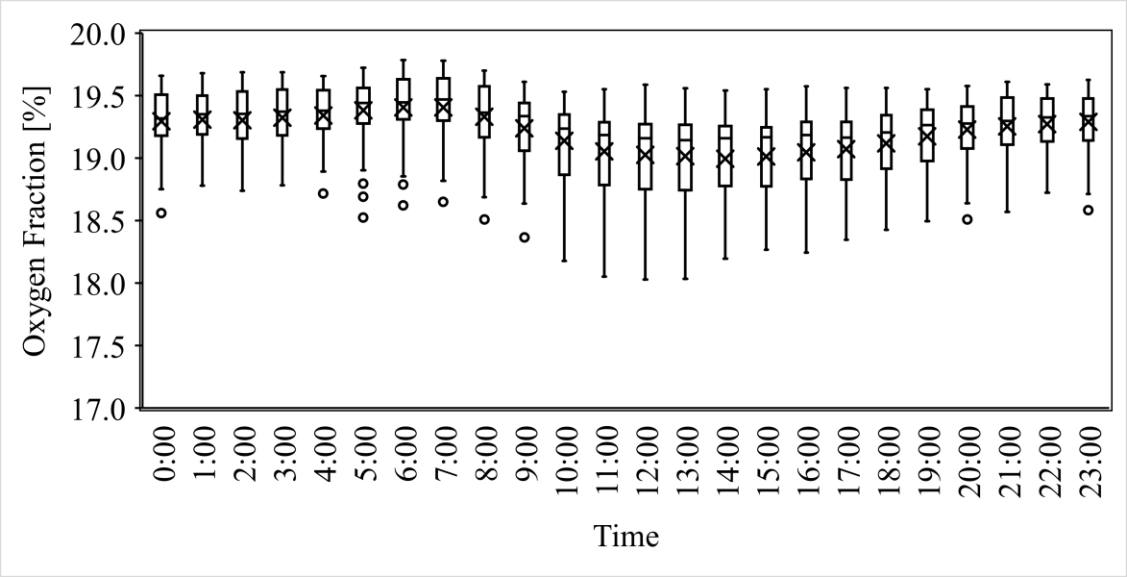
3.3.1 Data Preprocessing and Wastewater Characterization

All operational parameters used in this study were preprocessed with the DBSCAN clustering algorithm to remove outliers. After outlier removal, the average hourly data with standard deviation were plotted to assess each parameter's variability within 24 hours. The average value of off-gas oxygen fraction, airflow, aeration temperature, HRT, MLSS, COD, sCOD, primary effluent phosphorus and ammonia are 19.21±0.32%, 12545.79±1903.46 m^3/h, 20.46±0.58 C, 5.45±1.01 mg/L, 1380.92±51.97 mg/L, 276.50±46.14 mg/L, 142.70±23.63 mg/L, 3.50±0.59 mg/L and 13.67±3.37 mg/L, respectively (Table 3.1).

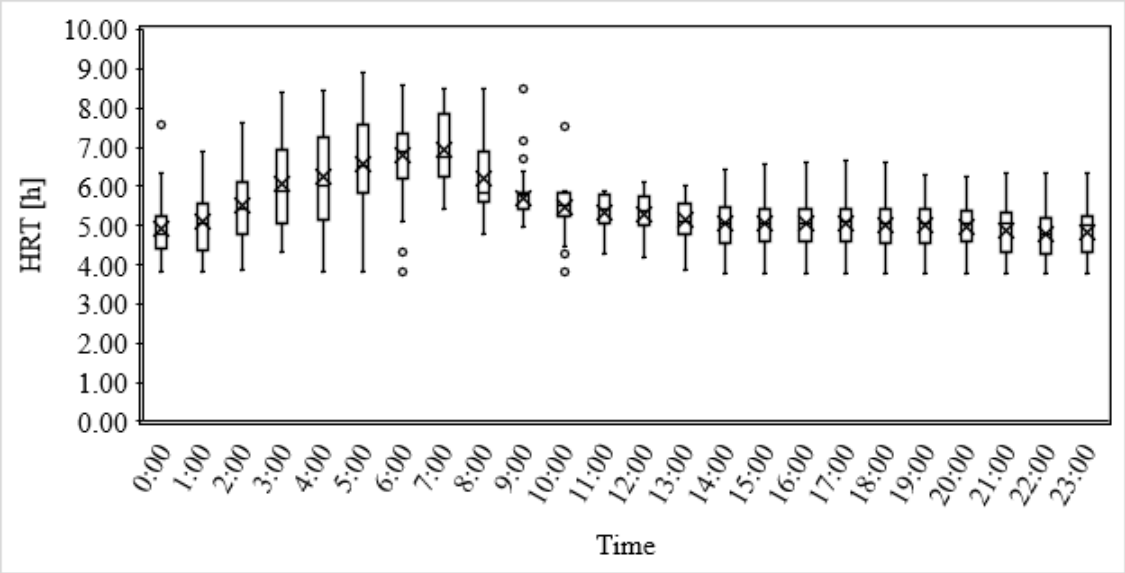
As shown in Fig. 3.2, significant daily fluctuations were observed in the parameter(s) used to estimate alpha, supporting alpha's dynamic nature in aeration systems. The off-gas oxygen fraction and the HRT showed the most robust daily fluctuation trend among the assessed parameters. The off-gas oxygen fraction was the highest early in the morning, reaching its peak at around 7:00 am, then decreasing to its minimum at around 1:00 pm before climbing back. A potential cause of such fluctuation could be the daily change in wastewater influent loading. The higher loading would cause an increased concentration in the aeration tanks, which leads to a lower oxygen concentration in the wastewater and, therefore, a higher oxygen transfer rate at the water surface. This is also supported by the daily average of HRT, which shows a similar trend as the off-gas oxygen fraction. Several other parameters, such as the temperature, MLSS, PO_4 and NH_3, did not show significant daily fluctuations (Figure SD2, in supporting document). This suggests that these parameters may

not reflect the daily fluctuation in the alpha-factor, and therefore models built with them may show limited predictability.

The off-gas fraction, alpha factor, NH_3, MLSS, COD, and temperature data were plotted against time to show the data variability during the field campaign period (Figures SD2a to SD2k). Most of the wastewater and operating parameters showed a higher variability during the field campaign period. The oxygen fraction and alpha-factor varied from 18 to 20% and 0.3 to 0.9, respectively, confirming alpha the dynamic nature of an alpha. When the overall dataset was considered, a strong relationship was observed between HRT, temperature, HRT, TSS and COD with the measured off-gas fraction (Figures SD2c to h).



(a)



(b)

Figure 3-3 Daily dynamic change of operational parameters in Adelaide PCP: (a) HRT and (b) exhaust air Oxygen fraction.

3.3.2 Optimal Sampling Frequency Determination

While developing models to describe wastewater treatment processes, one crucial question is how to deal with the uncertainties within the input dataset. Currently, existing alpha models are often built upon daily, monthly, or yearly average values from several wastewater treatment plants, causing them to be generalizable and lack precision when predicting the dynamic behaviour of a specific plant.

This study conducted an analysis to accurately identify an optimal sampling frequency to describe the oxygen transfer's dynamic behaviour accurately. The analysis of the performance of models was built with seven different sampling frequencies (15 minutes, 1 hour, 2 hours, 3 hours, 4 hours,

6 hours, and 8 hours), each with 5 replicates. The 1-hour interval group performs better than the 15-min interval group. The average RMSE, NSE and index of the agreement for the 1-hour interval group, were 0.06, 0.68 and 0.89, respectively (Figure 3.4). However, increasing the sampling frequency to 15 minutes lowered the model performance. The poor performance at 15 minutes data scale shows that the wastewater data collected at shorter times could be unstable, introducing noise when the sampling frequency was higher. In addition, some data were collected by sensors and analyzers; in this case, the sensors are distributed along with the treatment processes, causing unignorable time lag in some datasets hence supporting deficient performance/agreement at shorter timescales.

Reducing the sampling frequency to 2-hours caused the NSE to drop by 0.08 and the index of an agreement to drop by 0.05. The model performance for the 8 hours sampling frequency was almost halved compared to that of the 1-hour interval models, decreasing the accuracy of the models. The latter result showed that decreasing sampling frequency affected the prediction accuracy, as it may have caused oversimplification of the input data. In summary, the models developed based on the input dataset of 7 sampling frequencies indicated that the optimal sampling interval for alpha modelling in Adelaide PCP is 1 hour.

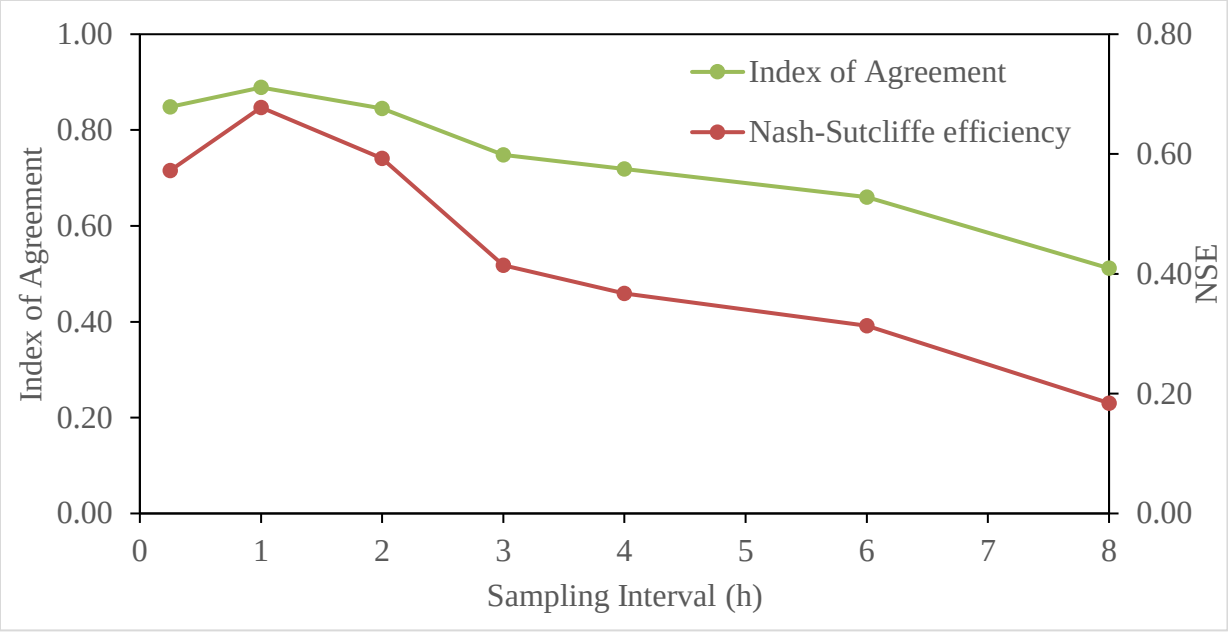


Figure 3-4 Goodness-of-fit (Index of Agreement and Nash-Sutcliffe efficiency) performance of AutoML built with input data of different sampling frequencies.

3.3.3 Model Performance

As aeration is both a critical process and is heavily energy-intensive in wastewater treatment, considerable effort has been made to develop models to predict the dynamic behaviour of the process. Among numerous factors involved in oxygen mass transfer efficiency in aeration, the alpha factor, which is dynamic by nature, is the most uncertain parameter. One popular theory is that the alpha factor is most closely related to the concentration of organic surfactants (Rosso & Stenstrom, 2006b; Sardeing et al., 2006; Gillot & Héduit, 2008). Based on this theory, various attempts to model the alpha factor with parameters reflecting surfactants have been developed

(Günder, 2001; Krampe & Krauth, 2003; Amerlinck et al., 2016; L. Jiang et al., 2017; Baquero‐Rodríguez et al., 2018). Most of these models are empirically derived; however, they often lack precision due to the time-varying and highly nonlinear characteristics of wastewater treatment processes.

On the other hand, machine learning and deep learning are practical and efficient when dealing with complicated systems. Like the previous alpha models, these models do not depend on specific theories and generally do not make assumptions about underlying mechanisms except when selecting input parameters. In this study, the AutoML models were used to develop two different models that are robust with extreme input values, an essential characteristic when dealing with wastewater measurements. The first set of models referred to as the "direct AutoML alpha model" directly estimated alpha from some measurable wastewater parameters (model output=alpha; model inputs=Aeration NH_3, MLSS, HRT, COD, Post Primary TSS, Post Primary COD, Post Primary PO_4). The second set of AutoML models (referred to herein as the hybrid AutoML-mechanistic alpha model) was developed to estimate the off-gas aeration fraction from wastewater parameters and subsequently calculated alpha using the estimated off-gas fraction using equation 3.4. The off-gas fraction was developed using wastewater parameters inputs (aeration temperature, aeration NH_3, MLSS, HRT, COD, Post Primary TSS, Post Primary COD and Post Primary PO_4) and an off-gas fraction as an output.

Figures 3.5(a) to 3.5(c) show the performance of three models built in this study: (a) the direct AutoML alpha model and (b) the hybrid AutoML-mechanistic alpha model, and (c) recalibrated published regression models (regression models are shown in Table 3-2). The results were based

on the 20% test datasets. It can be seen how the hybrid AutoML-mechanistic and AutoML models showed a slight discrepancy throughout the long-term sampling period between the measured (scattered-blue) and modelled (line-orange) alpha factor values, indicating a successful prediction for the long term. The hybrid alpha model (Figure 3.5(b); RMSE = 0.03, NSE = 0.94, d = 0.98) performed better than the direct model (Figure 3.5(a); RMSE = 0.07, NSE = 0.67, d = 0.88), possibly because of the limited ability of the latter AutoML algorithm to model extreme values. Several alpha-factor peaks were often recorded in the afternoon (i.e., from 13:00 to 17:00), reaching a maximum of 0.96. The calculation from the off-gas oxygen fraction to the alpha factor could amplify the extreme values from the oxygen fraction outputs, allowing a better fit for more extreme value points. It is important to recognize that the hybrid alpha model estimated the off-gas fraction using AutoML models and used the modelled off-gas fraction to calculate further the alpha factor based on established mechanistic equations. The modelling approach primarily provided an alternative to the actual measurement of off-gas fractions which were deemed difficult to conduct routinely in the wastewater treatment environment. To the authors' best knowledge, the approach of estimating air fraction has never been considered before and introduced in this study for the first time.

In this study, the empirical equations were recalibrated using the current dataset for Adelaide PCP based on these formulations by fitting the coefficients through residual minimization. The resulting recalibrated models and performance of the models are shown in Table 3-2 and Figure 3.5(c), respectively. Comparatively, the recalibrated regression models reflect only the average of the alpha factor compared to the AutoML models that successfully captured the daily dynamics of alpha.

The alpha values directly estimated with these equations have shown significant deviation from the measured values, possibly because of the difference in the range of input parameters. Three out of the four equations listed above estimated the alpha factors with MLSS measurements (Günder, 2001; Krampe & Krauth, 2003; Baquero‐Rodríguez et al., 2018). The average value for MLSS at Adelaide PCP is 1.4 g/L. Although most studies that reported an Alpha-MLSS relationship claimed to be applicable for ranges between 0-30 g/L of MLSS, the input dataset they used to develop these models are often from high strength wastewater or processes that make use of membranes that yields above 7 g/L for MLSS, which is significantly higher than the measurements at Adelaide PCP (Günder, 2001; Krampe & Krauth, 2003). Baquero-Rodriguez *et al.* (2018) supplemented their input dataset with data collected from a full-scale test by Diago Rosso in 2005, with MLSS in the range of 1-4 g/L. They also introduced a second coefficient in the equation. However, when adjusting the coefficients with the dataset from Adelaide WWTP (this study), it is found that removing the second coefficient results in a better fit, which results in a similar formula to the ones developed by Gunder (2001) and Krampe & Kreauth (2003). In comparison with the models developed with AutoML, it is evident that these suggested empirical models based on MLSS captured less dynamic behaviour of the alpha factor in the long term. Thus, results from such models can only be used to estimate average alpha values that may not be beneficial for dynamic aeration modelling; however, the values can still be beneficial for design purposes that require average alpha values.

Jiang *et al.* (2017) suggested an empirical relationship between the alpha factor and post-primary COD to estimate the dynamic behaviour of the former. The COD raw data they used to train their model was the range of 200-1600 mg/L (Jiang et al., 2017), which covers the COD level of 130-

400 mg/L in the Adelaide plant. Although developed in a similar range, this model also did not exhibit a strong enough prediction of the dynamic alpha factor. One possible explanation is that COD at Adelaide PCP lacked variation in value over the long term, which affected its ability to act as an indicator of the alpha factor. Thus, the complexity of the water condition in full-scale wastewater treatment facilities such as Adelaide PCP might require the involvement of multiple parameters to make an accurate prediction of the alpha factor.

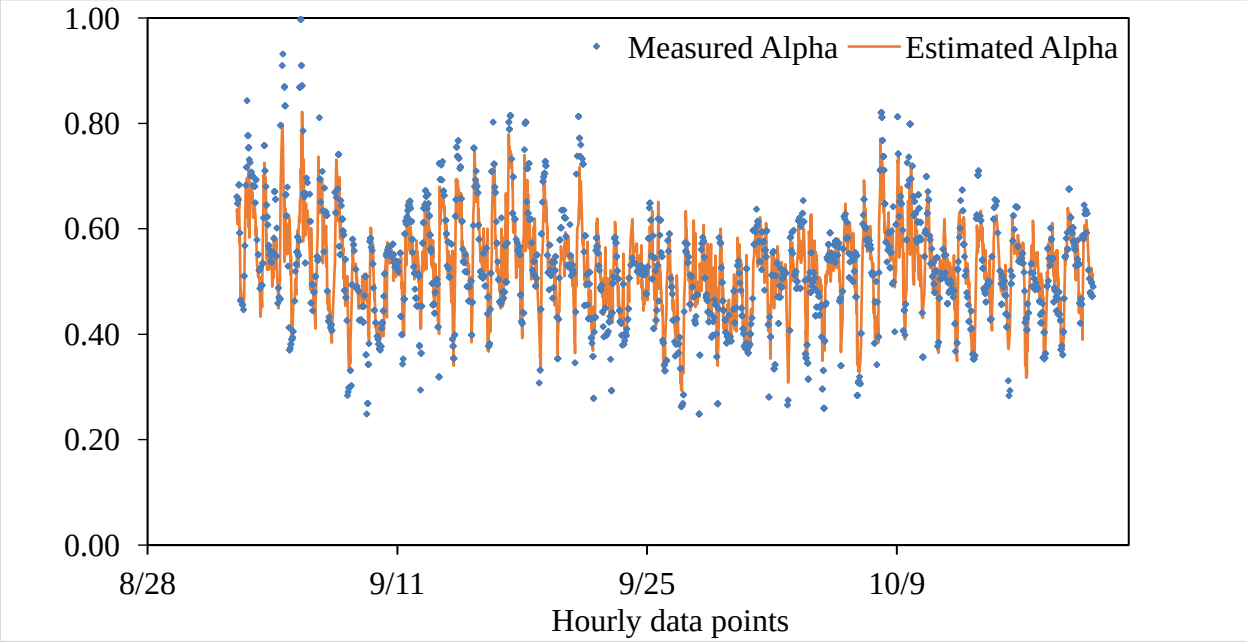
A novel approach using the bioreactor sCOD as an indicator of the dynamic changes in the alpha factor was recently reported by Ahmed et al. (2021). Our investigation demonstrated that although showing some short-term dynamics in alpha, this empirical equation still fails to capture the extreme measurements. Moreover, potentially due to a combined effect of the lack of coverage in sCOD values over the entire study period and the fact that the sCOD-alpha correlation equation was developed in batch rather than WWTP, it also tends to be fluctuating around a constant, average alpha level, instead of reflecting its dynamic long-term change.

Table 3-4 Re-calibrated published regression equations for dynamic alpha-factor modelling (MLSS: concentration of mixed liquor suspended solids, COD: chemical oxygen demand)

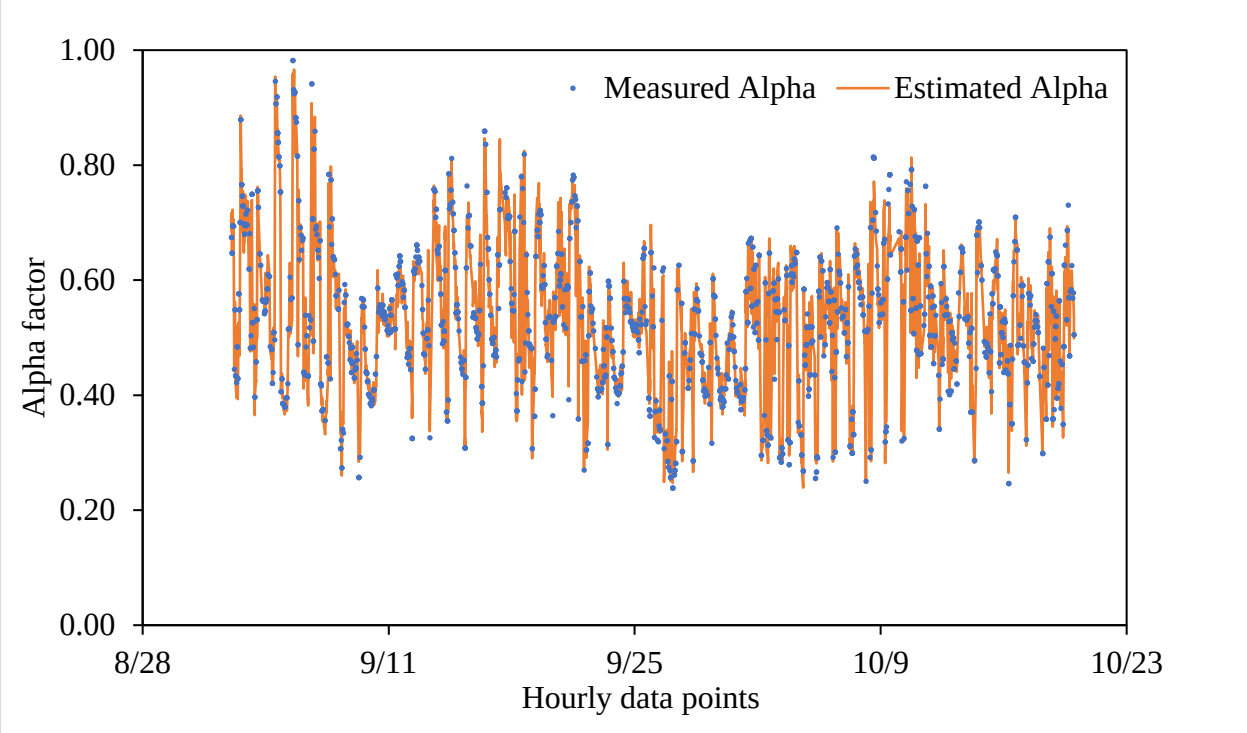
Empirical Relationship	Original Coefficient	Fitted Coefficient	Performance	Source
$\alpha = e^{-k \cdot MLSS}$, MLSS [g/L]	k = 0.0771	k = 5.08E^{-4}	RMSE = 0.1284	(Günder, 2001)

$\alpha = e^{-k \cdot MLSS}$, MLSS [g/L]	k = 0.08788		NSE = -0.0154 d = 0.1350	(Krampe & Krauth, 2003)
$\alpha = \frac{u}{u-v}(e^{-v \cdot MLSS} - e^{-u \cdot MLSS})$ MLSS [g/L]	u = 0.5072 v = 0.1044	u = 0.4551 v = 0.0001	RMSE = 0.2839 NSE = -3.9644 d = 0.3830	(Baquero - Rodríguez et al., 2018)
$\alpha = e^{k \cdot COD - b}$, COD [mg/L]	k = -1.82E^{-3} b = 0.213	k = 3.34E^{-4} b = 0.792	RMSE = 0.1284 NSE = -0.0154 d = 0.1072	(L. Jiang et al., 2017)

(a)


1.00
0.80
0.60
0.40
0.20
0.00
Measured Alpha
Estimated Alpha
8/28
9/11
9/25
10/9
Hourly data points

(b)


1.00
0.80
0.60
0.40
0.20
0.00
Alpha factor
Measured Alpha
Estimated Alpha
8/28
9/11
9/25
10/9
10/23
Hourly data points

(c)

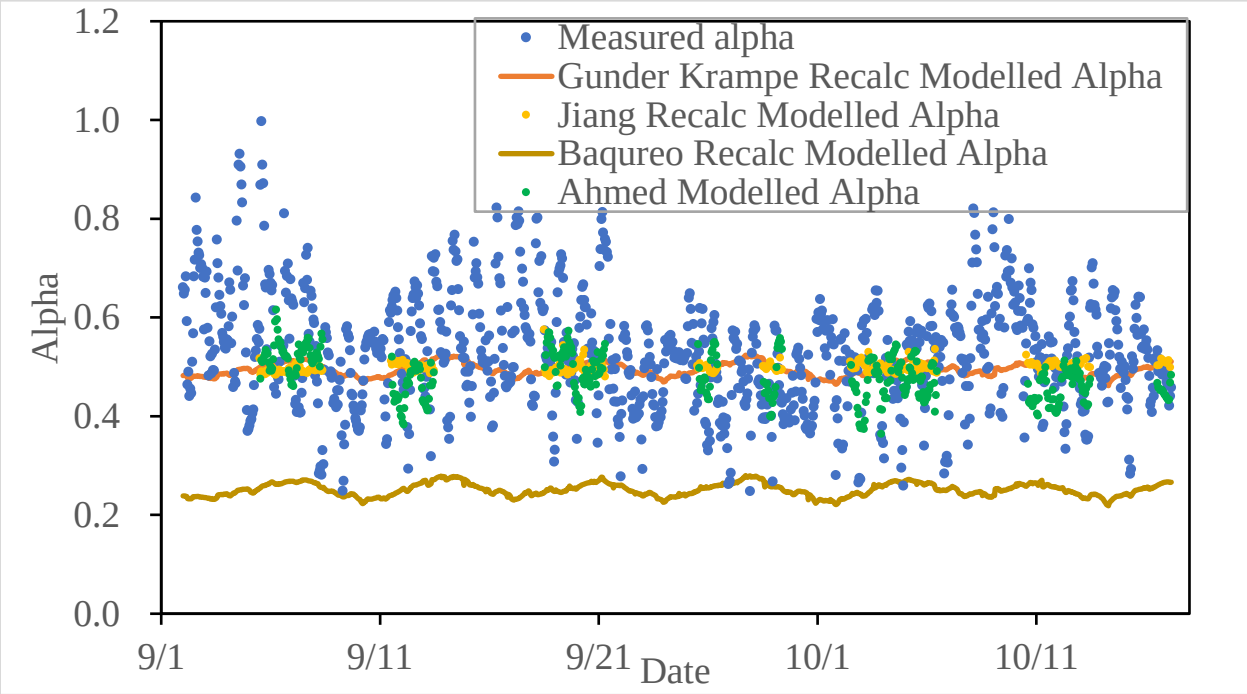


Figure 3-5 Comparison of the modelled and measured alpha-factor (a) direct alpha AutoML model (b) hybrid AutoML – mechanistic alpha model (c) alpha estimated using recalibrated published regression models

3.3.4 Parameter Sensitivity Analysis

In the current study, the AutoML and hybrid AutoML-mechanistic models used several parameters to estimate and capture alpha's dynamic nature successfully. The study also assessed the sensitivity of the models to each input parameter. The current study used the ML method to plot the desired factor (alpha) against all potential parameters and then remove the least important ones through

sensitivity analysis. Several parameters were identified to be correlated with the alpha factor; some of these parameters were never considered in previous studies. A total of ten input parameters were fed into the ML models: HRT, Temperature, primary effluent (PE) total COD, PE sCOD (based on influent fractionation), and aeration tank sCOD (simulated), PE PO_4, PE and aeration tank NH_3, PE TSS and MLSS concentration. Several other parameters, such as DO and airflow, were not used to estimate the alpha factor as those were directly involved in calculating the experimentally determined alpha values. The feature importance rankings for the AutoML and hybrid models built in this paper are summarized in Figures 3.7(a) and (b). For prediction of the alpha factor directly (AutoML model), aeration NH_3 was always the most critical parameter, followed by aeration COD and primary effluent PO_4 (Figures 3.7a). Regarding oxygen fraction prediction (hybrid model), the influence of temperature overweighted all other parameters, but HRT and aeration TSS was still crucial for prediction accuracy (Figures 3.7(b)).

The results showed that parameters measured within the unit (aeration) processes matter significantly. In alpha prediction models (direct AutoML alpha models), the importance of NH_3 dropped when replacing the aeration dataset with primary effluent NH_3 measurements. Aeration sCOD calculated from GPS-X simulation was not as crucial as measured aeration COD. This is observed in both alpha and oxygen fraction models. Assorted studies showed that surfactants significantly impact bubble diameter and reduce Kla (Gillot & Héduit, 2008; Rosso & Stenstrom, 2006a; Sardeing et al., 2006; Stenstrom & Gilbert, 1981). MLSS, TSS, COD and sCOD could therefore be seen as appropriate parameters to characterize the behaviour of oxygen transfer rate in the aeration process. Several other parameters that were not reported to impact oxygen transfer were also correlated with the alpha factor in this study. The concentration of NH_3 in the aeration

tank depends on the nitrification efficiency of the plant; the ammonia concentration may influence the oxygen demand by altering the chemical driving force in the reaction, therefore influencing the alpha factor. HRT has been proved to influence the COD and NH_3 load to the plant, and it could influence the DO needed for the biological reactions and influence the oxygen transfer rate (Sözüdoğru et al., 2020; Dong et al., 2016). The temperature of the wastewater also impacts the oxygen transfer rate. The temperature would influence the rate of the reactions during the aeration process, which changes the DO consumption rate and the dissolution rate of oxygen. In summary, it was found that the data measured from influent flow or primary effluent flow, the data directly measured from the aeration tank, would improve the modelling accuracy. Since most WWTP focus more on the quality of influent and effluent flow, many fail to access the information within the aeration section. Therefore, it is recommended for modern WWTPs to install more sensors to monitor the aeration process to provide a more accurate prediction of the oxygen transfer rate or have a virtual WWTP running in parallel that could provide information on the aeration tank process parameters.

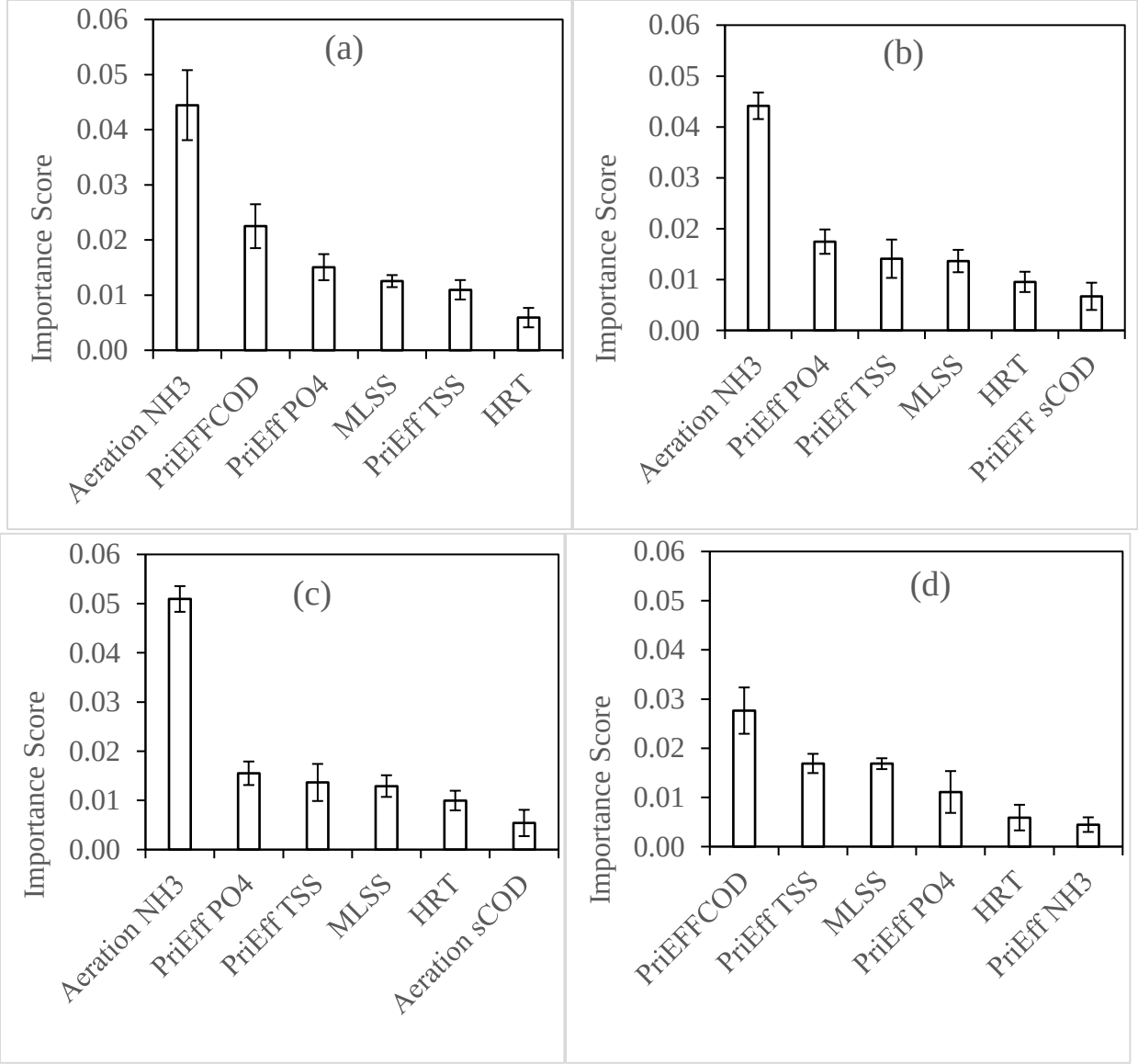


Figure 3-6 Sensitivity rank of input variables estimating alpha using direct alpha AutoML model: (a) inputs=aeration NH_3, PE total COD, TSS and PO_4, MLSS and HRT ; (b) replaced PE tCOD with PE sCOD; (c) replaced PE total COD with aeration sCOD; (d) replaced aeration NH_3 with PE NH_3

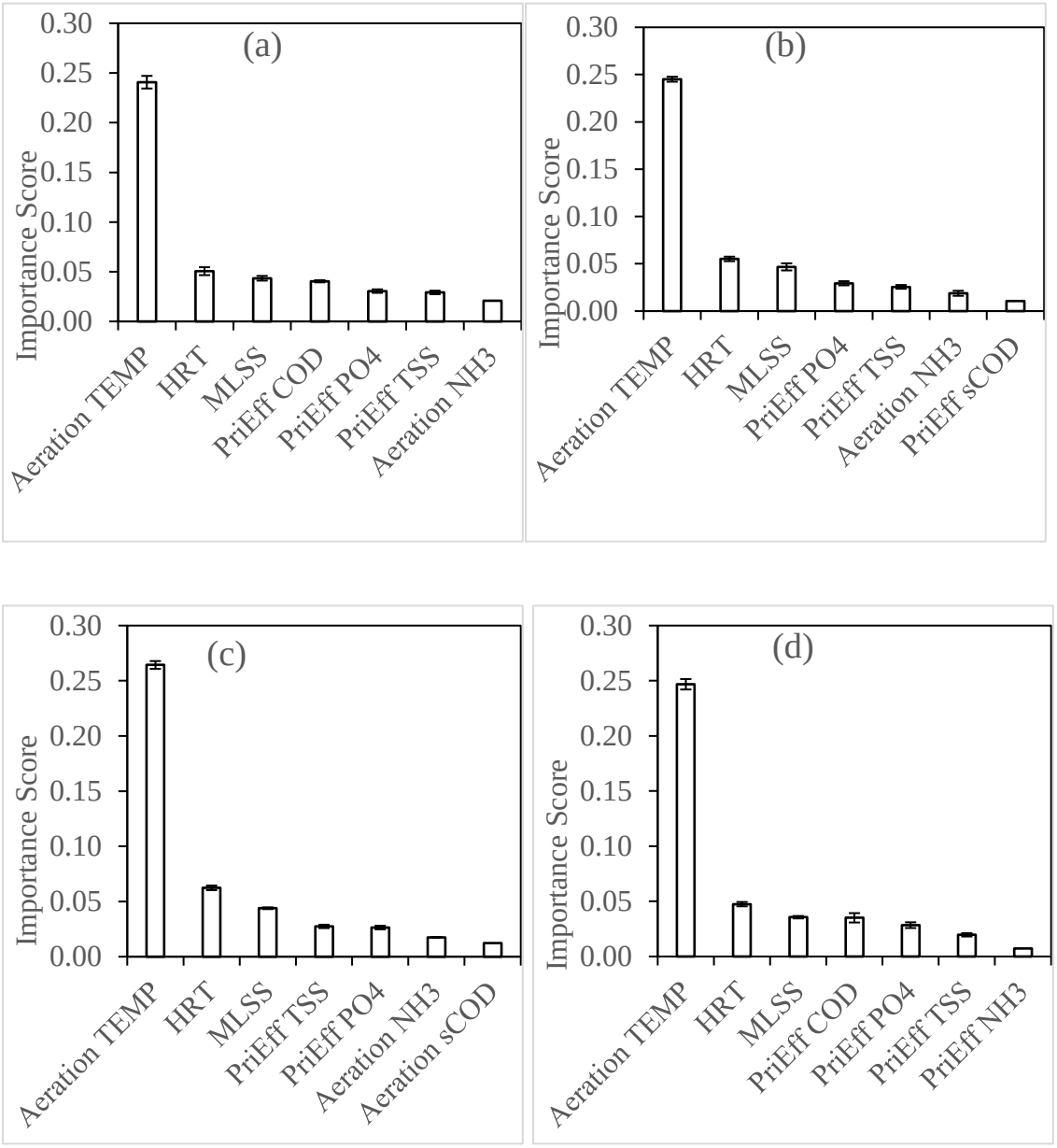


Figure 3-7 Sensitivity rank of input variables estimating alpha using hybrid AutoML-mechanistic alpha model: (a) inputs=aeration TMP and NH_3, PE total COD, TSS and PO_4, MLSS and HRT ; (b) replaced PE tCOD with PE sCOD; (c) replaced PE total COD with aeration sCOD; (d) replaced aeration NH_3 with PE NH_3

3.4 Conclusion

This study developed a new framework for efficient and robust modelling of the dynamic alpha-factor by estimating the off-gas oxygen fraction from activated sludge plants using machine learning. The estimated off-gas fraction was then used to calculate alpha-factor and compared with the alpha estimates directly from wastewater characteristics using published recalibrated regression equations and a machine learning approach. The results showed that: 1) the proposed approach of estimating off-gas fraction showed superior performance compared to directly estimating the alpha; 2) the hybrid-AutoML and AutoMLmodels were reliable to predict trends of the alpha factor in aeration; 3) except for the traditionally acknowledged parameters that were used to predict dynamic alpha change, aeration tank and plant parameters such as temperature, HRT, NH_3 and PO_4 were found crucial to predict dynamic off-gas fraction, and 4) the optimal sampling frequency for model development could be different from the raw sampling frequency of the sensors for the input parameters.

This study showed that AutoML models could be an economical, efficient, and reliable method for alpha modelling to optimize the aeration process. Although AutoML models have shown high efficiency and accuracy in alpha prediction, it has the limitation to produce only black-box models, and therefore there has yet to elucidate the cause-and-effect relationship in the aeration process. Therefore, a future effort is necessary to validate the models in other plants under different process conditions. Furthermore, the underlying mechanisms of the correlation between alpha and the new measurable parameters found in

this study could be explored in further investigations. Parameter analyses based on a larger sample pool of plants could be conducted, which may also allow for the identification of more potential input parameters for alpha modelling with increased accuracy.

Chapter 4

4 Dynamic Airflow and Energy Modelling

4.1 Introduction

In conventional wastewater treatment plants, the secondary treatment is conducted biologically where the microorganisms degrade the organic matter and nutrients in aeration tanks. The aeration tank affords the bioenvironmental conditions for the microorganisms to survive, reproduce and utilize the carbon and ammonia to grow, decreasing the biochemical oxygen demand (BOD) and ammonia level in the wastewater. In this process, ammonia is converted to nitrate during aerobic nitrification, followed by the subsequent anaerobic denitrification, which removes the resulting nitrate and releases the ammonia as the nitrogen gas (Samer, 2005). Dissolved oxygen (DO) concentration in the aeration tanks is of vital importance to the biodegradation process, and therefore greatly influences the overall performance of the wastewater treatment system (Fan et al., 2017). DO dynamics in aeration tanks is generally described with the oxygen mass balance equation:

$$\frac{dC}{dt} = In - Out + Generation - Consumption \ (Equation\ 4.1)$$

As municipal wastewater influent is often poor in oxygen, the input term in this equation can usually be ignored (Holman & Wareham, 2005).. The output term in this equation stands for the quantity of oxygen present in the effluent water of the aeration tank. DO concentration in the effluent flow is considered to be always the same as that in the aeration

tank and therefore cannot be directly altered. The term generation in this equation refers to the dissolved oxygen transferred from the gas phase as supplied by the airflow. The microorganisms consume oxygen in the tank to allow biodegradation to occur. The oxygen consumption rate by the microorganisms is also known as the oxygen uptake rate (OUR), that is, the respiration rate of organisms in the aeration tank. OUR is directly related to the degradation speed of the organic matter in the wastewater. As the organic matter and nutrients being treated in the aeration process mainly consists of the biochemical oxygen demand (BOD) and ammonium nitrogen (NH_4-N), OUR can usually be estimated from the consumption rate of BOD and NH_4-N by microorganism (Garcia-Ochoa et al., 2010).

To minimize energy waste during the aeration process, various research has been done to provide an estimation of the optimum airflow needed to meet the effluent quality standards (Thunberg et al., 2009). The optimum airflow supply is traditionally calculated from the estimated oxygen demand with this equation:

$$G = \frac{OTR}{OTE \cdot CF} \qquad (Equation\ 4.2)$$

Oxygen transfer rate (OTR) is the measurement of the amount of oxygen gas that can be dissolved in water over a given time. It is often used as an indicator for oxygen supply. Oxygen transfer efficiency (OTE) reflects the proportion of oxygen transferred by blowers. This equation also contains a conversion factor (CF) to account for air density, molecular weight, and O2 fraction under standard conditions. These parameters can be calculated

with the following equations, respectively (Hydromantis Environmental Software Solutions Inc., n.d.):

$$OTR = \alpha\theta^{(T-20)}F(k_L a)_{20°C}(\beta C^S - C)\cdot V \qquad (Equation\ 4.3)$$

$$OTE = \frac{Y_R - Y_{OG}}{Y_R} \qquad (Equation\ 4.4)$$

$$CF = \frac{Y_R MW_{O2} P_d}{RT} \ (Equation\ 4.5)$$

Where

G	Airflow rate under field condition	(Dynamic measured or modelled)	[m^3/h]
OTR	Oxygen transfer rate	(Calculated)	[g/h]
OTE	Oxygen transfer efficiency under field condition	(Calculated)	[-]
CF	Conversion factor	(Calculated)	[m^3/g]
α	Alpha factor	(Calculated or modelled)	[-]
θ	Theta factor	(Constant)	[-]
T	Temperature of aeration tank	(Dynamic measured)	[°C]
F	Fouling factor	(Constant)	[-]

$(k_L a)_{20℃}$	Overall oxygen mass transfer coefficient of clean water at standard condition	(Constant)	[h^{-1}]
β	Beta factor	(Constant)	[-]
C^S	Saturation concentration of oxygen in test liquid under equilibrium	(Constant measured)	[mg/L]
C	The concentration of dissolved oxygen in test liquid	(Dynamic measured)	[mg/L]
V	The volume of aeration tank	(Constant)	[m^3]
Y_R	Volumetric fractions of oxygen gas in inlet gas	(Constant)	[-]
Y_{OG}	Volumetric fractions of oxygen gas in outlet gas	(Dynamic measured or modelled)	[-]
MW_{O2}	Molecular weight of oxygen	(Constant)	[g/mol]
P_d	Pressure of post blower gas	(Dynamic measured)	[kPa]
R	Ideal gas constant	(Constant)	[m^3·Pa/K·mol]

As discussed in chapter 3, the overall transfer coefficient (k_La) consists of various factors that are often dynamic under operational conditions:

$$(k_L a)_{ww} = \frac{\rho G}{V} \frac{(Y_R - Y_{OG})}{(C^S - C)} \qquad (Equation\ 4.6)$$

Where

$(K_L a)_{ww}$	Overall oxygen mass transfer coefficient of wastewater	(Calculated)	[h^{-1}]
ρ	Density of oxygen at temperature and pressure at which gas flow is expressed	(Constant)	[g/L]
G	Airflow rate under field condition	(Dynamic measured or modelled)	[m^3/h]
Y_R	Volumetric fractions of oxygen gas in inlet gas	(Constant)	[-]
Y_{OG}	Volumetric fractions of oxygen gas in outlet gas	(Dynamic measured or modelled)	[-]
V	The volume of aeration tank	(Constant)	[m^3]
C^S	Saturation concentration of oxygen in test liquid under equilibrium	(Constant measured)	[mg/L]
C	The concentration of dissolved oxygen in test liquid	(Dynamic measured)	[mg/L]

With most of the parameters mentioned earlier being dynamic and closely related to the real-time condition in the aeration tank, we would naturally find that the relationship between oxygen demand and the airflow supply is not strictly linear. As a result, the required airflow supply is estimated, assuming that all constant factors may lie far from

reality. To account for the discrepancy between airflow estimation and the actual airflow demand, wastewater treatment plants tend to supply excessive airflow to ensure the quality of the effluent flow meets the standards (Steven, 2006). This approach could bring tremendous energy waste to the system.

This chapter aimed to introduce a dynamic method to estimate the real-time airflow supply and energy use during the wastewater aeration process. The approach involved estimating airflow through a hybrid mechanistic and the machine learning method, with the dynamic nature of OTR in mind. The hybrid approach is further compared with another new method to predict the required amount of energy to deliver the estimated airflow supply with machine learning algorithms and mechanistic models.

4.2 Methods

4.2.1 Data Collection

All raw data used in this study was obtained from Adelaide Pollution Control Plant (PCP), located in London, Ontario. A detailed description of the plant can be found in Chapter 3. Adelaide WWTP is equipped with the Supervisory Control and Data Acquisition (SCADA) system, allowing it to present high-quality measurements with a raw resolution of 15 minutes. The field sampling campaign datasets and parameters from online sensors shown in Chapter 3 (Aeration NH_4-N, Temperature, DO, MLSS, Airflow; Post Primary: PO4-P, TSS, COD; and HRT) were used in this study. Further, additional datasets,

including post blower pressure, delivered power, modelled alpha, and off-gas oxygen fraction (Chapter 3 section 3.3), were acquired for dynamic aeration and energy modelling. All raw datasets used in this study were collected during September - October 2018. The raw dataset is summarized in Table 4-1.

4.2.2 Data Preprocessing

Data preprocessing in this study followed the methodology established in Chapter 3. All raw datasets were cleaned with the DBSCAN algorithm to remove outliers and minimize measurement errors' influence on the study results. The resolution of the model input was determined following the same procedure described in Chapter 3. In brief, the raw dataset with its original sampling frequency of 15 minutes was averaged down to new datasets with different time intervals, which were then used to build ML tester models. The time interval that provides the model of the highest performance among the tester models is chosen as the optimum data resolution, which in this case is one hour. The dataset with a one-hour resolution was used as the input dataset for all models built in this study. The induced parameters, including the alpha factor and the hydraulic retention time (HRT), were calculated following the same procedure as in Chapter 3.

Table 4-1 Chapter 4 raw dataset summary

Location	Parameters	Unit	Avg	Stdev	Median	Max	Min
Plant influent	Liquide flow rate	m3/d	28611.56	4954.84	28263.54	40356.42	767.15
Post primary clarifier/Aeration influent	Phosphate	mg P/L	3.50	0.59	3.41	5.96	1.78
	Total suspended solids	mg /L	141.49	30.27	141.69	213.94	62.23
	Chemical oxygen demand	mg/L	259.23	53.46	265	400	126
Post blower	Airflow	m3/d	12579.86	1870.3	13144.56	15716	7633.92
	Pressure	kPa	147.89	1.43	148.23	151.43	142.47
Aeration tank	Mixed liquor suspended solids	mg /L	1380.93	51.95	1381.29	1519.46	1272.26
	Ammonia	mg N/L	3.68	2.46	3.37	11.84	0.55

	Temperature		°C	20.46	0.58	20.54	21.58	19.07
	Dissolved oxygen		mg/L	2.10	0.46	2.24	3.92	0.71
N/A	Alpha factor	Measured	-	0.53	0.11	0.52	0.99	0.25
		Modelled	-	0.53	0.095	0.53	0.90	0.28
	Off-gas oxygen fraction	Measured	%	19.21	0.34	19.27	19.84	17.98
		Modelled	%	19.21	0.32	19.27	19.83	18.01
Other	Delivered power		kW	216.62	31.39	220.37	289.14	126.36

4.2.3 Model Buildup

Two sets of models were built in this study, aiming to estimate aeration airflow and energy consumption, respectively. Each set of models can be classified into mechanistic-based and machine learning models (Figures 4.1). All machine learning models shown in this study were implemented with the Python package AutoGluon (Nick et al., 2020). Details of the machine learning procedure used were similar to the one discussed earlier (Chapter 3 section 2.6).

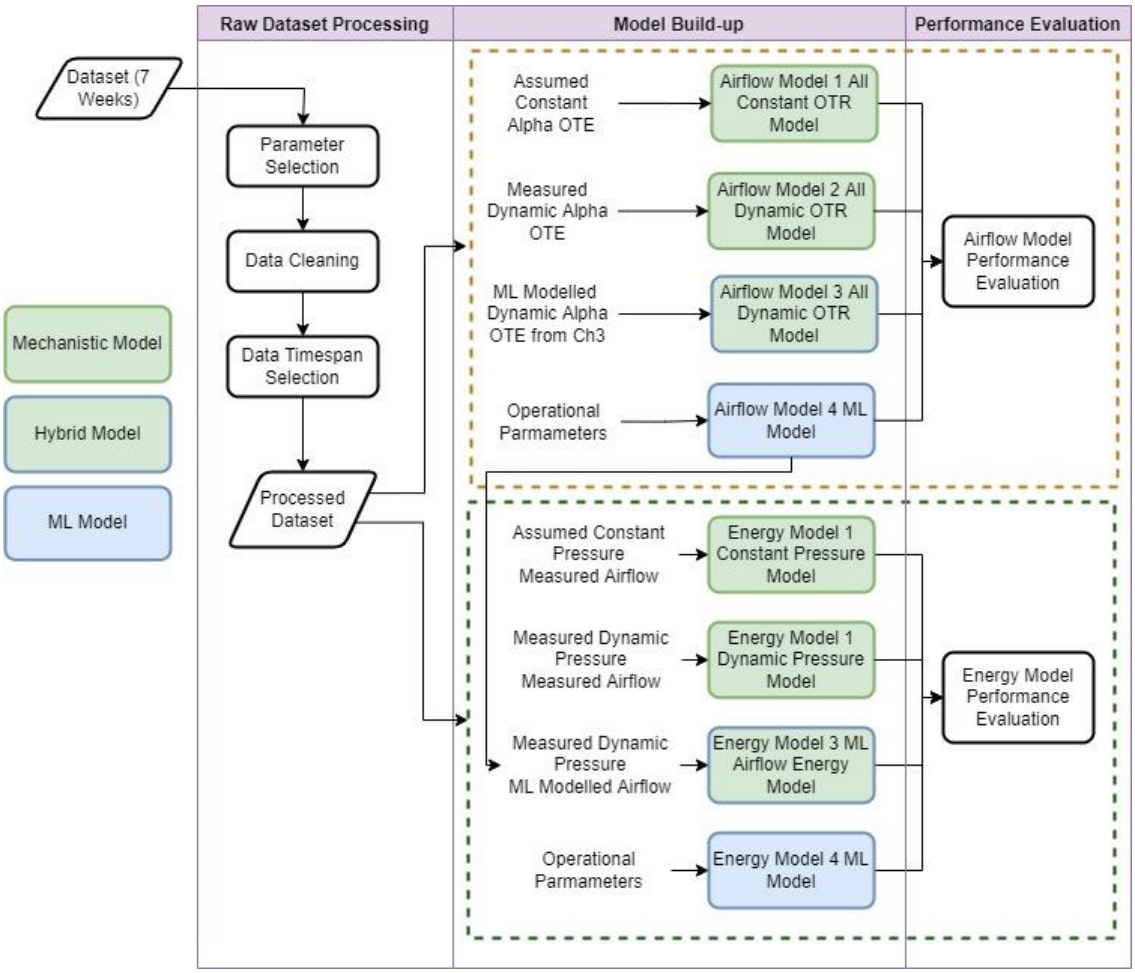


Figure 4-1 Flow chart of chapter 4 airflow and energy modelling

The mechanistic-based airflow models were built upon equations 4.7:

$$G = \frac{FR(k_L a)_{20℃} V}{MW_{O2}} \cdot \frac{\alpha T \theta^{(T-20)} (\beta C^S - C)}{(Y_R - Y_{OG}) P_d} \qquad (Equation\ 4.7)$$

G	Airflow rate under field condition	(Dynamic measured or modelled or calculated)	[m^3/h]
F	Fouling factor	(Constant)	[-]
R	Ideal gas constant	(Constant)	[m^3·Pa/K·mol]
$(k_L a)_{20℃}$	Overall oxygen mass transfer coefficient of clean water at standard condition	(Constant)	[h^{-1}]
V	The volume of aeration tank	(Constant)	[m^3]
MW_{O2}	Molecular weight of oxygen	(Constant)	[g/mol]
α	Alpha factor	(Dynamic calculated or modelled)	[-]
T	Temperature of aeration tank	(Dynamic measured)	[℃]
θ	Theta factor	(Constant)	[-]
β	Beta factor	(Constant)	[-]
C^S	Saturation concentration of oxygen in test liquid under equilibrium	(Constant measured)	[mg/L]

C	The concentration of dissolved oxygen in test liquid	(Dynamic measured)	[mg/L]
Y_R	Volumetric fractions of oxygen gas in inlet gas	(Constant)	[-]
Y_{OG}	Volumetric fractions of oxygen gas in outlet gas	(Dynamic measured or modelled)	[-]
P_d	Pressure of post blower gas	(Dynamic measured)	[kPa]

The independent variables in the mechanism-based models include the alpha factor, temperature, DO, off-gas oxygen fraction and pressure. Table 4-2 summarizes a list of the constants involved in the calculation and their assumed values. One model assumed all parameters but the DO was built as a baseline for model performance comparison (Airflow Model 1). There were two models taking all the independent variables in this study: one with all real-time measurements including alpha (Airflow Model 2), and another with its hybrid alpha model (Airflow Model 3) replaced by the estimations generated by the hybrid ML models presented in Chapter 3 (Section 3.3). The hybrid ML alpha model was developed in this study by estimating off-gas fraction using AutoML models and calculating alpha factor using mechanistic models. Using the modelled alpha in the analysis aimed to test the generalizability of the developed model in plants that lack off-gas measurements. The fourth model ignored using mechanistic airflow models and built complete ML-based airflow models that used airflow data and the same set of operational and process parameters used for the ML models in Chapter 3 (Aeration: NH4-N, Temperature, DO, MLSS; Post Primary: PO4-P, TSS, COD; HRT).

The development of energy models also follows the same strategy (Figure 4.1). With airflow data known, the delivered power (DP) of blowers can be estimated using the mechanistic model as (Hydromantis Environmental Software Solutions Inc., n.d.):

$$DP = \frac{wRT}{K}\left[\left(\frac{P_d}{P_a}\right)^K - 1\right] \quad (Equation\ 4.8)$$

Where

DP	Delivered power of blowers	(Dynamic measured or calculated)	[kW]
w	Molecular mass flow rate of air	(Dynamic measured or calculated or modelled)	[mol/s]
R	Ideal gas constant	(Constant)	[m^3·Pa/K·mol]
T	Temperature of aeration tank	(Dynamic measured)	[°C]
K	K=R/CP where Cp is the heat capacity of air at constant pressure	(Constant)	[-]
P_d	Absolute pressure downstream (outlet) of blower	(Dynamic measured)	[kPa]
P_a	Absolute pressure upstream (inlet) of blower	(Constant)	[kPa]

The mechanism-based energy model (equation 4.8) requires input air pressure, airflow, and temperature. The model performance was assessed by considering several options: Energy Model

1): measured dynamic airflow rates, constant downstream blower pressure and dynamic temperature; Energy Model 2): measured dynamic airflow rates, dynamic blower pressure and dynamic temperature; and Energy Model 3): hybrid model with modelled dynamic airflow (Airflow Model 4), dynamic pressure and dynamic temperature.

Further, a new ML energy model (Energy Model 4) was developed using the same set of wastewater process and operational parameters by using the same input parameters as Airflow Model 4.

Table 4-2 Constant values in chapter 4

Parameter	Description or Equation	Value	Source	Units	Reference
F	Fouling factor	1	GPS-X	-	(Hydromantis Environmental Software Solutions Inc., n.d.)
R	Ideal gas constant	8.314	Online Database	m^3·Pa/K·mol	(Engineering ToolBox, 2004)
β	Beta factor: salinity-surface tension	0.99	Literature	-	(Tchobanoglous et al., 2003)
$k_L a_{20℃}$	Overall oxygen mass transfer coefficient of clean water	12.53	Literature	h^{-1}	(Lee, 2017)
V	Volume of aeration tank	9484	Plant data	m^3	-
MW_{O2}	Molecular weight of oxygen	32	Literature	g/mol	(National Center for Biotechnology Information, n.d.)

Y_R	Volumetric fraction of oxygen in the air	20.946	Online Database	%	(Engineering ToolBox, 2003)
θ	Theta factor	1.024	Literature	-	(Iranpour et al., 2000)
C^S	Saturation concentration of oxygen in test liquid in equilibrium with exit gas	$C^S = 14.189 \cdot e^{-0.022T}$	Literature	g/m^3	(Benson & Krause, 1980)
K	For U.S. standard air	0.283	GPS-X	-	(Hydromantis Environmental Software Solutions Inc., n.d.)
P_a	Absolute pressure upstream of the blower	101.325	Online Database	kPa	(Engineering ToolBox, 2004)

4.2.4 Model Performance Evaluation

Similar to Chapter 3, the performance of each model followed the goodness-of-fit procedures, including three criteria: Root Mean Square Error (RMSE; Chapter 3 equation 3.6), Nash-Sutcliffe efficiency (NSE; Chapter 3 equation 3.7), and Index of Agreement (d; Chapter 3 equation 3.8).

4.3 Results and Discussion

4.3.1 Airflow and blower data characteristics

The daily trend and the dynamic fluctuation of the airflow, blower pressure and blower power during the sampling period are plotted in Figure SD2 and SD3 in the supporting document. A summary of their data characteristics is listed in Table 4-1. The aeration airflow came in a wide range, increasing from around 8,000 m^3/h to over 16,000 m^3/h during the sampling period. However, the airflow did not show a clear daily changing trend aside from its rapid increase. On the other hand, the airflow pressure is stable at around 147.89 kPa. Throughout the days, the pressure decreases in the morning, reaching its valley point of about 146 kPa at 8:00, then gradually increasing until its stable point at 15:00. The post blower power usage demonstrated a similar trend as the airflow, which increases in the long-term but does not change significantly throughout the days.

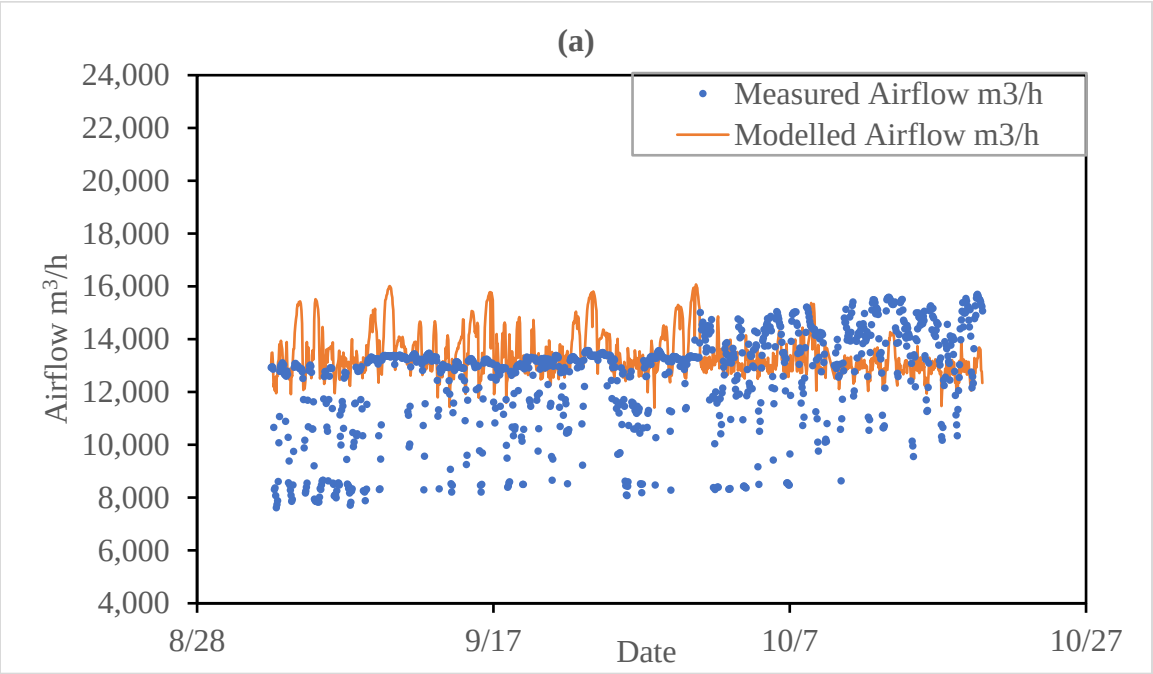
4.3.2 Airflow Models

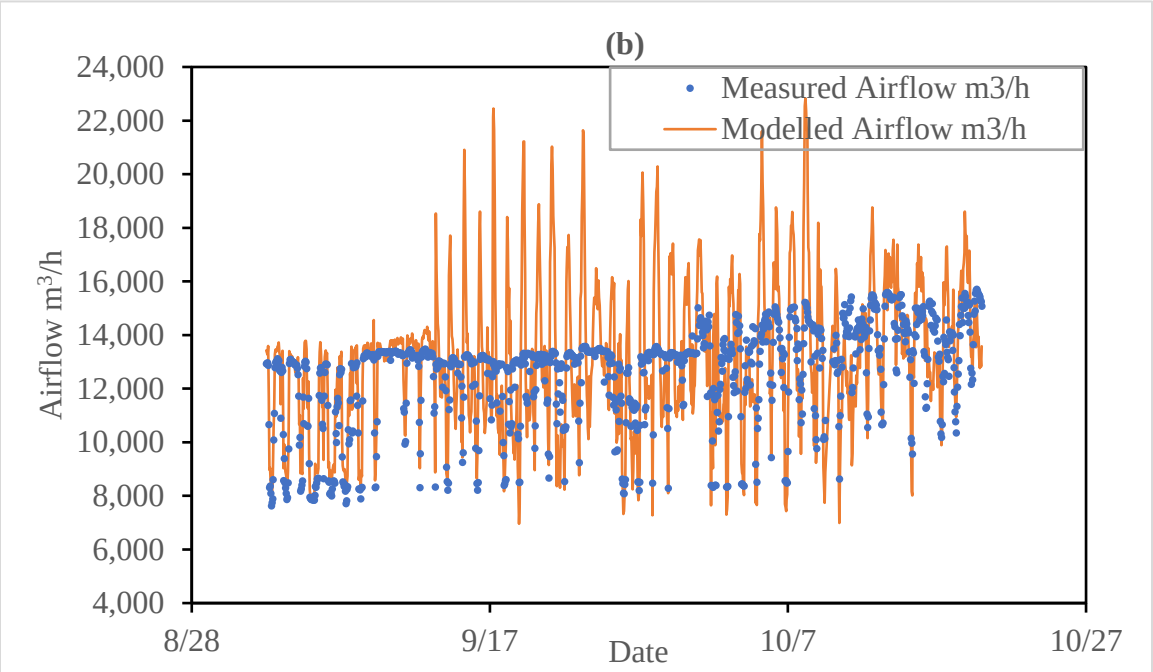
Compared with the traditional mechanism-based model that assumes both the alpha factor and the OTE to be constant (RMSE=2110.42, NSE=-0.26, d=0.38), the model built with dynamic measurement data (OTE and alpha factor) is substantially more accurate in predicting airflow (RMSE=2046.35, NSE=-0.19, d=0.79) (Figures 4.2 a and b). Not only did it describe the general rising trend of airflow over the months, but it also captured the daily fluctuations in the data (Figure 4.2 b). On the contrary, the airflow estimation made by the mechanism-based model with all constant inputs lie almost strictly around 13,400 m^3/h (11,410.2-16,076.4 m^3/h) over the study period and failed to predict the increasing trend in October (last month of the field campaign) (Figure 4.2 a). Although offering some insights on the daily change of the airflow (6,959.5-22,822.6 m^3/h), its predictions are far from the actual valley values measured (7,633.9-15,716.0 m3/h).

When real-time measurements of the alpha-factor and the off-gas oxygen fraction are not available in WWTP, airflow estimations made upon modelled parameters might be of interest. Our results showed that the mechanism-based airflow model based on hybrid ML predictions of the alpha-factor and OTE (generated with the AutoML model described in Chapter 3) demonstrated similar predictive power to the one built upon on-site measurements (RMSE=1974.16, NSE=-0.11, d=0.78) (Figure 4.3 c). While its estimations are generally more dispersive (e.g. Figure 4.2 c during Sep 8 – Sep 10), it still offered decent predictions and captured most of the valley measurements.

The overshooting problem was one major drawback of both mechanism-based models with dynamic alpha and OTE inputs. This problem arose on September 13 and persisted until October

14 (Figure 4.2 b). The vast majority of such events took place from the midnights, starting at 0:00 am, to the mornings around 10:00 am. During this period, the mechanism-based models showed many prediction points, mostly with values above 16,000 m^3/h, exceeding their targets. It is suspected that this was caused by the assumption of the constant K_La_{20} value involved in the calculation of the OTR. As reported by Özbek and Gayik (2001), K_La_{20} (clean water) can take a wide range of values from 0 to 360 h-1, depending on many parameters, including the mixing speed, airflow rate, and water viscosity. However, due to the lack of a well-tested universal equation describing the change of K_La_{20} in clean water, we treated the K_La_{20} as a constant in all mechanism-based models in this study. The influent flow for Adelaide WWTP usually carries a significantly smaller amount of organic matter during the nights, which can be reflected in the dynamic change of the ammonia level. A lower loading would result in a smaller OTE level, as the lower oxygen demand pushes the oxygen solubility balance away from dissolution. The constant parameters we assumed in the calculation of the airflow would not have the ability to reflect such dramatic changes, which could therefore cause the deviation. Future research on such overshooting phenomenon might extend our explanation.


(a)
Measured Airflow m3/h
Modelled Airflow m3/h
Airflow m3/h
24,000
22,000
20,000
18,000
16,000
14,000
12,000
10,000
8,000
6,000
4,000
8/28
9/17
10/7
10/27
Date


(b)
Measured Airflow m3/h
Modelled Airflow m3/h
Airflow m3/h
24,000
22,000
20,000
18,000
16,000
14,000
12,000
10,000
8,000
6,000
4,000
8/28
9/17
10/7
10/27
Date

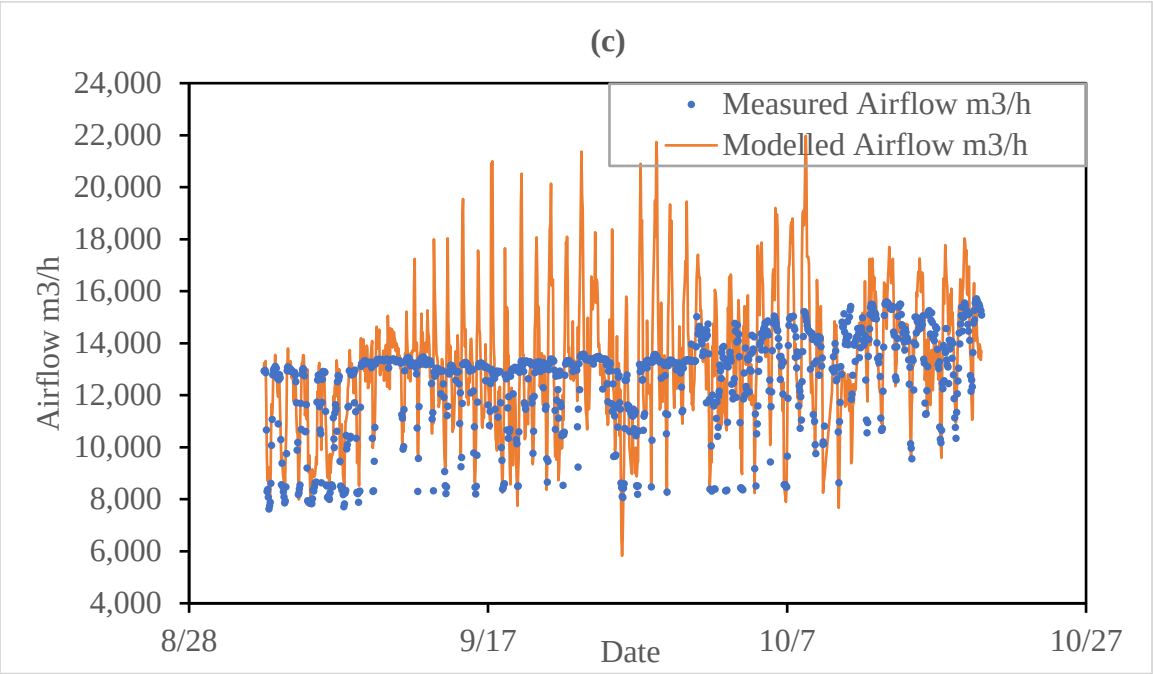


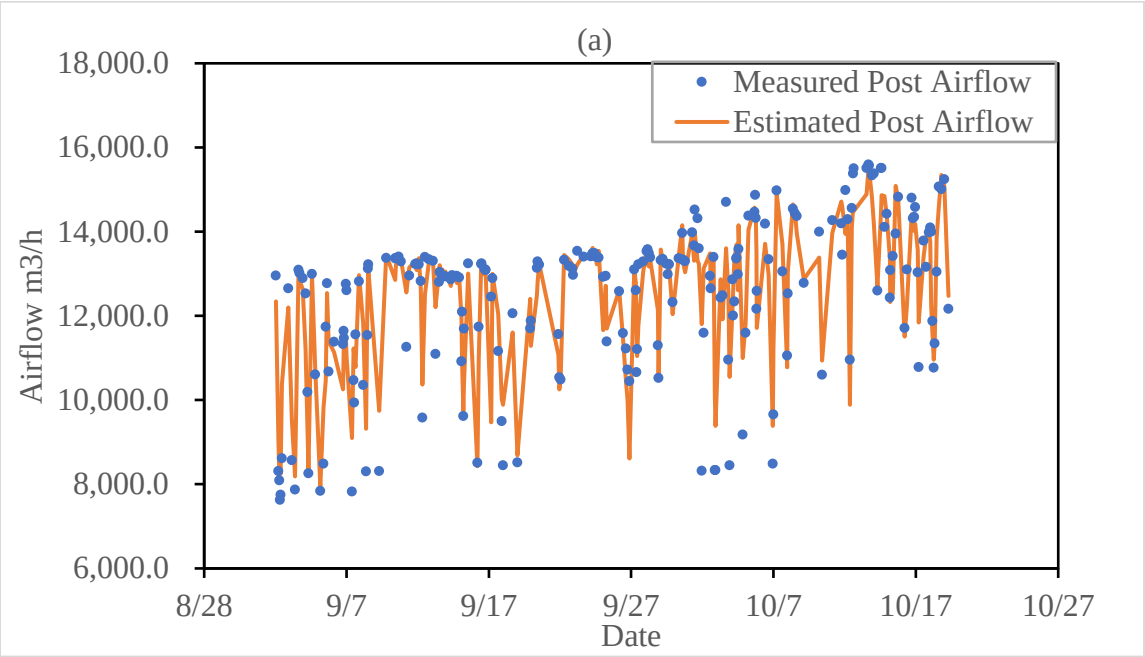
Figure 4-2 Dynamic airflow simulation using mechanistic airflow equation: (a) Constant (average) alpha and OTE (RMSE=2110.42, NSE=-0.26, d=0.38); (b) Measured dynamic alpha and OTE (RMSE=2046.35, NSE=-0.19, d=0.79); and (c) ML Modelled dynamic alpha and OTE inputs (RMSE=1974.16, NSE=-0.11, d=0.78)

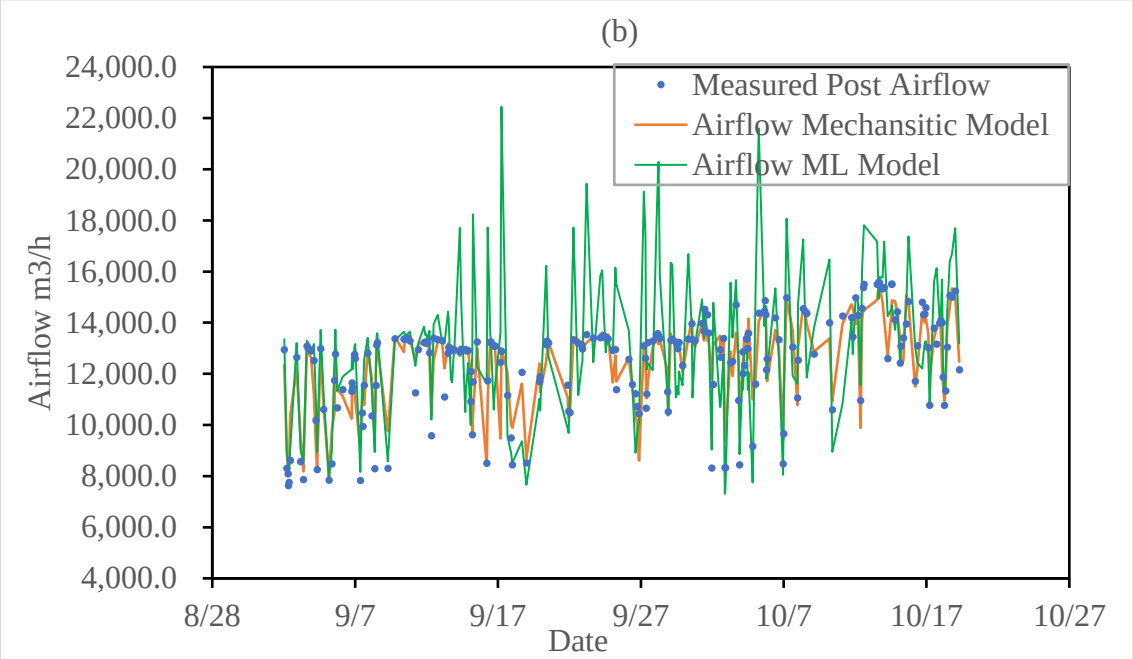
The machine learning airflow model, on the other hand, demonstrated the highest predictability among all airflow models built in this study (Figure 4.3 a; RMSE=744.71, NSE=0.84, d=0.95). The ML airflow model successfully captured both the long-term rising trend and the short-term daily fluctuations in airflow. It also showed high robustness throughout the study period.

The NH_4-N recording in the aeration tank provided the most predictive power among all input parameters, followed by temperature and the TSS measurement from the influent flow of the aeration tank (Figure 4.3 c). As one of the primary purposes of the aeration process, ammonia-nitrogen removal efficiency is directly correlated with the oxygen demand of the microorganisms

present in the aeration tank. TSS and COD played similar roles showing the level of loading in the wastewater. This observation demonstrated a close relationship between the airflow and the nitrogen level present in the aeration tank, which confirmed the importance of the demand-related operational parameters in airflow prediction models. Our findings on the correlation between temperature and airflow agree with previous research on the factors affecting OTE (Chapter 3 section 3.4). The results confirm that the inclusion of the simple temperature measurement can dramatically improve the overall quality of airflow predictions for wastewater practitioners.

It is worth discussing these interesting facts revealed by the low importance of DO in airflow predictions. Although directly involved in the calculation of the mechanism-based models, the contributions DO make to the overall predictive power of the airflow ML model are very limited. Such observation could potentially relate to the fact that the blowers in Adelaide WWTP are controlled in response to the DO level detected. Due to this characteristic of the plant, the DO level in Adelaide WWTP is relatively more stable in comparison to other operational parameters, causing it to supply less information for the prediction. This contrary might be of interest in future studies, where this modelling framework can be applied to other plants with more significant variability in DO.


(a)
18,000.0
16,000.0
14,000.0
12,000.0
10,000.0
8,000.0
6,000.0
Airflow m3/h
8/28
9/7
9/17
9/27
10/7
10/17
10/27
Date
Measured Post Airflow
Estimated Post Airflow


(b)
24,000.0
22,000.0
20,000.0
18,000.0
16,000.0
14,000.0
12,000.0
10,000.0
8,000.0
6,000.0
4,000.0
Airflow m3/h
8/28
9/7
9/17
9/27
10/7
10/17
10/27
Date
Measured Post Airflow
Airflow Mechansitic Model
Airflow ML Model

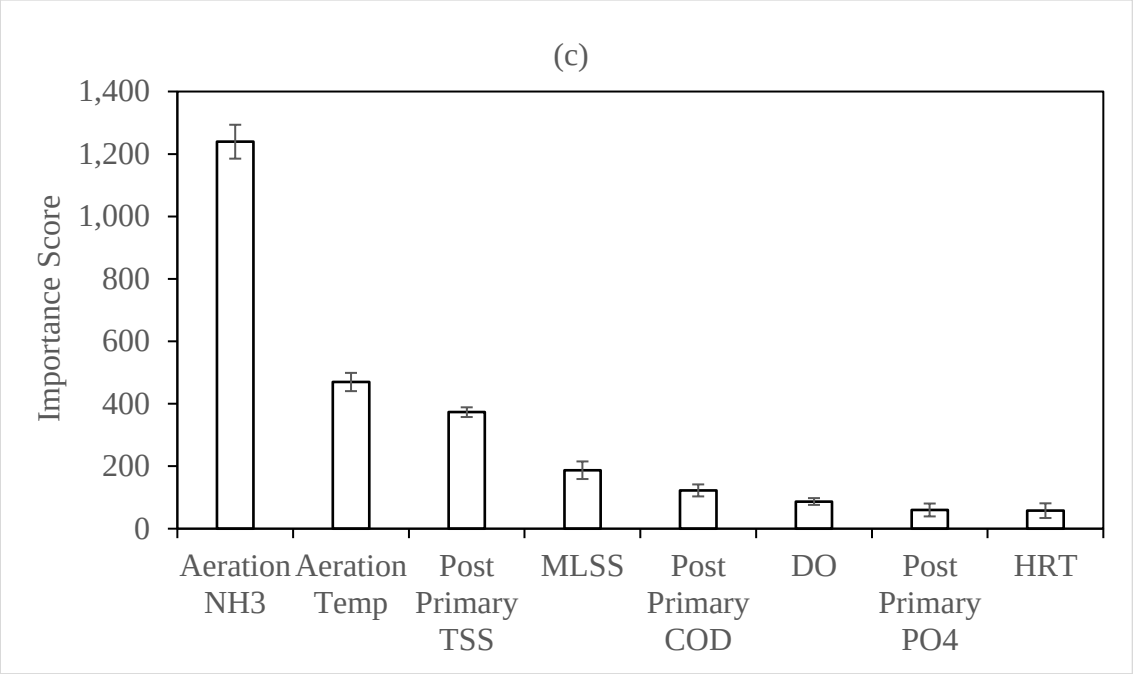
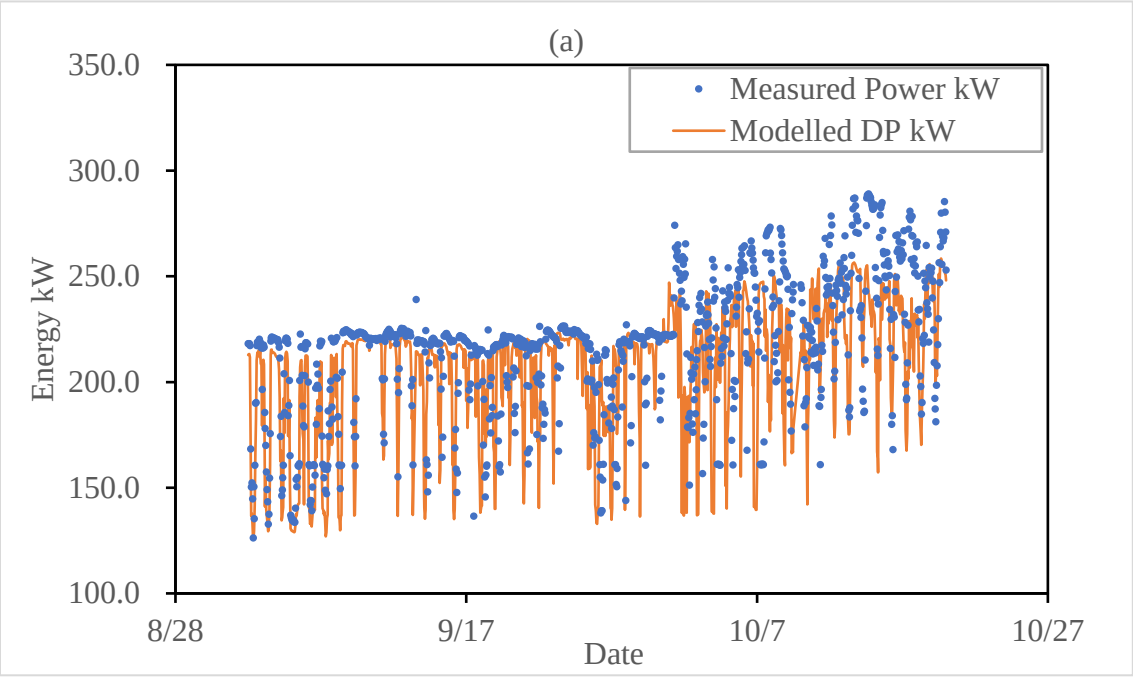


Figure 4-3 Dynamic wastewater aeration airflow consumption simulation using: (a) Using ML (RMSE=744.71, NSE=0.84, d=0.95); (b)Airflow Mechansitic Model (RMSE=2046.35, NSE=-0.19, d=0.79) and ML Model compare with measured airflow; and (c) Input parameter importance score

4.3.3 Energy Models

The mechanism-based energy model provided decent predictability even with the assumed constant post-blower pressure input (Figure 4.4a; RMSE=31.88, NSE=-0.03, d=0.70). The mechanism-based energy model built with dynamic post-blower airflow pressure, in fact, showed slightly lower performance when the turbo energy was relatively stable before October (Figure 4.4 b; RMSE=29.62, NSE=0.11, d=0.77). However, the dynamic model appears to be more accurate

as energy usage rises. Based on the goodness-of-fit measurements, we can tell that the adaptation of dynamic parameters does lead to good results for the energy model as well, even if the improvement is negligible.



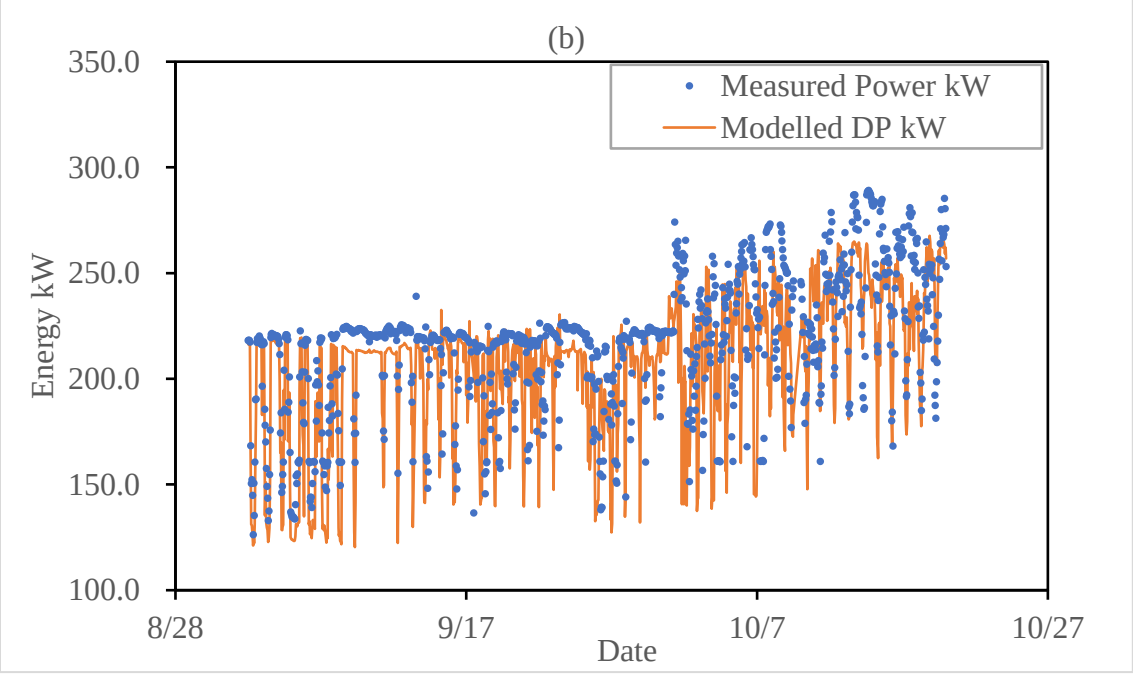
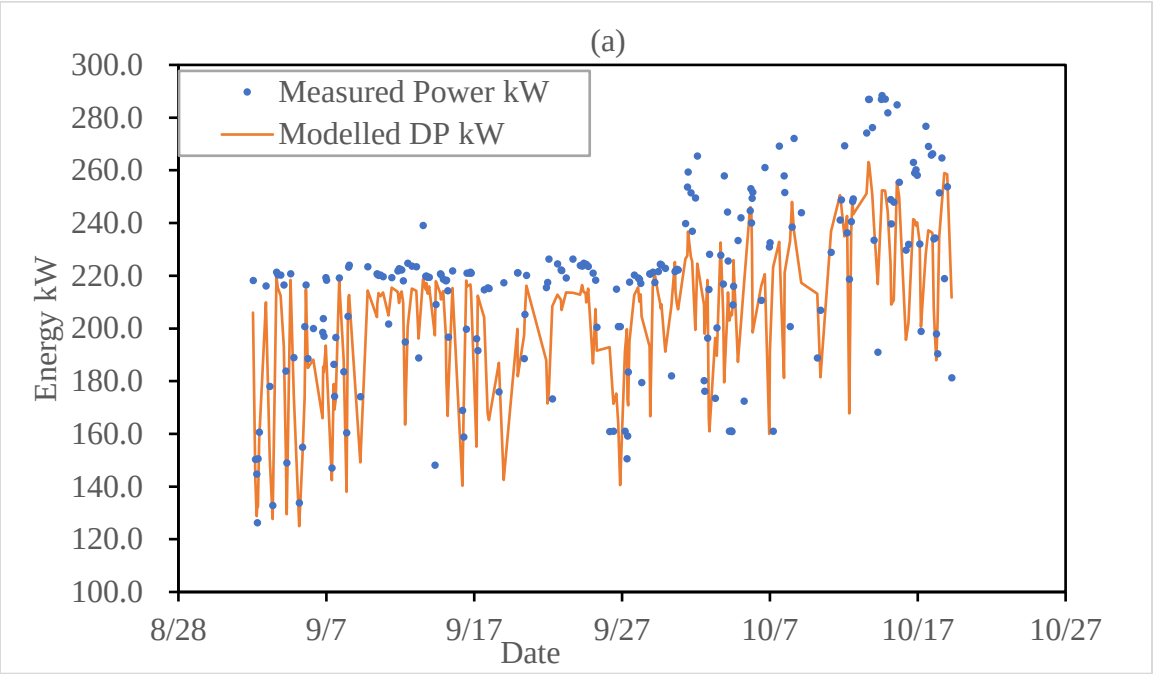
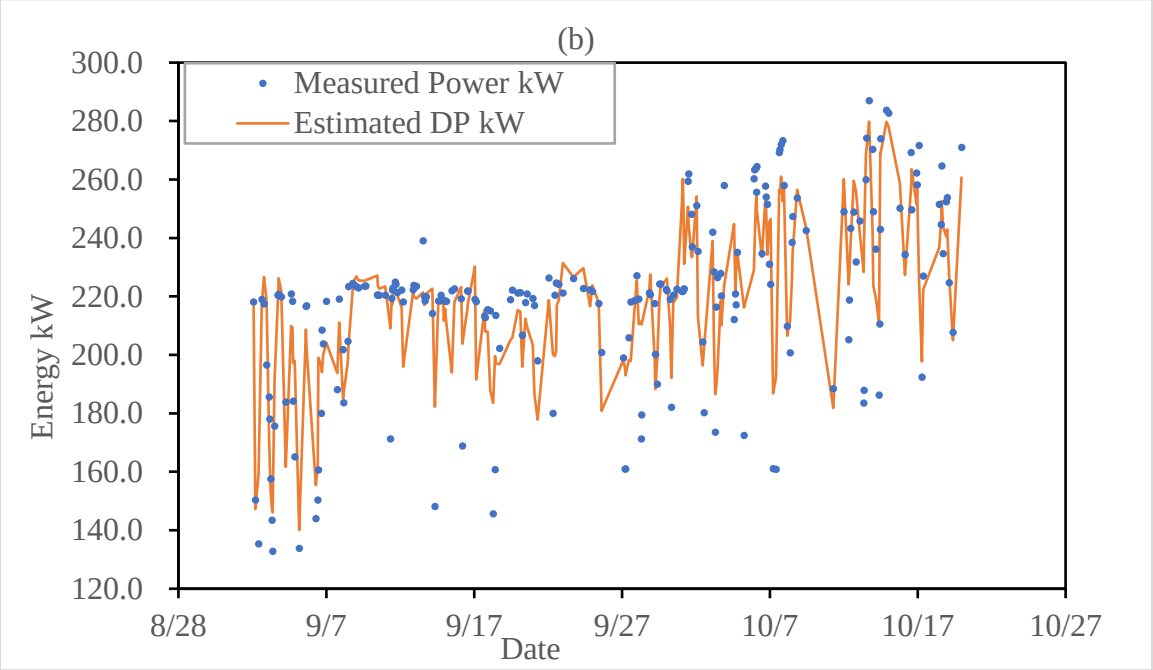


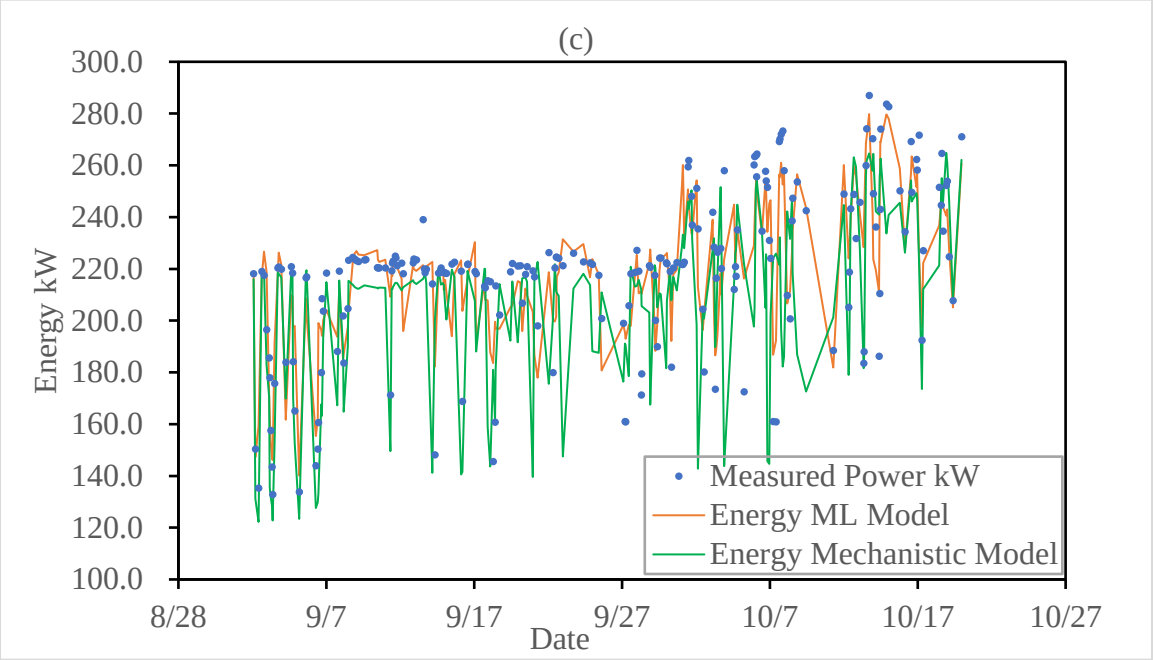
Figure 4-4 Dynamic wastewater aeration energy consumption simulation using: (a) mechanistic model with constant pressure and measured dynamic airflow (RMSE=31.88, NSE=-0.03, d=0.70) (b) mechanistic model with measured dynamic pressure and dynamic airflow (RMSE=29.62, NSE=0.11, d=0.77)

The two ML-based energy models showed similar predictive power (Figures 4.5 a and b). The model derived from the airflow ML estimations devotedly followed the overall trend of the dynamic change in energy; however, there seems to exist a small gap between the measured peak values and the predictions (RMSE=26.35, NSE=0.34, d=0.81). The ML model directly built upon the operational parameters showed better performance. Although losing points below 180 kW, this model accurately predicted the overall energy change during the study period, mainly featuring the rising period starting from October, where all other models fail to capture most of the peaks (RMSE=16.56, NSE=0.70, d=0.91). Comparisons revealed that the ML models directly built upon operational parameters have the highest prediction accuracy and can provide reliable estimations even in the face of substantial changes in energy.

The importance scores of the input parameters revealed that the temperature, HRT and aeration NH_3-N are the top three most important factors to predict energy. The importance of the temperature poses no surprise as it is also an important factor involved in the mechanism equations of blower energy usage. The importance ranks of NH_3-N, TSS and MLSS agree with the ML airflow model. However, we do not yet know the theoretical reason behind the high correlation between HRT and blower energy. We suspect that it is related to the extra energy expenditure used to stir up the wastewater at a higher flow rate, but future investigations are necessary for this issue.


(a)
Measured Power kW
Modelled DP kW
Energy kW
300.0
280.0
260.0
240.0
220.0
200.0
180.0
160.0
140.0
120.0
100.0
8/28
9/7
9/17
9/27
10/7
10/17
10/27
Date


(b)
Measured Power kW
Estimated DP kW
Energy kW
300.0
280.0
260.0
240.0
220.0
200.0
180.0
160.0
140.0
120.0
8/28
9/7
9/17
9/27
10/7
10/17
10/27
Date


(c)
300.0
280.0
260.0
240.0
220.0
200.0
180.0
160.0
140.0
120.0
100.0
Energy kW
8/28
9/7
9/17
9/27
10/7
10/17
10/27
Date
Measured Power kW
Energy ML Model
Energy Mechanistic Model

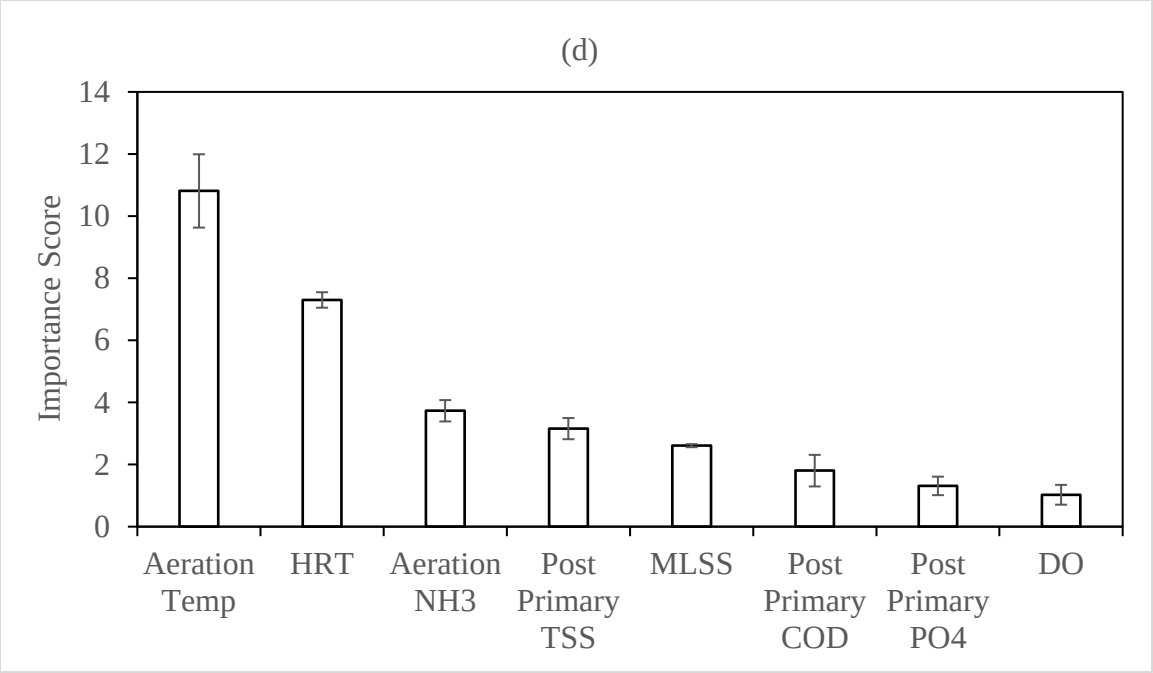

(d)
14
12
10
8
6
4
2
0
Importance Score
Aeration Temp
HRT
Aeration NH3
Post Primary TSS
MLSS
Post Primary COD
Post Primary PO4
DO

Figure 4-5 Dynamic wastewater aeration energy consumption simulation using: (a) mechanistic model with measured pressure and ML modelled dynamic airflow (Airflow Model 4) (RMSE=26.35, NSE=0.34, d=0.81); (b) ML aeration energy model using process and operating paramters (RMSE=16.56, NSE=0.70, d=0.91); (c) Energy mechanistic model (RMSE=29.62, NSE=0.11, d=0.77) and energy ML model compare with measured energy; and (d) Dynamic aeration energy consumption ML input parameter importance score

4.4 Conclusion

To optimize oxygen supply and minimize energy waste in the aeration process, it is necessary to have an accurate real-time estimation of the airflow supply requirement for successful control strategies. While most traditional airflow models tend to estimate the airflow supply based on the oxygen demands of the microorganisms as a correlation to the level of nutrients present, this method fails to take the dynamic nature of the oxygen transfer efficiency into consideration. As a result, the traditional airflow models often lack accuracy and precision.

This work proposed a new modelling approach utilizing both the conventional mechanism-based equations and the state-of-the-art machine learning algorithms to predict the dynamic airflow supply and the corresponding blower energy consumption. The developed model emphasizes the importance of dynamic oxygen transfer rate and the operational parameters in calculation. With the support of the machine learning method, it can be adapted in plants without long-term off-gas measurements. The promising results confirm that the machine learning method is significantly more accurate than pure mechanism-based estimations. It is able to estimate long-term changes as

well as daily fluctuation cycles. Besides, the models developed also cast a new light on the correlation between airflow and some new parameters that have not been previously acknowledged to show effects on the biodegradation process. Regardless, the presented modelling approach provides an accurate alternative to monitor real-time airflow demand and blower energy consumption, which improves aeration operation control strategies.

Future investigations are necessary to validate the new relationships discovered in this study. Furthermore, we believe that more work could be established in the field of wastewater modelling with the aid of machine learning methods.

Chapter 5

5 Conclusion and Recommendation

To minimize energy consumption while meeting the stringent requirements for effluent quality, modern WWTPs must optimize the control strategies for the aeration process with advanced modelling approaches. Although widely studied, traditional mechanistic-based models often fail to cover all the complicated interactions among the various parameters involved in the dynamic process conditions. Therefore, hybrid machine learning methods have been used to model aeration processes to avoid such shortcomings.

In this study, a modelling framework that combines mechanism-based mathematical models and ML algorithms has been developed to estimate the dynamic changes of various parameters involved in the aeration process. The proposed models demonstrated significantly increased accuracy over traditional models in capturing the daily fluctuations, long-term changes and peak values, as demonstrated by a case study on the Adelaide WWTP. The generalizability of the models was tested on parameters including the alpha factor in oxygen transfer rate calculation, the airflow demand of the aeration tank, and the power consumption of the aeration blowers. The significant findings were as follows:

- The optimum sampling frequency for maximum model performance is not necessarily the highest available. Excess data resolution may cause the model to suffer from overfitting and reduce its accuracy.
- The alpha factor is dynamic and can be best estimated with the hybrid ML model using a novel concept of predicting off-gas fractions used as an input for

mechanistic alpha models. ML alpha-factor models can predict the real-time fluctuations of the alpha factor based on operational measurements from full-scale WWTP.

- The dynamic airflow demand in aeration can be estimated based on the oxygen transfer rate and other operational parameters. While the ML model demonstrated the best performance among all the modelling methods tested, the combination of ML and mechanism-based models also offered decent predictions.
- The ML modelling framework can also be applied to estimate the energy consumption of the airflow blowers in the aeration process. Furthermore, our results suggest that the parameters that provided the most predictability for the energy and the airflow models could be different.

The resulting model of our work could be implemented in the control loops of the daily workflow for the WWTPs to test its ability on energy saving in a real-life scenario. Further research on the use of machine learning modelling methods might expand the research field to other parameters in the wastewater treatment processes. The methodology developed in this study can also be applied with input datasets from different plants, sites, or periods to further validate the models' robustness. Furthermore, the study suggests that the combination of machine learning methods and mechanistic equations may produce higher predictability and increased interpretability. We encourage future researchers and wastewater treatment practitioners to validate further the combined modelling approach in different areas of wastewater treatment modelling. Our findings on the new operational factors that influence aeration can also be interesting. Research should further develop and confirm these initial findings by analyzing the theory behind them.

Potential limitations of this study mainly rooted in the data collection process. The estimations of the model are made based on the operational data collected only in the Adelaide WWTP. The model and the associated modelling method may therefore lack generalizability for other plants. A model developed through a similar process to ours for another WWTP may show limited accuracy because of but not limited to the following conditions: (1) when manual control is extensively used; (2) when the sensors installed fail to capture the parameters that are more closely related to the outcome; and (3) when the randomness in the recorded data outweighs the trend. Similarly, the temporal limitation on the input dataset for the model also limited the generalizability of the model to other times of the year. The input dataset used in this study covers only the period of Sep-02 to Oct-17, i.e., less than two months in Autumn. Many parameters, e.g., temperature, may be heavily correlated with the sampling time of the year. Thus, data collected in such a short time frame may not accurately reflect all operational conditions throughout the year. The model induced from such data therefore may be prone to bias and lose accuracy when predicting a different season. Moreover, the types of input data are limited by the types of parameters available in Adelaide WWTP. We might be unaware of some other measurable parameters that more accurately reflect the dynamic changes in our desired outcome parameter. Finally, the lack of previous research studies on the topic may hinder the credibility and scope of this project. Prior research studies that focused on aeration modelling with the machine learning approach are limited. We therefore suggest further research to continue exploring this field and to assess the performance of the ML model with an enlarged dataset.

6 Bibliography

Ahmed, A. S., Rosso, D., Santoro, D., & Nakhla, G. (2021). Influence of substrates concentrations on the dynamics of oxygen demand and aeration performance in ideal bioreactors. *Process Safety and Environmental Protection*, *153*, 339-353.

Ahnert, M., Blumensaat, F., Langergraber, G., Alex, J., & Woerner, D. (2010). Goodness-of-fit measures for numerical modelling in urban water management – a summary to support practical applications. *10th IWA Specialised Conference on Design, Operation and Economics of Large Wastewater Treatment Plants*.

Åmand, L., Olsson, G., & Carlsson, B. (2013). Aeration control - A review. In *Water Science and Technology* (Vol. 67, Issue 11). https://doi.org/10.2166/wst.2013.139

Amerlinck, Y., Bellandi, G., Amaral, A., Weijers, S., & Nopens, I. (2016). Detailed off-gas measurements for improved modelling of the aeration performance at the WWTP of Eindhoven. *Water Science and Technology*, *74*(1). https://doi.org/10.2166/wst.2016.200

Atanasova, N., & Kompare, B. (2002). Modelling of wastewater treatment plant with decision and regression trees. *3 Rd Workshop on Binding Environmental Sciences and Artificial Intelligence (BESAI'2002)*, *23*(9), 3.

Baquero-Rodríguez, G. A., Lara-Borrero, J. A., Nolasco, D., & Rosso, D. (2018). A Critical Review of the Factors Affecting Modeling Oxygen Transfer by Fine-Pore Diffusers in Activated Sludge. *Water Environment Research*, *90*(5). https://doi.org/10.2175/106143017x15131012152988

Barker, P. S., & Dold, P. L. (1997). General model for biological nutrient removal activated-sludge systems: model presentation. *Water Environment Research*, *69*(5). https://doi.org/10.2175/106143097x125669

Benson, B. B., & Krause, D. (1980). The concentration and isotopic fractionation of gases dissolved in freshwater in equilibrium with the atmosphere. 1. Oxygen. *Limnology and Oceanography*, *25*(4). https://doi.org/10.4319/lo.1980.25.4.0662

Borzooei, S., Miranda, G. H. B., Abolfathi, S., Scibilia, G., Meucci, L., & Zanetti, M. C. (2020). Application of unsupervised learning and process simulation for energy optimization of a WWTP under various weather conditions. *Water Science and Technology*, *81*(8). https://doi.org/10.2166/wst.2020.220

Boyle, W. C., Hellstrom, B. G., & Ewing, L. (1989). Oxygen transfer efficiency measurements using off-gas techniques. *Water Science and Technology*, *21*(10-11–11 pt 4). https://doi.org/10.2166/wst.1989.0327

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1). https://doi.org/10.1023/A:1010933404324

Chern, J. M., Chou, S. R., & Shang, C. S. (2001). Effects of impurities on oxygen transfer rates in diffused aeration systems. *Water Research*, *35*(13). https://doi.org/10.1016/S0043-1354(01)00031-8

City of London. (2018). *Adelaide wastewater treatment plant annual report*.

de Clercq, D., Wen, Z., & Fei, F. (2019). Determinants of efficiency in anaerobic bio-waste co-digestion facilities: A data envelopment analysis and gradient boosting approach. *Applied Energy*, *253*. https://doi.org/10.1016/j.apenergy.2019.113570

Deepnarain, N., Nasr, M., Kumari, S., Stenström, T. A., Reddy, P., Pillay, K., & Bux, F. (2019). Decision tree for identification and prediction of filamentous bulking at full-scale activated sludge wastewater treatment plant. *Process Safety and Environmental Protection*, *126*. https://doi.org/10.1016/j.psep.2019.02.023

Dong, Q., Parker, W., & Dagnew, M. (2016). Influence of SRT and HRT on Bioprocess Performance in Anaerobic Membrane Bioreactors Treating Municipal Wastewater. *Water Environment Research*, *88*(2). https://doi.org/10.2175/106143016x14504669767175

Drewnowski, J., Remiszewska-Skwarek, A., Duda, S., & Łagód, G. (2019). Aeration process in bioreactors as the main energy consumer in a wastewater treatment plant. Review of solutions and methods of process optimization. *Processes*, *7*(5). https://doi.org/10.3390/pr7050311

Engineering ToolBox. (2003). *Air - Composition and Molecular Weight*. Engineering ToolBox.

Engineering ToolBox. (2004a). *Pressure*. Engineering ToolBox.

Engineering ToolBox. (2004b). *Universal and Individual Gas Constants*. Engineering ToolBox.

Erickson, N., Mueller, J., Shirkov, A., Zhang, H., Larroy, P., Li, M., & Smola, A. (2020). AutoGluon-Tabular: Robust and Accurate AutoML for Structured Data. *ArXiv*.

Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining*.

Garcia-Ochoa, F., & Gomez, E. (2009). Bioreactor scale-up and oxygen transfer rate in microbial processes: An overview. In *Biotechnology Advances* (Vol. 27, Issue 2). https://doi.org/10.1016/j.biotechadv.2008.10.006

Garcia-Ochoa, F., Gomez, E., Santos, V. E., & Merchuk, J. C. (2010). Oxygen uptake rate in microbial processes: An overview. In *Biochemical Engineering Journal* (Vol. 49, Issue 3). https://doi.org/10.1016/j.bej.2010.01.011

Germain, E., Nelles, F., Drews, A., Pearce, P., Kraume, M., Reid, E., Judd, S. J., & Stephenson, T. (2007). Biomass effects on oxygen transfer in membrane bioreactors. *Water Research*, *41*(5), 1038–1044. https://doi.org/10.1016/j.watres.2006.10.020

Gigante, D., Oliveira, P., Fernandes, B., Lopes, F., & Novais, P. (2021). Unsupervised Learning Approach for pH Anomaly Detection in Wastewater Treatment Plants. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *12886 LNAI*. https://doi.org/10.1007/978-3-030-86271-8_49

Gillot, S., & Héduit, A. (2008). Prediction of alpha factor values for fine pore aeration systems. *Water Science and Technology*, *57*(8). https://doi.org/10.2166/wst.2008.222

Government of Ontario. (2016). *Water and Energy Conservation Guidance Manual for Sewage Works*.

Granata, F., Papirio, S., Esposito, G., Gargano, R., & de Marinis, G. (2017). Machine learning algorithms for the forecasting of wastewater quality indicators. *Water (Switzerland), 9*(2). https://doi.org/10.3390/w9020105

Gujer, W., Henze, M., Mino, T., Matsuo, T., Wentzel, M. C., & Marais, G. v. R. (1995). The Activated Sludge Model No. 2: Biological phosphorus removal. *Water Science and Technology*, *31*(2). https://doi.org/10.1016/0273-1223(95)00175-M

Günder, B. (2000). *The membrane-coupled activated sludge process in municipal wastewater treatment*. CRC Press.

Guo, H., Jeong, K., Lim, J., Jo, J., Kim, Y. M., Park, J. pyo, Kim, J. H., & Cho, K. H. (2015). Prediction of effluent concentration in a wastewater treatment plant using machine learning models. *Journal of Environmental Sciences (China), 32*. https://doi.org/10.1016/j.jes.2015.01.007

Harpaz, C., Russo, S., Leitão, J. P., & Penn, R. (2022). Potential of supervised machine learning algorithms for estimating the impact of water efficient scenarios on solids accumulation in sewers. *Water Research, 216,* 118247. https://doi.org/10.1016/J.WATRES.2022.118247

Henderson, M. (2002). Energy Reduction Methods in the Aeration Process at Perth Waste Water Treatment Plant. *University of Strathclyde, Renewable Energy Systems and the Environment MSc Individual Theses*.

Henkel, J., Cornel, P., & Wagner, M. (2011). Oxygen transfer in activated sludge – new insights and potentials for cost saving. *Water Science and Technology, 63*(12), 3034–3038. https://doi.org/10.2166/wst.2011.607

Henriques, J., & Catarino, J. (2017). Sustainable value – An energy efficiency indicator in wastewater treatment plants. *Journal of Cleaner Production*, *142*. https://doi.org/10.1016/j.jclepro.2016.03.173

Henze, M., Grady, C. P. L., Gujer, W., Marais, G., & Matsuo, T. (1987). Activated sludge model No. 1, IAWQ Scientific and Technical report. *N*.

Henze, M., Gujer, W., Mino, T., Matsuo, T., Wentzel, M. C., Marais, G. V. R., & van Loosdrecht, M. C. M. (1999). Activated Sludge Model No.2d, ASM2d. *Water Science and Technology*, *39*(1). https://doi.org/10.1016/S0273-1223(98)00829-4

Holman, J. B., & Wareham, D. G. (2005). COD, ammonia and dissolved oxygen time profiles in the simultaneous nitrification/denitrification process. *Biochemical Engineering Journal*, *22*(2). https://doi.org/10.1016/j.bej.2004.09.001

Hydromantis Environmental Software Services, I. (2015). *GPS-X Technial Reference* . Hydromantis Environmental Software Services, Inc.

Hydromantis Environmental Software Solutions. (2013). *Adelaide Wastewater Treatment Plant - GPS-X Model Calibration Report*.

Iranpour, R., Magallanes, A., Zermeño, M., Moghaddam, O., Wilson, J., & Stenstrom, M. K. (2000). Assessment of Aeration System Performance Efficiency: Frequent Sampling for Damage Detection. *Water Environment Research*, *72*(3). https://doi.org/10.2175/106143000x137590

Iranpour, R., Magallanes, A., Zermeño, M., Varsh, V., Abrishamchi, A., & Stenstrom, M. (2000). Assessment of aeration basin performance efficiency: Sampling methods and tank coverage. *Water Research*, *34*(12). https://doi.org/10.1016/S0043-1354(00)00065-8

Jenkins, T. E. (2013). *Aeration control system design: a practical guide to energy and process optimization*. John Wiley & Sons.

Jiang, L. M., Garrido-Baserba, M., Nolasco, D., Al-Omari, A., DeClippeleir, H., Murthy, S., & Rosso, D. (2017). Modelling oxygen transfer using dynamic alpha factors. *Water Research*, *124*. https://doi.org/10.1016/j.watres.2017.07.032

K. Al-Ahmady, K. (2011). Mathematical Model for Calculating Oxygen Mass Transfer Coefficient in Diffused Air Systems. *AL-Rafdain Engineering Journal (AREJ)*, *19*(4). https://doi.org/10.33899/rengj.2011.26795

Kang, H. (2013). The prevention and handling of the missing data. In *Korean Journal of Anesthesiology* (Vol. 64, Issue 5). https://doi.org/10.4097/kjae.2013.64.5.402

Khan, M. B., Nisar, H., Ng, C. A., Lo, P. K., & Yap, V. V. (2018). Generalized classification modeling of activated sludge process based on microscopic image analysis. *Environmental Technology (United Kingdom)*, *39*(1). https://doi.org/10.1080/09593330.2017.1293166

Kim, M., Kim, Y., Kim, H., Piao, W., & Kim, C. (2016). Evaluation of the k-nearest neighbor method for forecasting the influent characteristics of wastewater treatment plant. *Frontiers of Environmental Science and Engineering*, *10*(2). https://doi.org/10.1007/s11783-015-0825-7

Kişi, Ö. (2004). River Flow Modeling Using Artificial Neural Networks. *Journal of Hydrologic Engineering*, *9*(1). https://doi.org/10.1061/(asce)1084-0699(2004)9:1(60)

Krampe, J., & Krauth, K. (2003). Oxygen transfer into activated sludge with high MLSS concentrations. *Water Science and Technology*, *47*(11). https://doi.org/10.2166/wst.2003.0618

Lee, J. (2017). Development of a model to determine mass transfer coefficient and oxygen solubility in bioreactors. *Heliyon*, *3*(2). https://doi.org/10.1016/j.heliyon.2017.e00248

Leu, S.-Y., Rosso, D., Larson, L. E., & Stenstrom, M. K. (2009). Real-Time Aeration Efficiency Monitoring in the Activated Sludge Process and Methods to Reduce Energy Consumption and Operating Costs. *Water Environment Research*, *81*(12). https://doi.org/10.2175/106143009x425906

Lorenzo-Toja, Y., Vázquez-Rowe, I., Amores, M. J., Termes-Rifé, M., Marín-Navarro, D., Moreira, M. T., & Feijoo, G. (2016). Benchmarking wastewater treatment plants under an eco-efficiency perspective. *Science of the Total Environment*, *566–567*. https://doi.org/10.1016/j.scitotenv.2016.05.110

Martin, C., & Vanrolleghem, P. A. (2014). Analysing, completing, and generating influent data for WWTP modelling: A critical review. *Environmental Modelling & Software*, *60*, 188–201. https://doi.org/10.1016/j.envsoft.2014.05.008

Meng, F., Yang, A., Zhang, G., & Wang, H. (2017). Effects of dissolved oxygen concentration on photosynthetic bacteria wastewater treatment: Pollutants removal, cell growth and

pigments production. *Bioresource Technology*, *241*. https://doi.org/10.1016/j.biortech.2017.05.183

Mueller, J. A., Boyle, W. C., & Pöpel, H. J. (2002). Aeration: Principles and practice. In *Aeration: Principles and Practice* (Vol. 11).

Newhart, K. B., Holloway, R. W., Hering, A. S., & Cath, T. Y. (2019). Data-driven performance analyses of wastewater treatment plants: A review. In *Water Research* (Vol. 157). https://doi.org/10.1016/j.watres.2019.03.030

Nowak, O. (1999). Considerations on costs and implementation of nutrient removal technologies in CEE countries. *International Conference on EU Water Management Framework Directive and Danubian Countries*.

Olden, J. D., & Jackson, D. A. (2002). Illuminating the "black box": A randomization approach for understanding variable contributions in artificial neural networks. *Ecological Modelling*, *154*(1–2). https://doi.org/10.1016/S0304-3800(02)00064-9

Panepinto, D., Fiore, S., Zappone, M., Genon, G., & Meucci, L. (2016). Evaluation of the energy efficiency of a large wastewater treatment plant in Italy. *Applied Energy*, *161*. https://doi.org/10.1016/j.apenergy.2015.10.027

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*.

Pittoors, E., Guo, Y., & van Hulle, S. W. H. (2014). Modeling Dissolved Oxygen Concentration for Optimizing Aeration Systems and Reducing Oxygen Consumption in Activated Sludge Processes: A Review. In *Chemical Engineering Communications* (Vol. 201, Issue 8). https://doi.org/10.1080/00986445.2014.883974

PubChem. (n.d.). *PubChem Compound Summary for CID 977, Oxygen*. National Center for Biotechnology Information.

Ribeiro, D., Sanfins, A., & Belo, O. (2013). Wastewater treatment plant performance prediction with support vector machines. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *7987 LNAI*. https://doi.org/10.1007/978-3-642-39736-3_8

Rieger, L., Jones, R. M., Dold, P. L., & Bott, C. B. (2013). Ammonia-Based Feedforward and Feedback Aeration Control in Activated Sludge Processes. *Water Environment Research*, *86*(1). https://doi.org/10.2175/106143013x13596524516987

Rieger, L., Takács, I., Villez, K., Siegrist, H., Lessard, P., Vanrolleghem, P. A., & Comeau, Y. (2010). Data Reconciliation for Wastewater Treatment Plant Simulation Studies-Planning for High-Quality Data and Typical Sources of Errors. *Water Environment Research*, *82*(5). https://doi.org/10.2175/106143009x12529484815511

Rodríguez, F. A., Reboleiro-Rivas, P., Osorio, F., Martínez-Toledo, M. v., Hontoria, E., & Poyatos, J. M. (2012). Influence of mixed liquid suspended solids and hydraulic retention time on oxygen transfer efficiency and viscosity in a submerged membrane bioreactor

using pure oxygen to supply aerobic conditions. *Biochemical Engineering Journal*, *60*. https://doi.org/10.1016/j.bej.2011.10.016

Rosso, D., Iranpour, R., & Stenstrom, M. K. (2005). Fifteen Years of Offgas Transfer Efficiency Measurements on Fine-Pore Aerators: Key Role of Sludge Age and Normalized Air Flux. *Water Environment Research*, *77*(3). https://doi.org/10.2175/106143005x41843

Rosso, D., Larson, L. E., & Stenstrom, M. K. (2006). Surfactant effects on alpha factors in full-scale wastewater aeration systems. *Water Science and Technology*, *54*(10). https://doi.org/10.2166/wst.2006.768

Rosso, D., Larson, L. E., & Stenstrom, M. K. (2008). Aeration of large-scale municipal wastewater treatment plants: State of the art. In *Water Science and Technology* (Vol. 57, Issue 7). https://doi.org/10.2166/wst.2008.218

Rosso, D., & Stenstrom, M. K. (2006). Economic Implications of Fine-Pore Diffuser Aging. *Water Environment Research*, *78*(8). https://doi.org/10.2175/106143006x101683

Rossum, G. van, & Drake, F. L. (2006). Python Reference Manual. *October*, *22*.

Samer, M. (2015). Biological and Chemical Wastewater Treatment Processes. In *Wastewater Treatment Engineering*. https://doi.org/10.5772/61250

Sardeing, R., Painmanakul, P., & Hébrard, G. (2006). Effect of surfactants on liquid-side mass transfer coefficients in gas-liquid systems: A first step to modeling. *Chemical Engineering Science*, *61*(19). https://doi.org/10.1016/j.ces.2006.05.051

Schierholz, E. L., Gulliver, J. S., Wilhelms, S. C., & Henneman, H. E. (2006). Gas transfer from air diffusers. *Water Research, 40*(5). https://doi.org/10.1016/j.watres.2005.12.033

Sözüdoğru, O., Massara, T. M., Çalık, S., Yılmaz, A. E., Bakırdere, S., Katsou, E., & Komesli, O. T. (2020). Influence of Hydraulic Retention Time (HRT) upon the Treatment of Wastewater by a Laboratory-Scale Membrane Bioreactor (MBR). *Analytical Letters, 54*(10). https://doi.org/10.1080/00032719.2020.1815756

Stenstrom, M. K., & Gilbert, R. G. (1981). Effects of alpha, beta and theta factor upon the design, specification and operation of aeration systems. *Water Research, 15*(6). https://doi.org/10.1016/0043-1354(81)90156-1

Stenstrom, M. K., Leu, S.-Y. (Ben), & Jiang, P. (2014). Theory to Practice: Oxygen Transfer and the New ASCE Standard. *Proceedings of the Water Environment Federation*, *2006*(7). https://doi.org/10.2175/193864706783762931

Steven A. Bolles. (2006). *MODELING WASTEWATER AERATION SYSTEMS TO DISCOVER ENERGY SAVINGS OPPORTUNITIES* . Retrieved April 5, 2022, from https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.539.9955&rep=rep1&type=pdf

Thunberg, A., Sundin, A., & Carlsson, B. (2009). Energy optimization of the aeration process at Käppala wastewater treatment plant. *10th IWA Conference on Instrumentation, Control and Automation*.

United States Environmental Protection Agency. (1989). *Fine Pore Aeration Systems*.

United States Environmental Protection Agency. (2010). *Evaluation of Energy Conservation Measures for Wastewater Treatment Facilities*. EPA.

Vestby, M. L. (2020). *Data Mining in Norwegian Level-of-Living Survey Data* [Master's thesis]. The University of Bergen.

VLMP. (2014). Maximum dissolved oxygen concentration saturation table. *lakestewardsofmaine.org*. https://lakestewardsofmaine.org/wp-content/uploads/2014/01/Maximum-Dissolved-Oxygen-Concentration-Saturation-Table.pdf

Wang, R., Yu, Y., Chen, Y., Pan, Z., Li, X., Tan, Z., & Zhang, J. (2022). Model construction and application for effluent prediction in wastewater treatment plant: Data processing method optimization and process parameters integration. *Journal of Environmental Management*, *302*. https://doi.org/10.1016/j.jenvman.2021.114020

Water Environment Federation. (1998). Design of municipal wastewater treatment plants: WEF manual of practice 8. In *WEF Manual of Practice No. 8*.

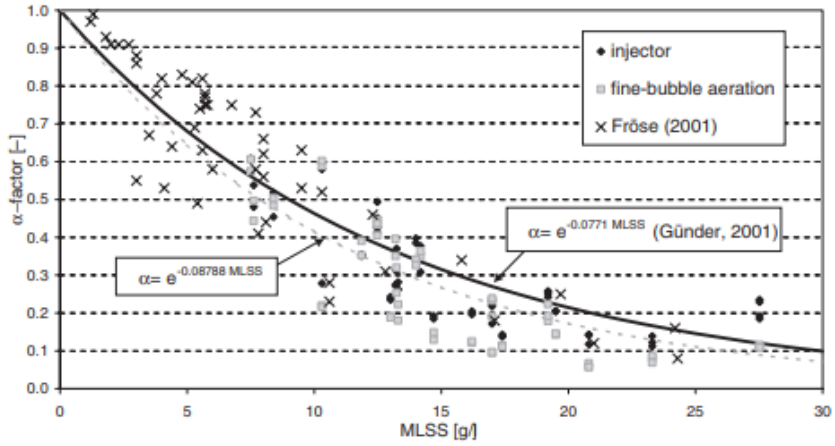
Yang, Y. H. (2006). Support Vector Machines for Environmental Informatics: Application to Modelling the Nitrogen Removal Processes in Wastewater Treatment Systems. *Journal of Environmental Informatics*, *7*(1). https://doi.org/10.3808/jei.200600063

Zuluaga-Bedoya, C., Ruiz-Botero, M., Ospina-Alarcón, M., & Garcia-Tirado, J. (2018). A dynamical model of an aeration plant for wastewater treatment using a phenomenological based semi-physical modeling methodology. *Computers and Chemical Engineering*, *117*. https://doi.org/10.1016/j.compchemeng.2018.07.008

Appendices

Appendix A Plots of alpha empirical equation studies

Plots of alpha empirical equation studies



a) (Krampe and Kreauth 2003)

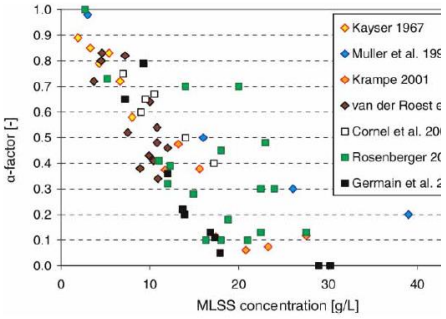


Figure 1 | α-factor dependency on MLSS concentration.

b) (Henkel, Cornel and Wagner 2011)

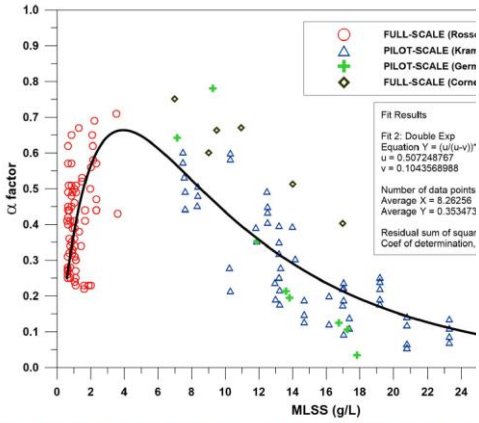


Figure 1—Dependence of the α-factor on the MLSS concentration. http://wst.iw 47/11/313

c) (Baquero-Rodriguez, et al. 2018)

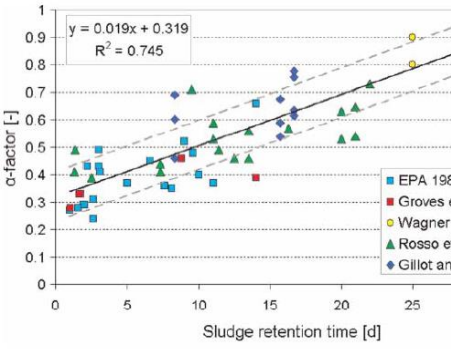
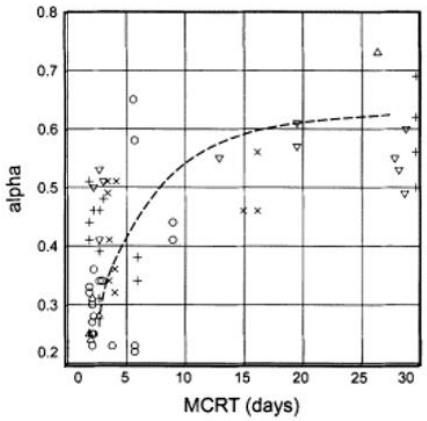


Figure 3 | Linear regression of α-factor and SRT (2 g/L $\leq$ MLVSS $\leq$ 4

d) (Henkel, Cornel and Wagner 2011)



e) (Rosso, Iranpour and Stenstrom 2005)

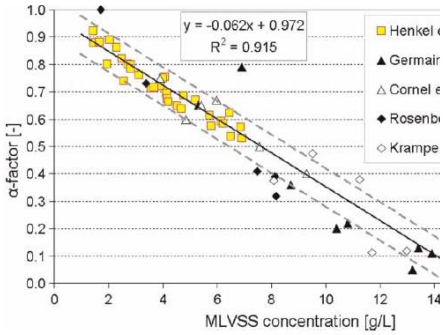
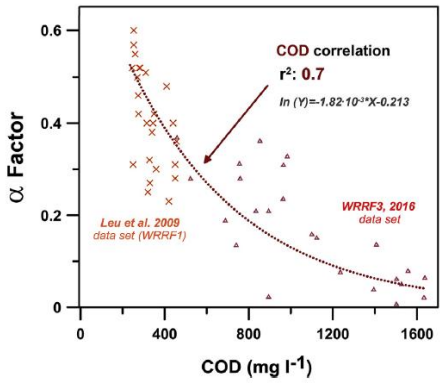


Figure 2 | Linear regression of α-factor and MLVSS concentration (S

f) (Henkel, Cornel and Wagner 2011)



g) (Jiang, Garrido-Baserba and Nolasco 2017)

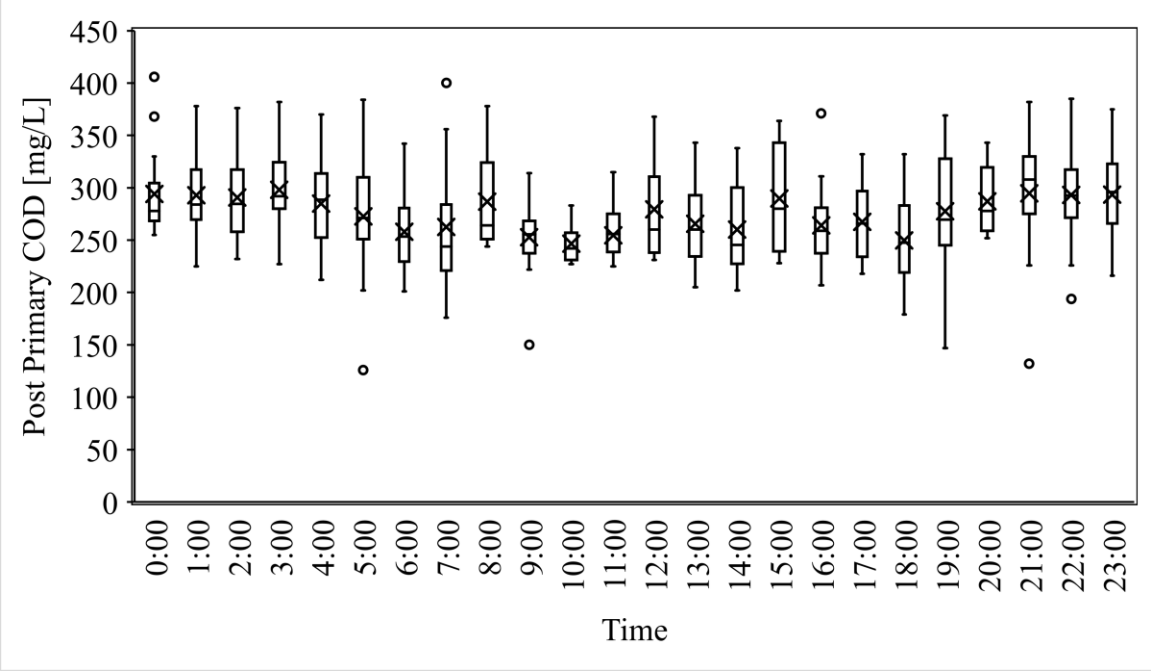
Figure SD 1. alpha factor dependency based on literature between 2003 and 2018

Table SD1. Typical K_La correction factors for domestic wastewater (Bewtra et al., 1970; Tewari and Bewtra, 1982)

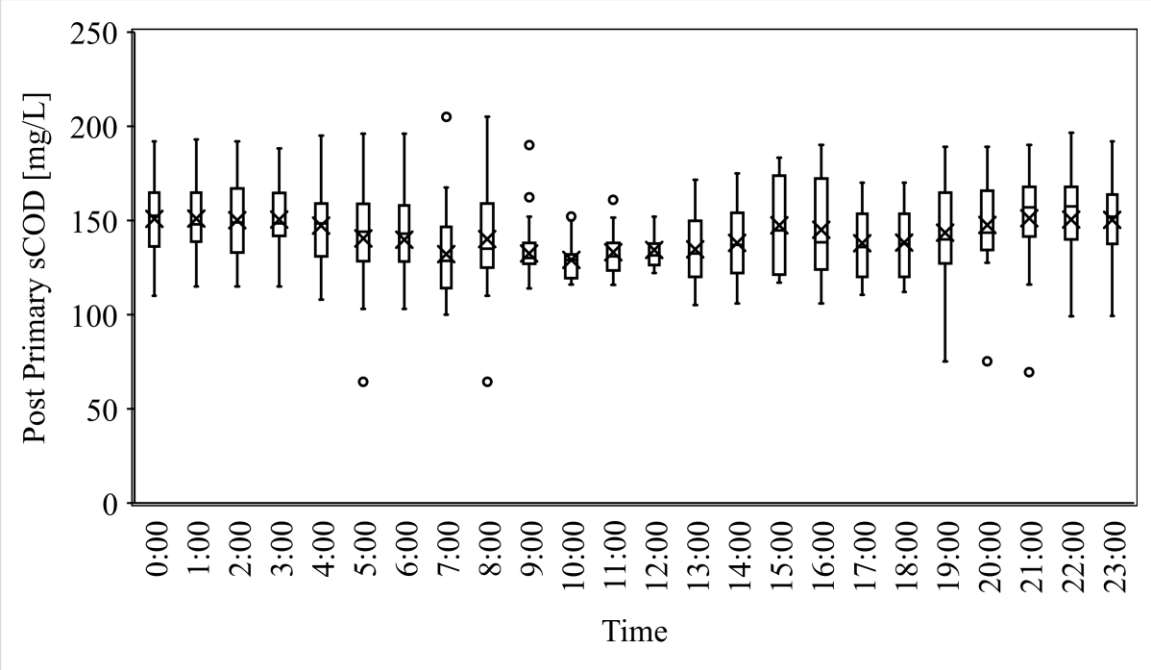
Correction factor	Equation	Typical values
Alpha factor	$\alpha = \frac{(K_L a)_{ww}}{K_L a}$	0.3-1.2
Beta factor: salinity-surface tension	$\beta = \frac{C^*_{wastewater}}{C^*_{cleanwater}}$	0.7-1.0
Gamma factor: temperature factor	$\gamma = \theta^{20°C-T}$	θ = 1.016-1.024

Appendix B Figures SD2(a) to SD2(i) show average daily fluctuations in wastewater, aeration tank and operating parameters

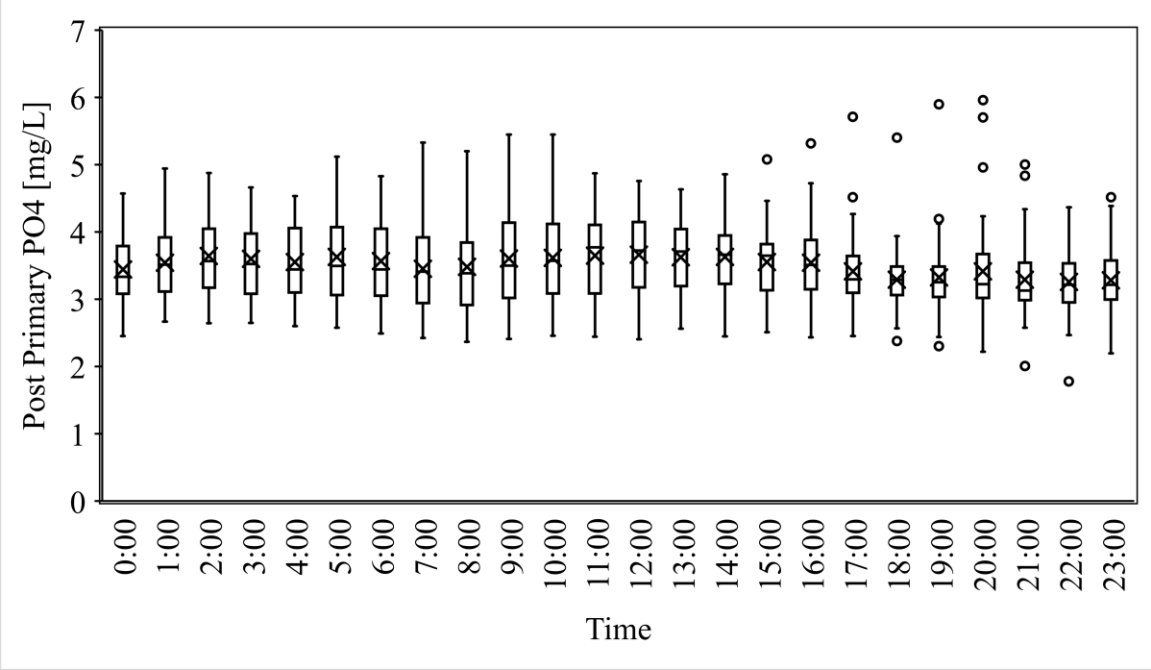
(a)



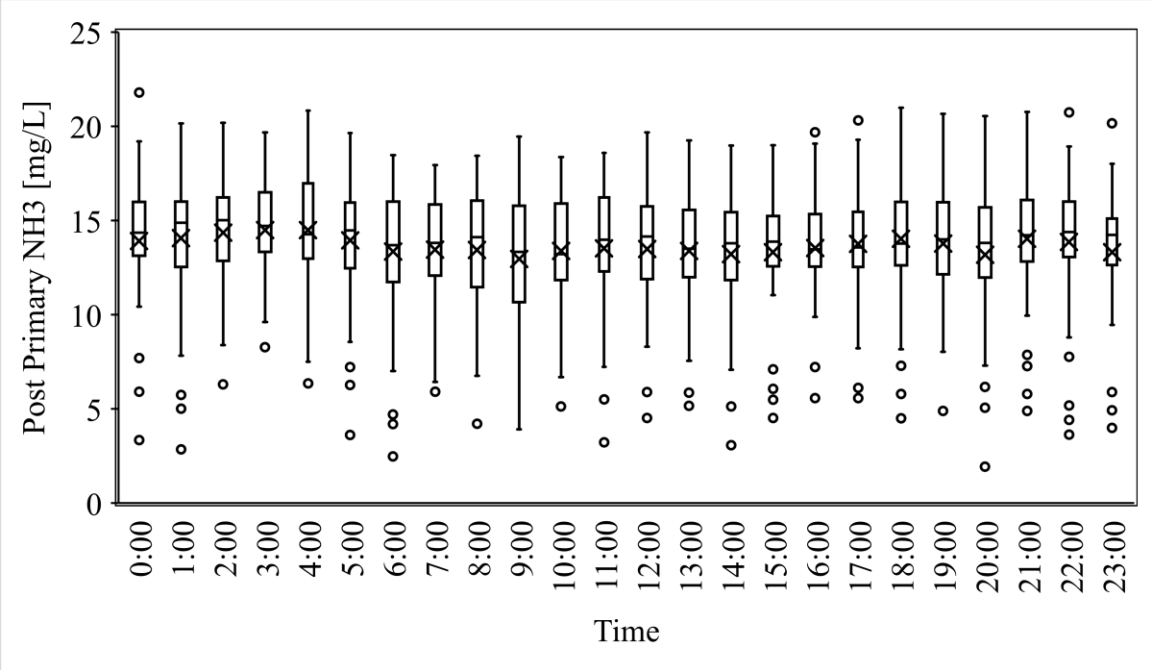
(b)


Post Primary sCOD [mg/L]
250
200
150
100
50
0
0:00
1:00
2:00
3:00
4:00
5:00
6:00
7:00
8:00
9:00
10:00
11:00
12:00
13:00
14:00
15:00
16:00
17:00
18:00
19:00
20:00
21:00
22:00
23:00
Time

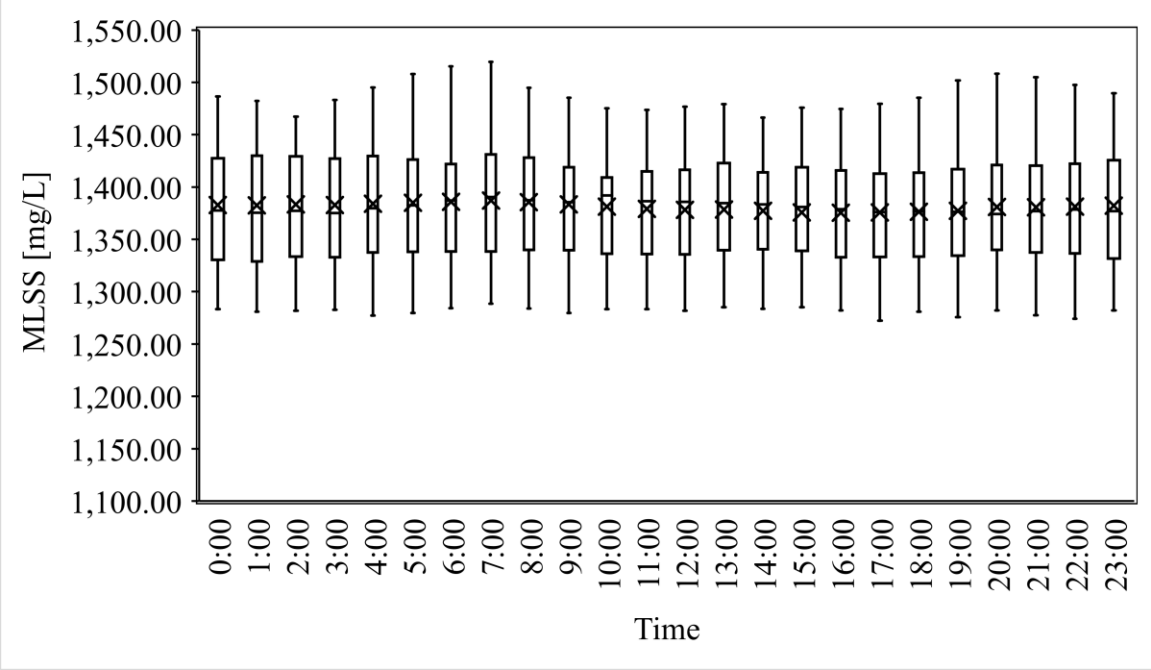
(c)


Post Primary PO4 [mg/L]
7
6
5
4
3
2
1
0
0:00
1:00
2:00
3:00
4:00
5:00
6:00
7:00
8:00
9:00
10:00
11:00
12:00
13:00
14:00
15:00
16:00
17:00
18:00
19:00
20:00
21:00
22:00
23:00
Time

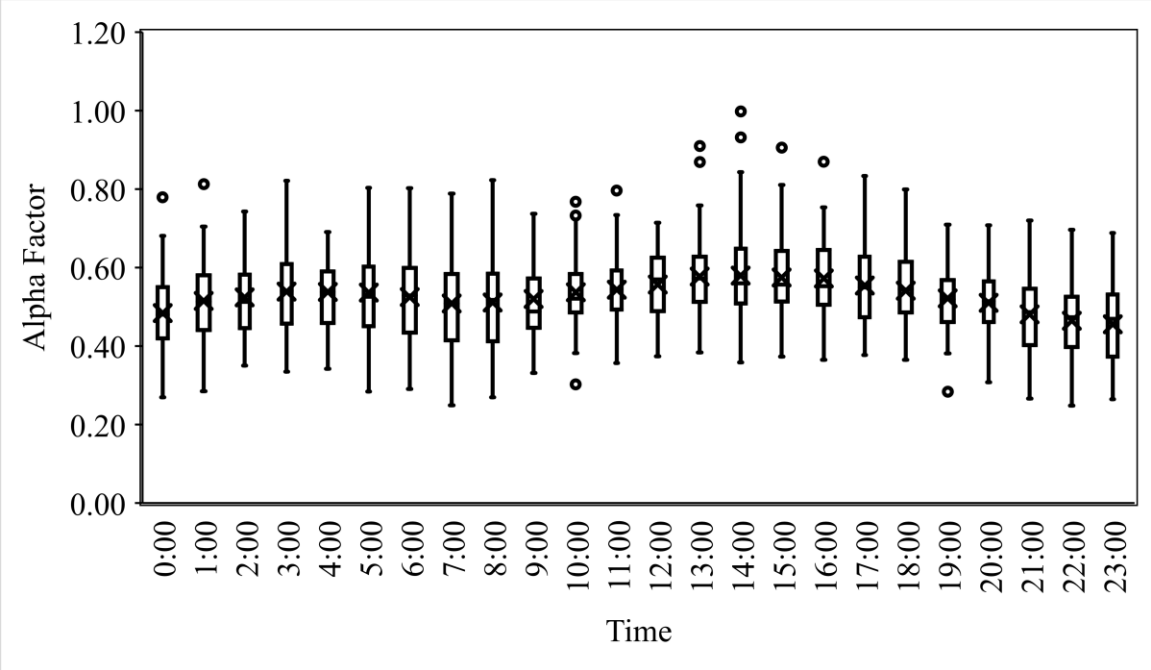
(d)


Post Primary NH3 [mg/L]
25
20
15
10
5
0
0:00
1:00
2:00
3:00
4:00
5:00
6:00
7:00
8:00
9:00
10:00
11:00
12:00
13:00
14:00
15:00
16:00
17:00
18:00
19:00
20:00
21:00
22:00
23:00
Time

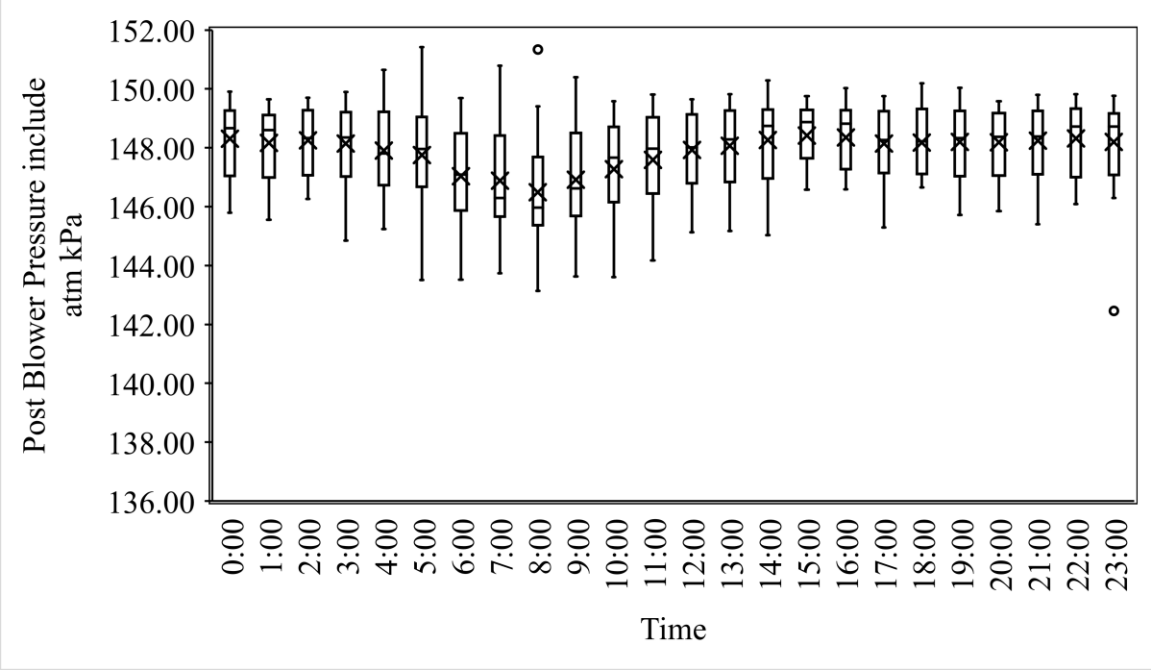
(e)


MLSS [mg/L]
1,550.00
1,500.00
1,450.00
1,400.00
1,350.00
1,300.00
1,250.00
1,200.00
1,150.00
1,100.00
0:00
1:00
2:00
3:00
4:00
5:00
6:00
7:00
8:00
9:00
10:00
11:00
12:00
13:00
14:00
15:00
16:00
17:00
18:00
19:00
20:00
21:00
22:00
23:00
Time

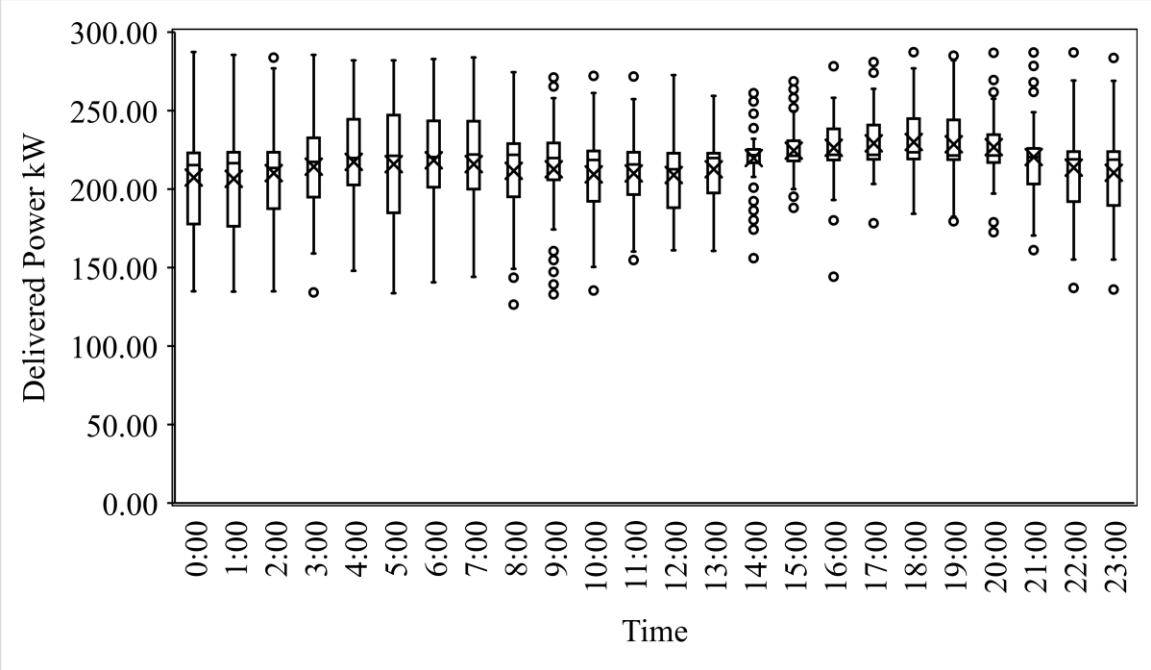
(f)


Alpha Factor
1.20
1.00
0.80
0.60
0.40
0.20
0.00
0:00
1:00
2:00
3:00
4:00
5:00
6:00
7:00
8:00
9:00
10:00
11:00
12:00
13:00
14:00
15:00
16:00
17:00
18:00
19:00
20:00
21:00
22:00
23:00
Time

(g)


Post Blower Pressure include atm kPa
152.00
150.00
148.00
146.00
144.00
142.00
140.00
138.00
136.00
0:00
1:00
2:00
3:00
4:00
5:00
6:00
7:00
8:00
9:00
10:00
11:00
12:00
13:00
14:00
15:00
16:00
17:00
18:00
19:00
20:00
21:00
22:00
23:00
Time

(h)


Delivered Power kW
300.00
250.00
200.00
150.00
100.00
50.00
0.00
0:00
1:00
2:00
3:00
4:00
5:00
6:00
7:00
8:00
9:00
10:00
11:00
12:00
13:00
14:00
15:00
16:00
17:00
18:00
19:00
20:00
21:00
22:00
23:00
Time

(i)

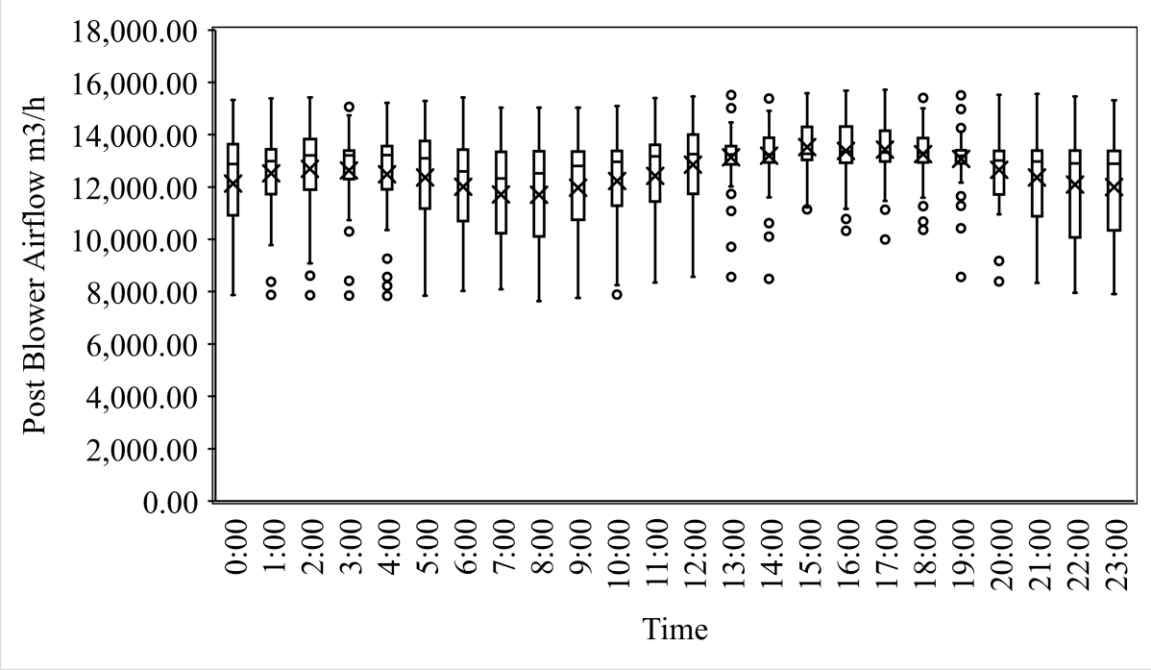
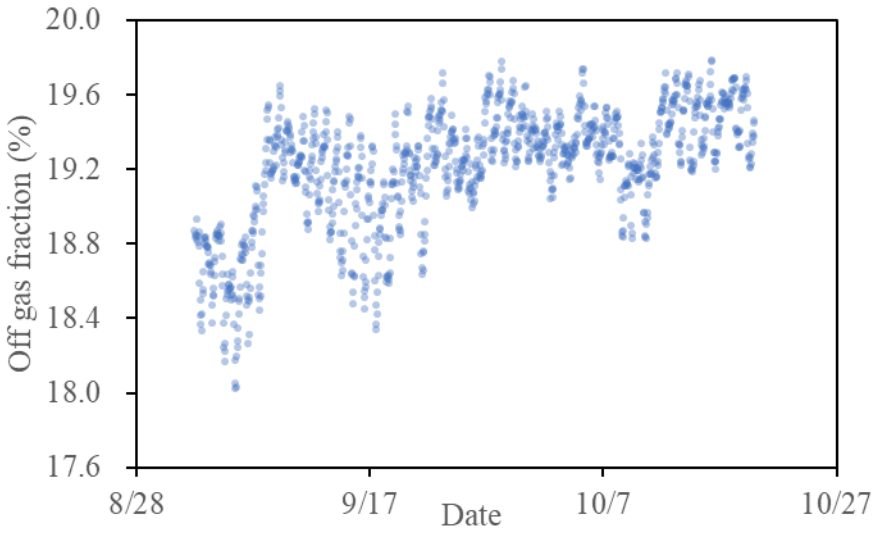

Post Blower Airflow m3/h
18,000.00
16,000.00
14,000.00
12,000.00
10,000.00
8,000.00
6,000.00
4,000.00
2,000.00
0.00
0:00
1:00
2:00
3:00
4:00
5:00
6:00
7:00
8:00
9:00
10:00
11:00
12:00
13:00
14:00
15:00
16:00
17:00
18:00
19:00
20:00
21:00
22:00
23:00
Time

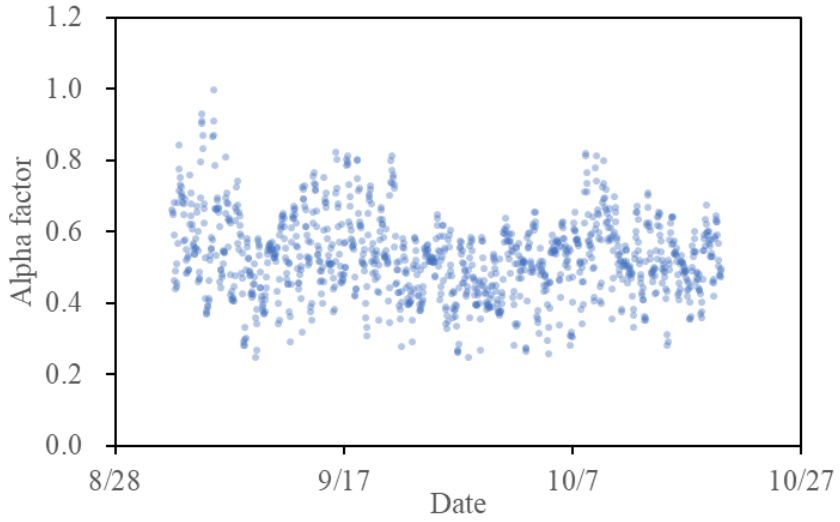
Figure SD2. Daily dynamic change of operational parameters in Adelaide PCP: (a) Post-primary COD (b) Post-primary sCOD (c) Post-primary PO4 (d) Post-primary NH_3 (e) MLSS (f) Alpha factor (g) Post blower pressure (h) Delivered power (i) Post blower airflow

Appendix C Figures SD3(a) to SD3(k) show hourly fluctuations in process and operating parameters during the study period.

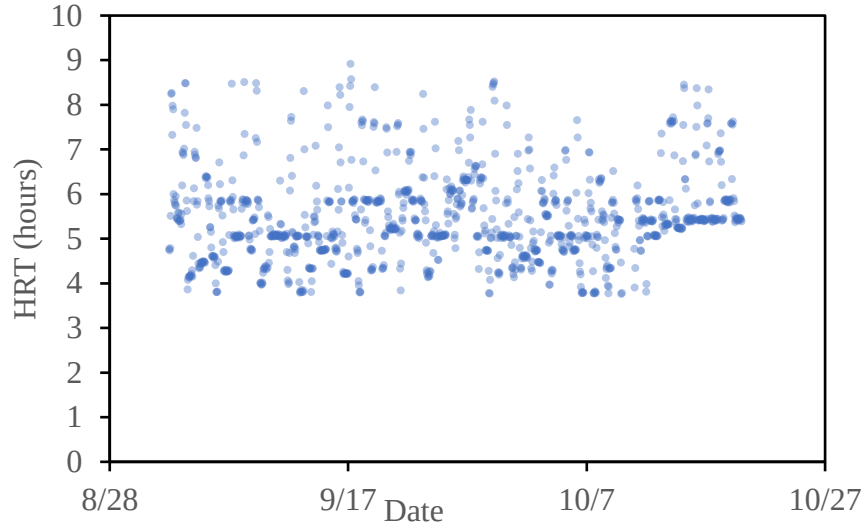
(a)



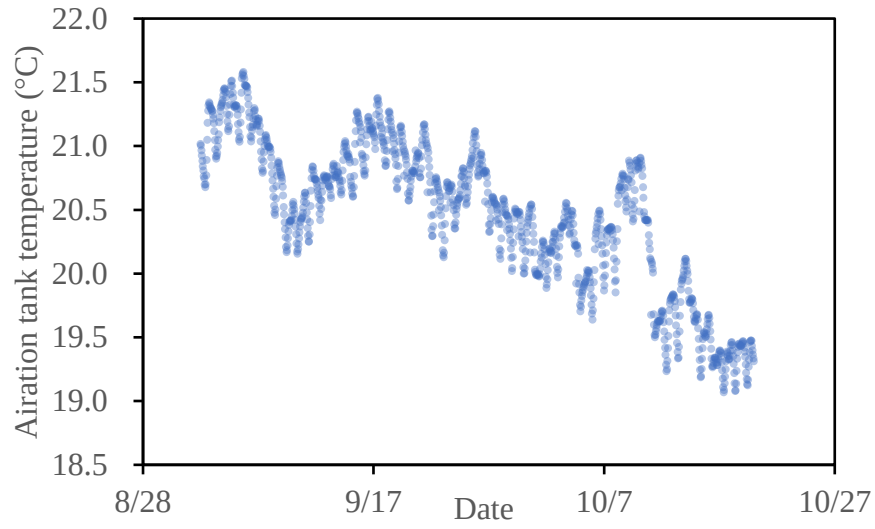
(b)


1.2
1.0
0.8
0.6
0.4
0.2
0.0
Alpha factor
8/28
9/17
10/7
10/27
Date

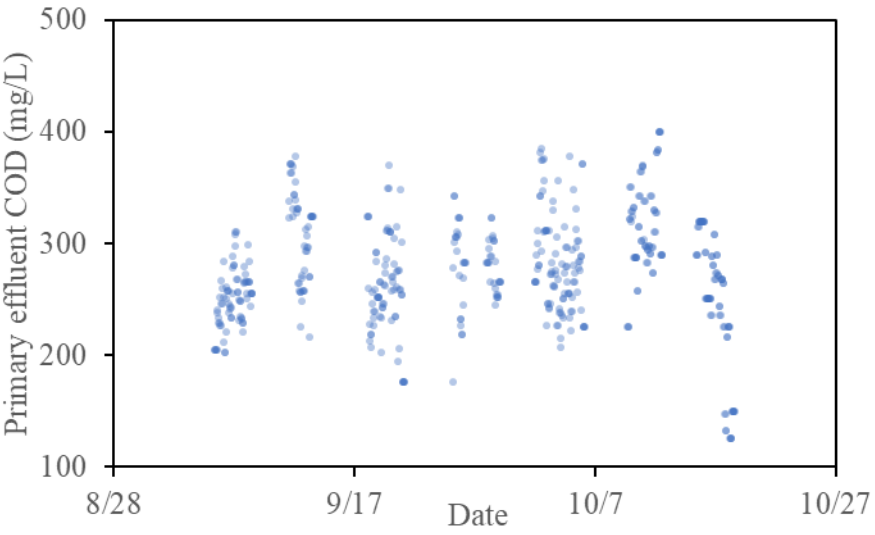
(c)


10
9
8
7
6
5
4
3
2
1
0
HRT (hours)
8/28
9/17
Date
10/7
10/27

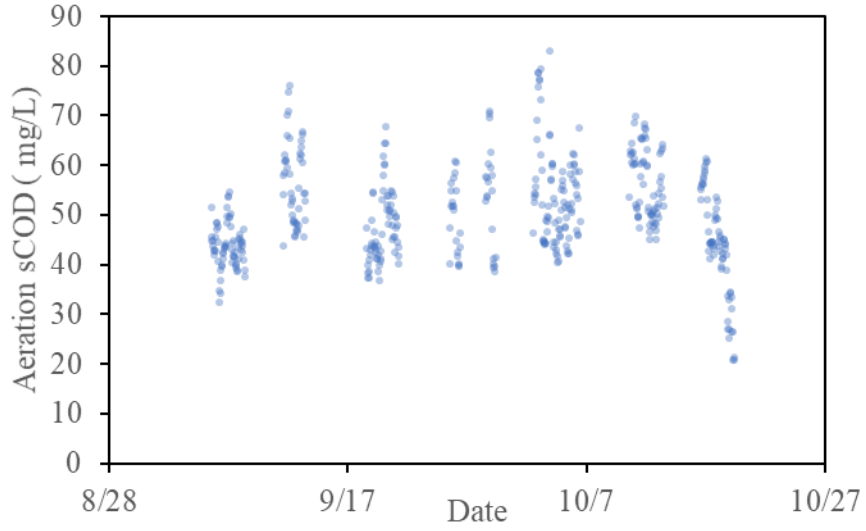
(d)


Airation tank temperature (°C)
22.0
21.5
21.0
20.5
20.0
19.5
19.0
18.5
8/28
9/17
Date
10/7
10/27

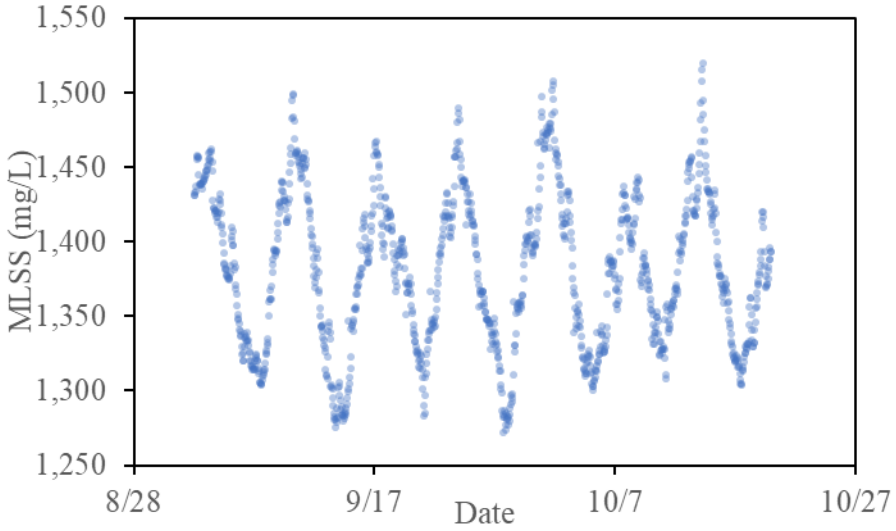
(e)


Primary effluent COD (mg/L)
500
400
300
200
100
8/28
9/17
Date
10/7
10/27

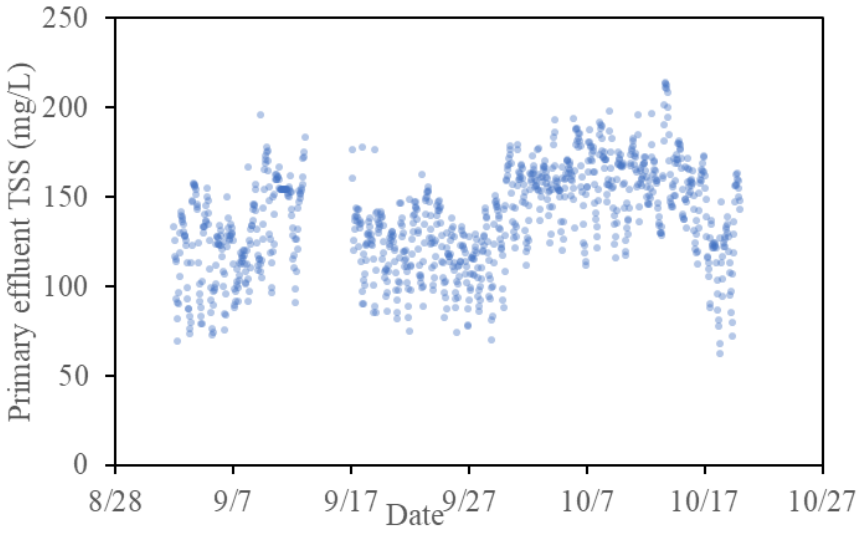
(f)


Aeration sCOD (mg/L)
90
80
70
60
50
40
30
20
10
0
8/28
9/17
Date
10/7
10/27

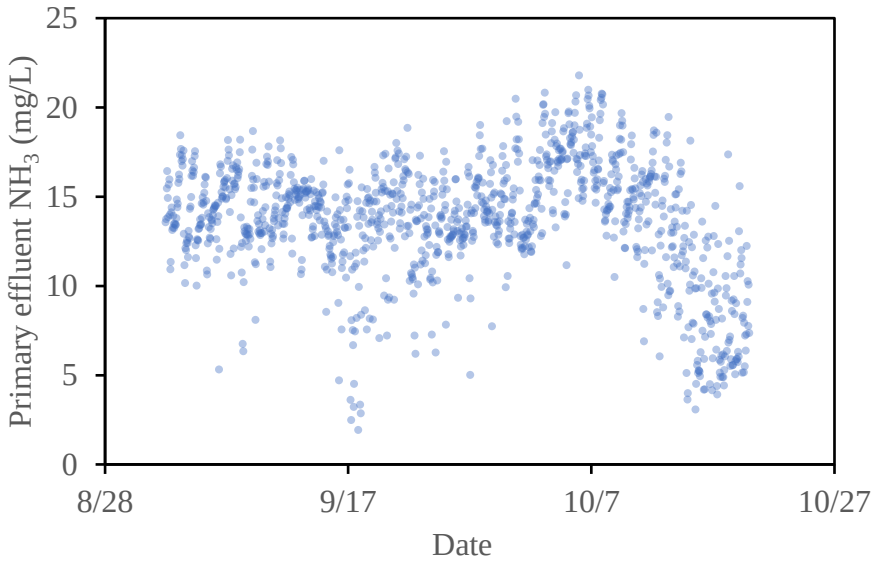
(g)


MLSS (mg/L)
1,550
1,500
1,450
1,400
1,350
1,300
1,250
8/28
9/17
Date
10/7
10/27

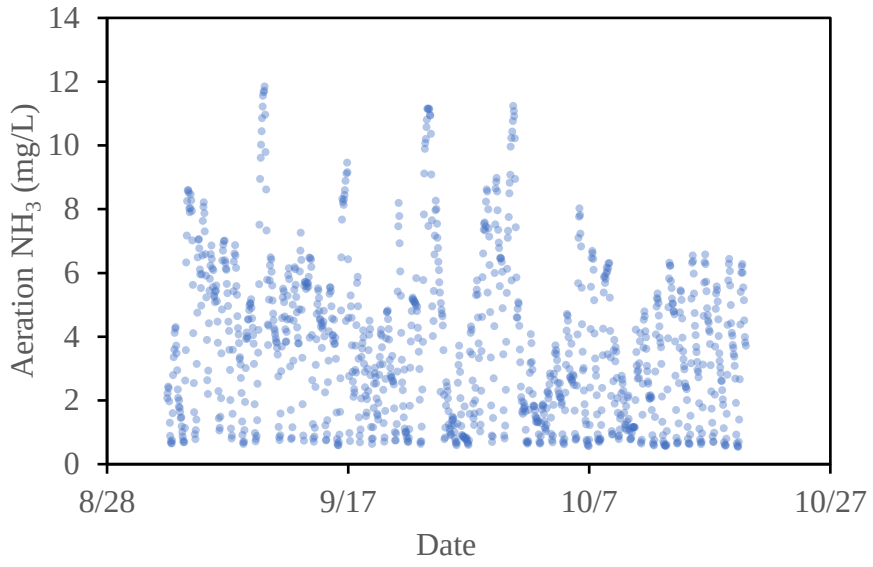
(h)


250
200
150
100
50
0
Primary effluent TSS (mg/L)
8/28
9/7
9/17
9/27
10/7
10/17
10/27
Date

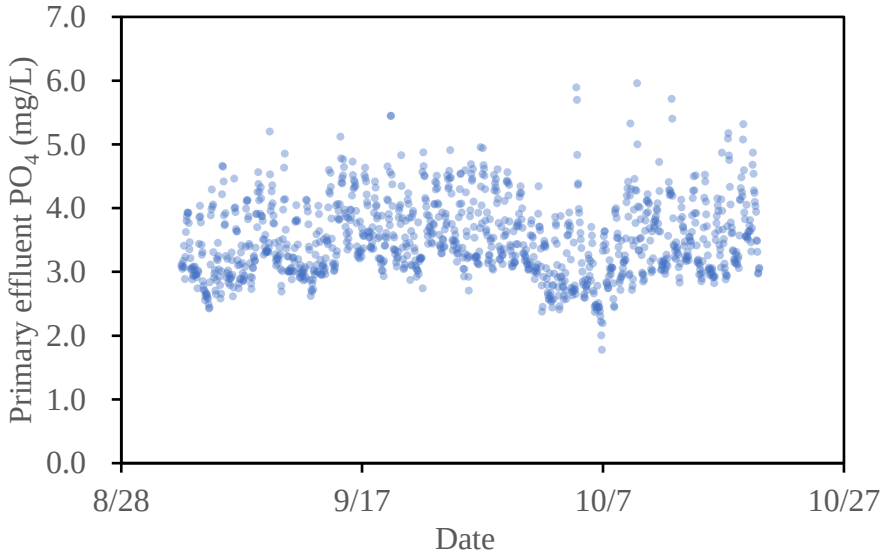
(i)


25
20
15
10
5
0
Primary effluent NH3 (mg/L)
8/28
9/17
10/7
10/27
Date

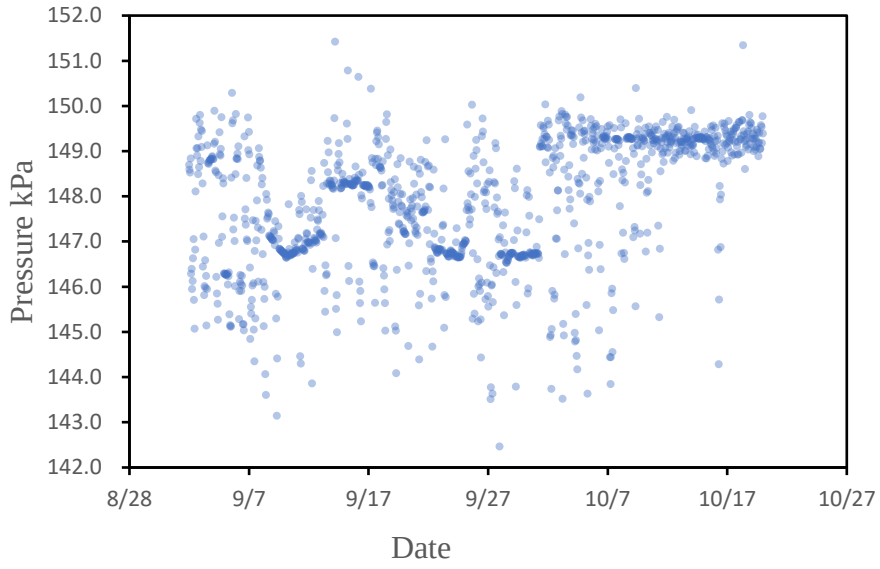
(j)


14
12
10
8
6
4
2
0
Aeration NH3 (mg/L)
8/28
9/17
10/7
10/27
Date

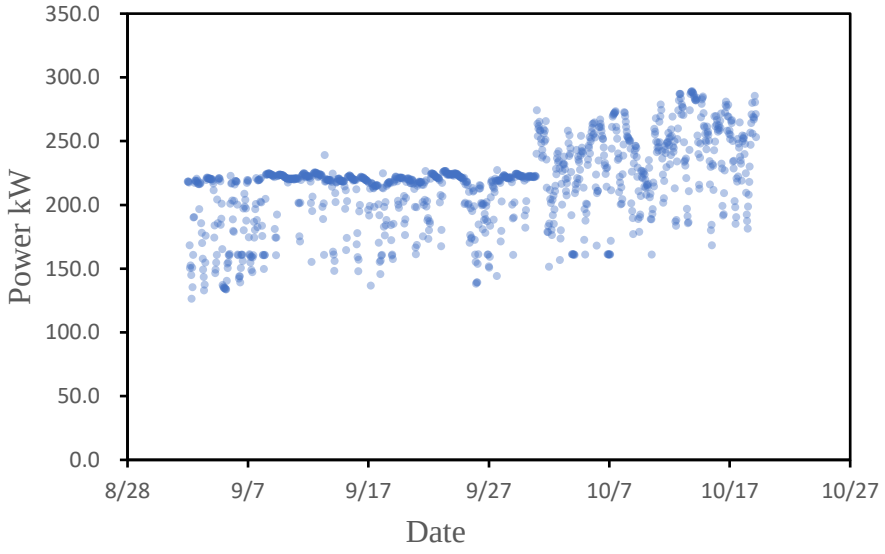
(k)


7.0
6.0
5.0
4.0
3.0
2.0
1.0
0.0
Primary effluent PO4 (mg/L)
8/28
9/17
10/7
10/27
Date

(l)


152.0
151.0
150.0
149.0
148.0
147.0
146.0
145.0
144.0
143.0
142.0
Pressure kPa
8/28
9/7
9/17
9/27
10/7
10/17
10/27
Date

(m)


350.0
300.0
250.0
200.0
150.0
100.0
50.0
0.0
Power kW
8/28
9/7
9/17
9/27
10/7
10/17
10/27
Date

(n)

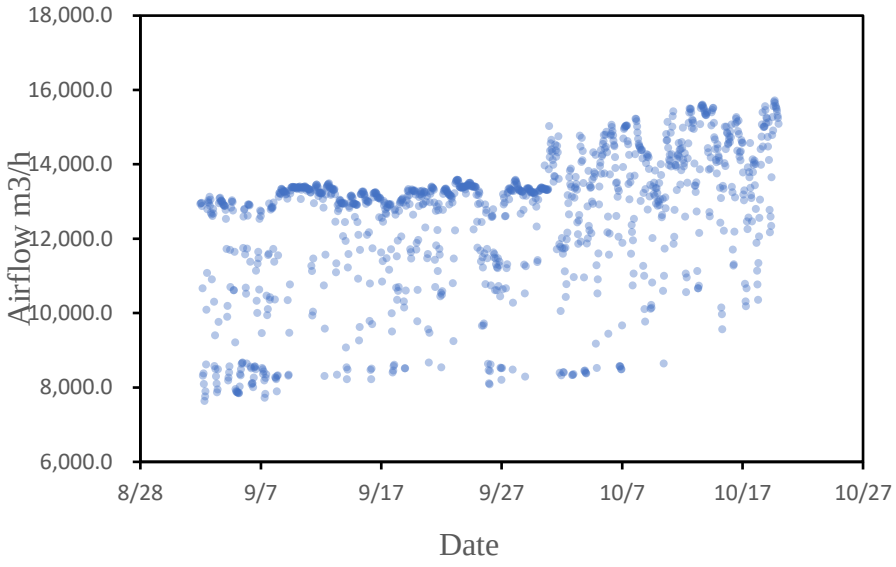
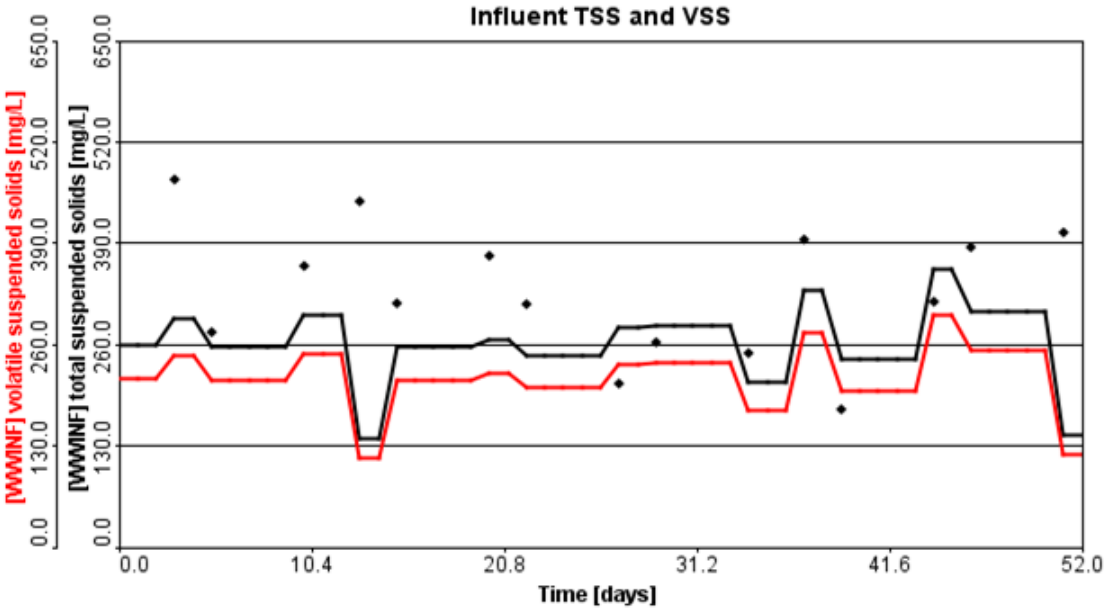


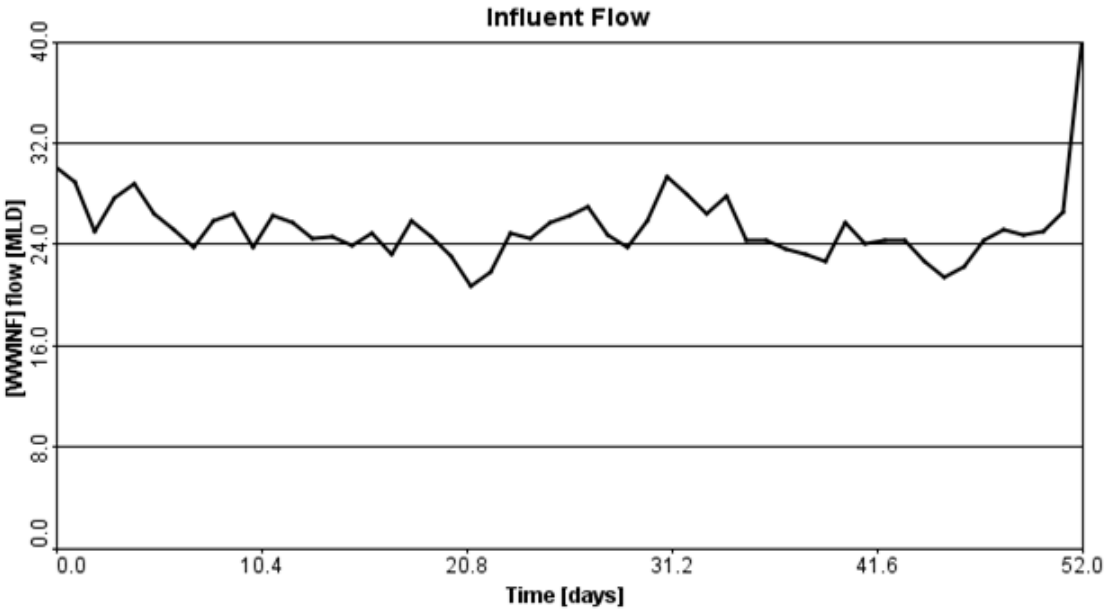
Figure SD3 Wastewater and operating condition profile during the study period: (a) off-gas fraction (b) dynamic alpha (c) HRT (d) Temperature (e) Primary effluent COD (f) Aeration tank sCOD (g) Primary effluent TSS (h) MLSS (i) Primary effluent NH_3 (j) Aeration tank NH_3 (k) Primary effluent PO_4. (l) Post blower pressure with atm (m) Delivered Power. (n) Airflow Figures SD3a to SD3k show hourly fluctuations in process and operating parameters during the study period.

Appendix D Figures SD4(a) to SD4(p) exported operational parameters from GPS-X model (input parameters: inf BOD, inf TSS, inf flow, inf PO4-P, inf NH4-N, Pri eff TSS, Pri eff BOD5, MLSS, MLVSS, DO, RAS, WAS, eff BOD, eff PO4-P, eff NH4-N, eff NO3-N, eff NO2-N, eff TSS, temperature)

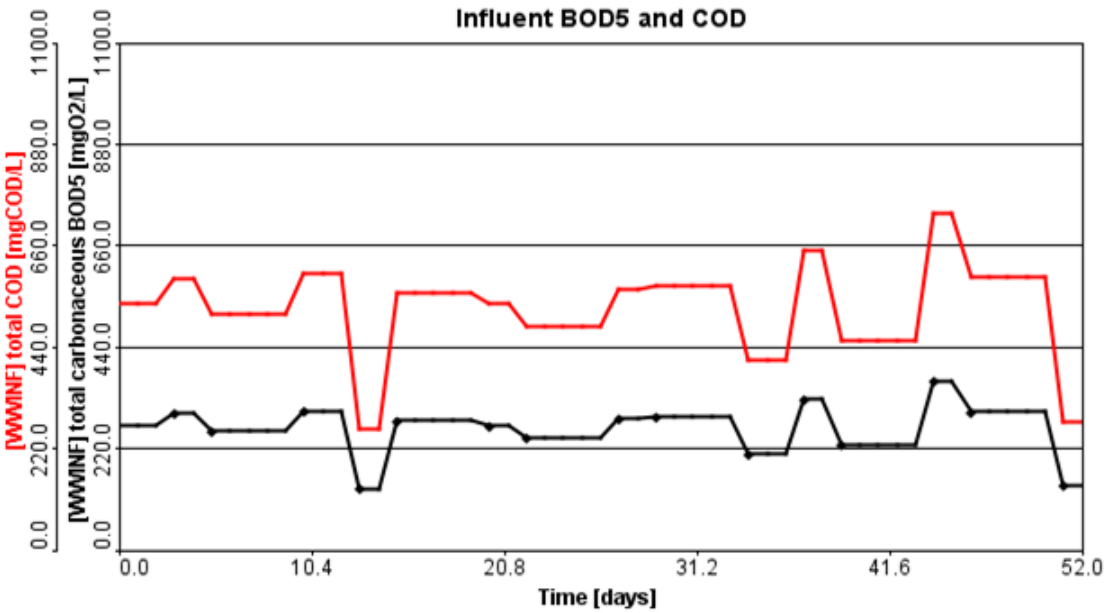
(a)



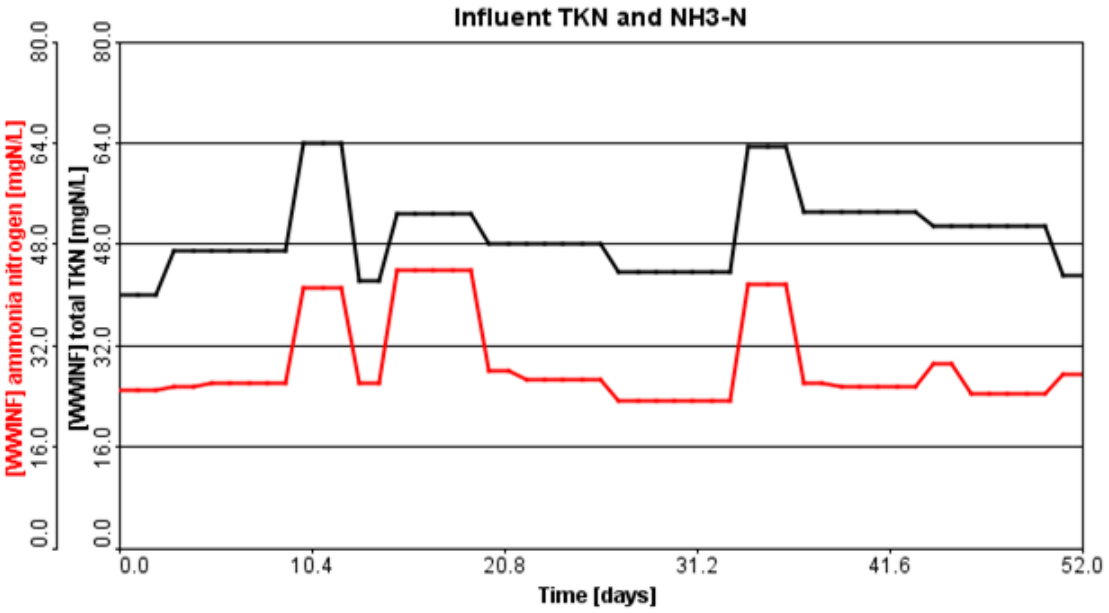
(b)


Influent Flow
[WWINF] flow [MLD]
0.0
8.0
16.0
24.0
32.0
40.0
0.0
10.4
20.8
31.2
41.6
52.0
Time [days]

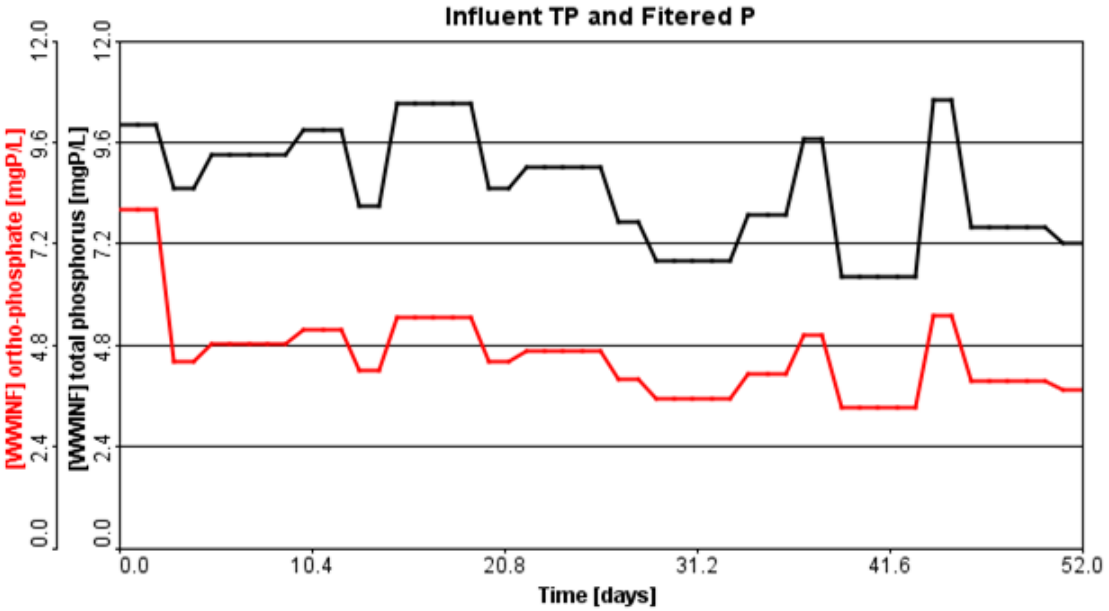
(c)


Influent BOD5 and COD
[WWINF] total COD [mgCOD/L]
[WWINF] total carbonaceous BOD5 [mgO2/L]
0.0
220.0
440.0
660.0
880.0
1100.0
0.0
10.4
20.8
31.2
41.6
52.0
Time [days]

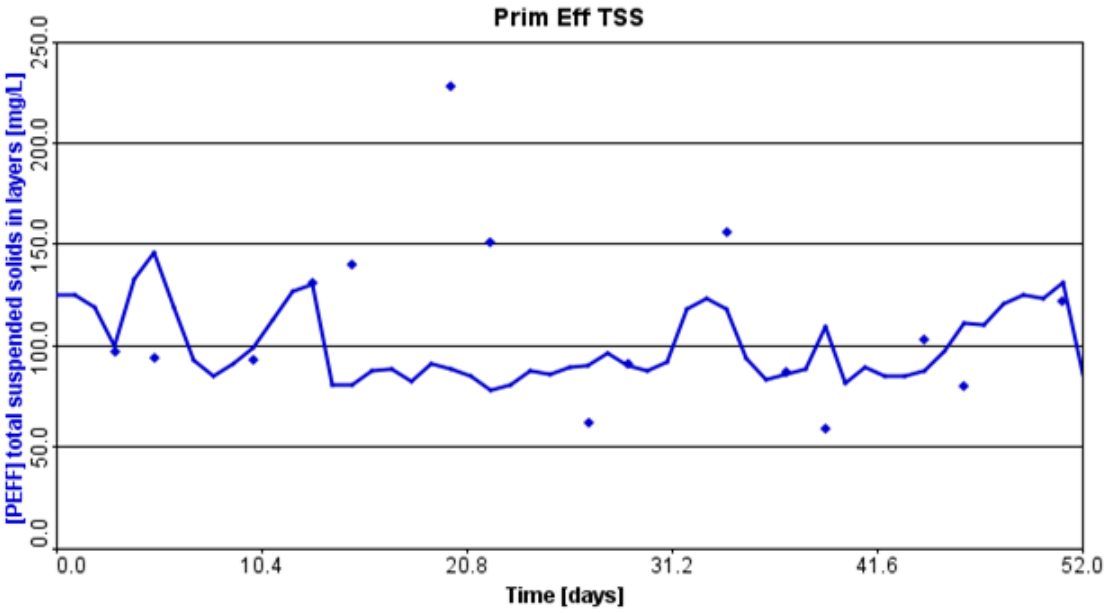
(d)


Influent TKN and NH3-N
[WWINF] ammonia nitrogen [mgN/L]
[WWINF] total TKN [mgN/L]
0.0
16.0
32.0
48.0
64.0
80.0
0.0
10.4
20.8
31.2
41.6
52.0
Time [days]

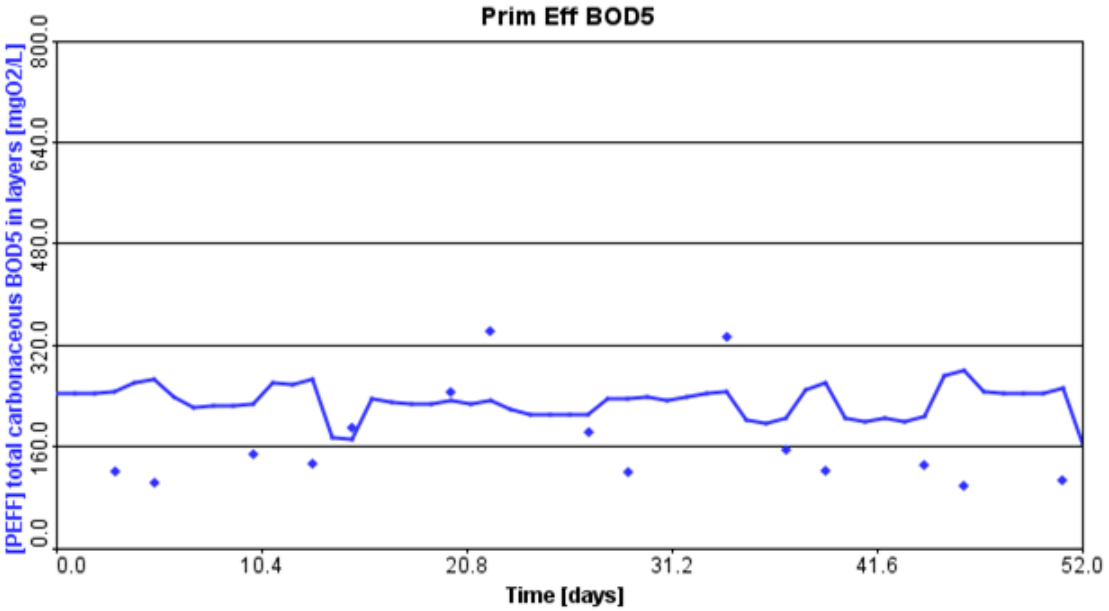
(e)


Influent TP and Fitered P
[WWINF] ortho-phosphate [mgP/L]
[WWINF] total phosphorus [mgP/L]
0.0
2.4
4.8
7.2
9.6
12.0
0.0
10.4
20.8
31.2
41.6
52.0
Time [days]

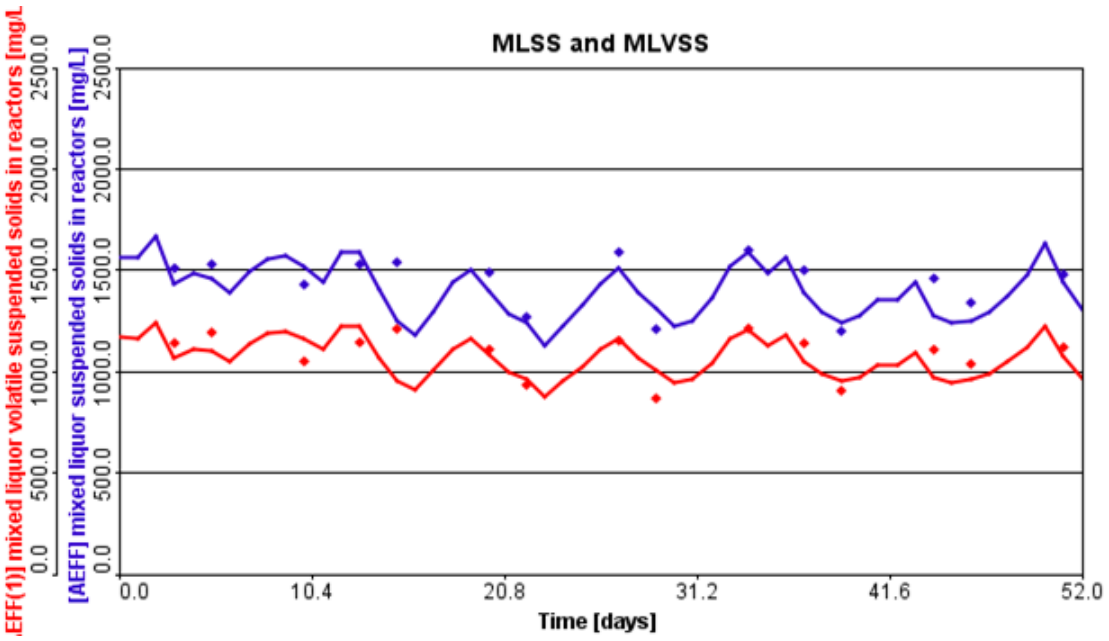
(f)


Prim Eff TSS
[PEFF] total suspended solids in layers [mg/L]
250.0
200.0
150.0
100.0
50.0
0.0
0.0
10.4
20.8
31.2
41.6
52.0
Time [days]

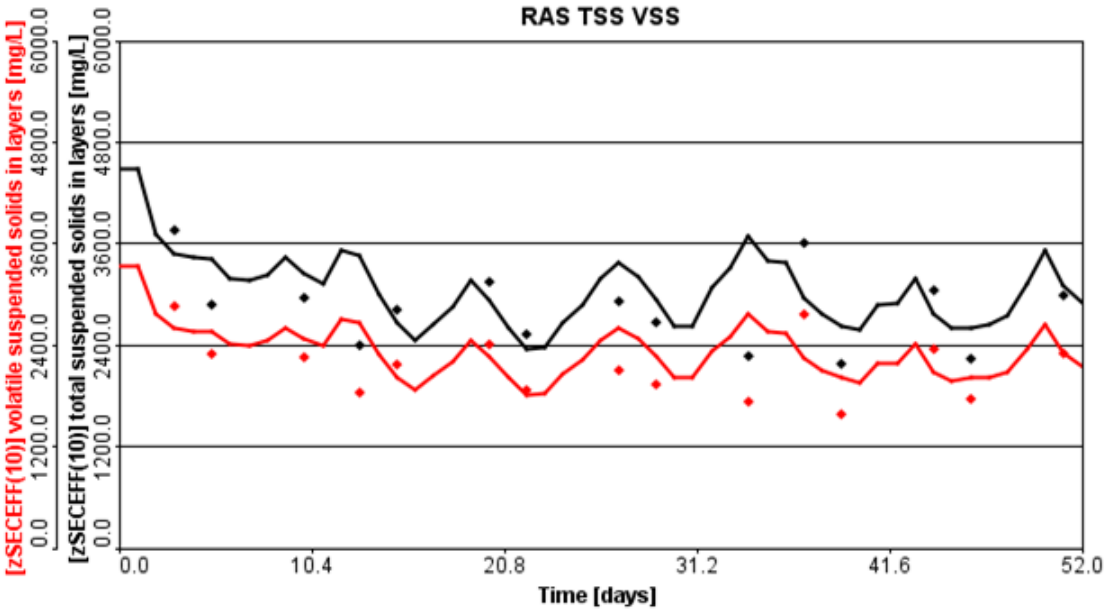
(g)


Prim Eff BOD5
[PEFF] total carbonaceous BOD5 in layers [mgO2/L]
800.0
640.0
480.0
320.0
160.0
0.0
0.0
10.4
20.8
31.2
41.6
52.0
Time [days]

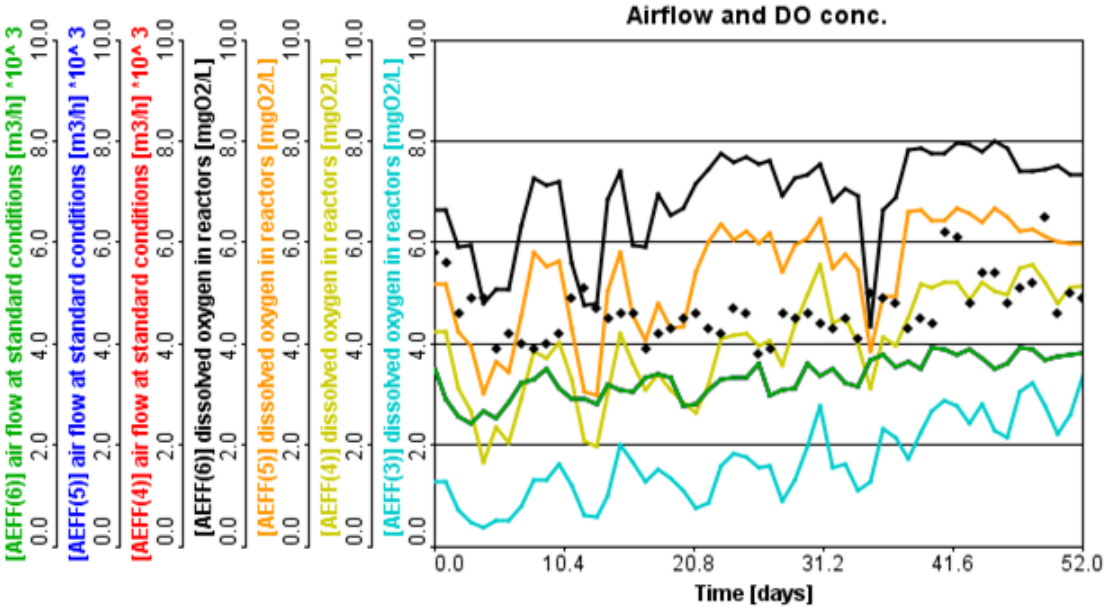
(h)


MLSS and MLVSS
[AEFF] mixed liquor suspended solids in reactors [mg/L]
0.0
500.0
1000.0
1500.0
2000.0
2500.0
0.0
10.4
20.8
31.2
41.6
52.0
Time [days]

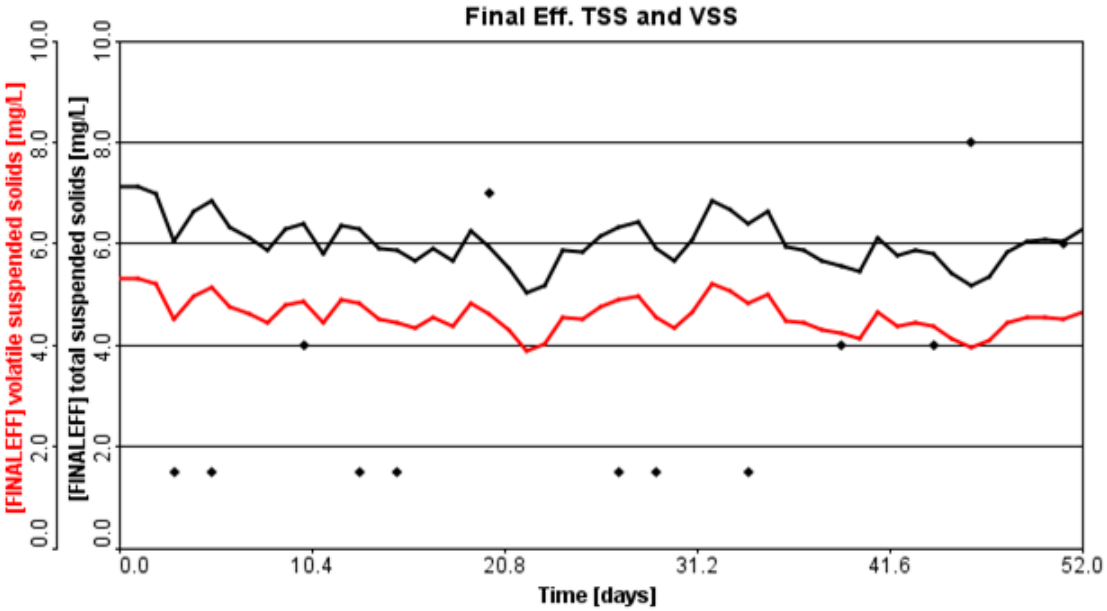
(i)


RAS TSS VSS
[zSECEFF(10)] volatile suspended solids in layers [mg/L]
[zSECEFF(10)] total suspended solids in layers [mg/L]
0.0
1200.0
2400.0
3600.0
4800.0
6000.0
0.0
10.4
20.8
31.2
41.6
52.0
Time [days]

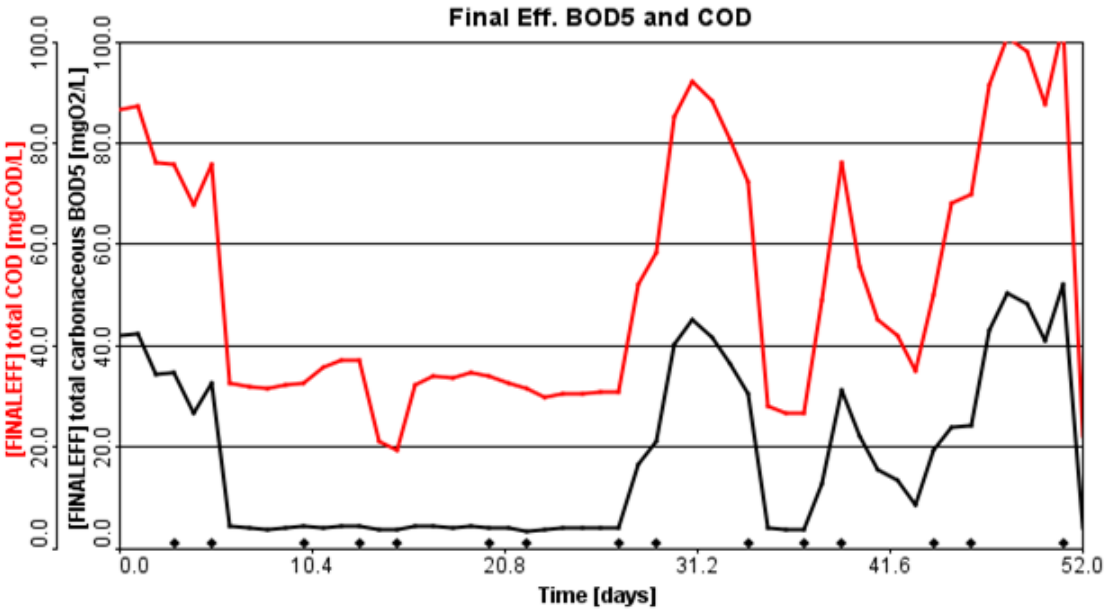
(j)


Airflow and DO conc.
[AEFF(6)] air flow at standard conditions [m3/h] *10^3
[AEFF(5)] air flow at standard conditions [m3/h] *10^3
[AEFF(4)] air flow at standard conditions [m3/h] *10^3
[AEFF(6)] dissolved oxygen in reactors [mgO2/L]
[AEFF(5)] dissolved oxygen in reactors [mgO2/L]
[AEFF(4)] dissolved oxygen in reactors [mgO2/L]
[AEFF(3)] dissolved oxygen in reactors [mgO2/L]
0.0
2.0
4.0
6.0
8.0
10.0
0.0
10.4
20.8
31.2
41.6
52.0
Time [days]

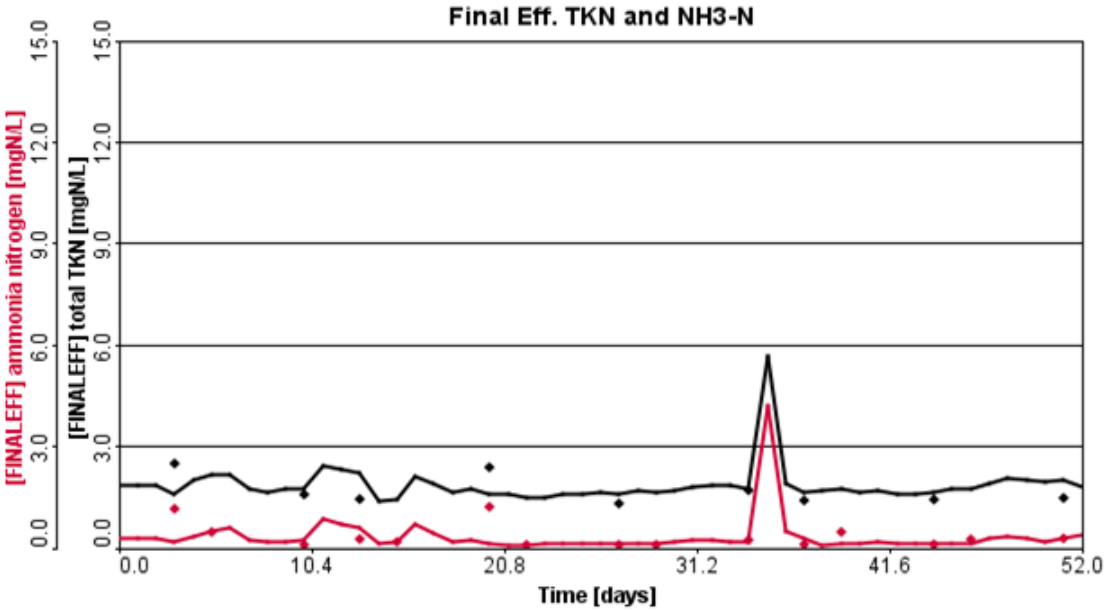
(k)


Final Eff. TSS and VSS
[FINALEFF] volatile suspended solids [mg/L]
[FINALEFF] total suspended solids [mg/L]
0.0
2.0
4.0
6.0
8.0
10.0
0.0
10.4
20.8
31.2
41.6
52.0
Time [days]

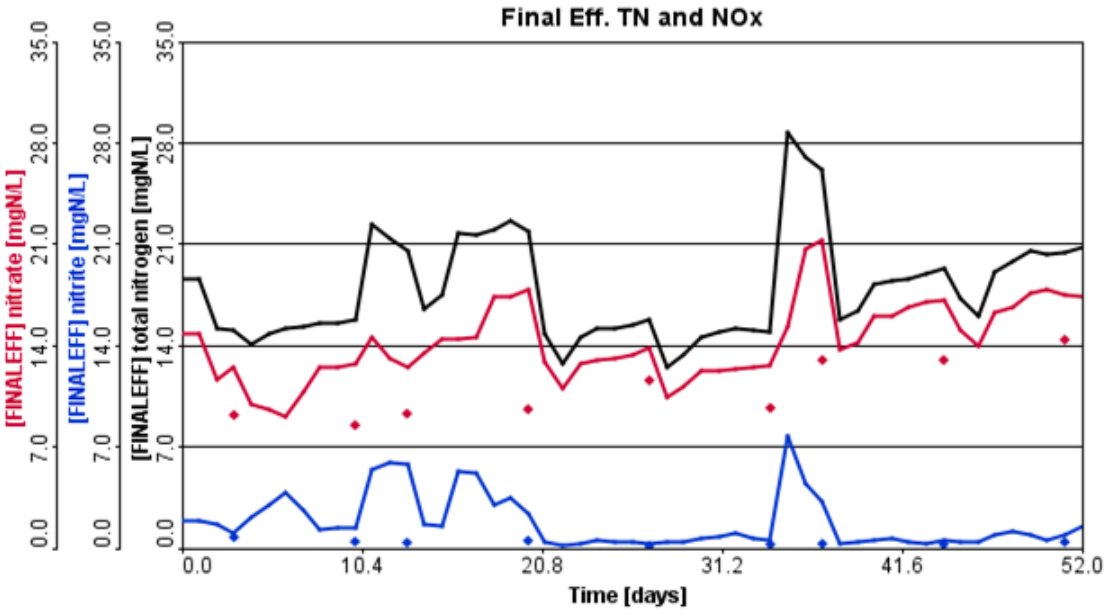
(l)


Final Eff. BOD5 and COD
[FINALEFF] total COD [mgCOD/L]
[FINALEFF] total carbonaceous BOD5 [mgO2/L]
0.0
20.0
40.0
60.0
80.0
100.0
0.0
10.4
20.8
31.2
41.6
52.0
Time [days]

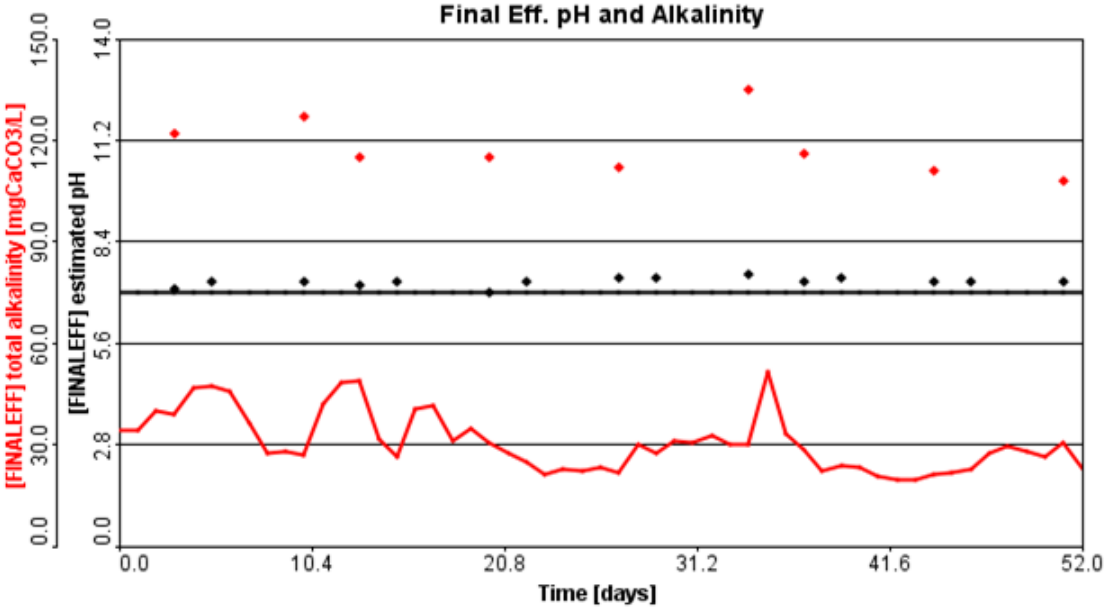
(m)


Final Eff. TKN and NH3-N
[FINALEFF] ammonia nitrogen [mgN/L]
[FINALEFF] total TKN [mgN/L]
0.0
3.0
6.0
9.0
12.0
15.0
0.0
10.4
20.8
31.2
41.6
52.0
Time [days]

(n)


Final Eff. TN and NOx
[FINALEFF] nitrate [mgN/L]
[FINALEFF] nitrite [mgN/L]
[FINALEFF] total nitrogen [mgN/L]
0.0
7.0
14.0
21.0
28.0
35.0
0.0
10.4
20.8
31.2
41.6
52.0
Time [days]

(o)


Final Eff. pH and Alkalinity
[FINALEFF] total alkalinity [mgCaCO3/L]
[FINALEFF] estimated pH
0.0
30.0
60.0
90.0
120.0
150.0
0.0
2.8
5.6
8.4
11.2
14.0
0.0
10.4
20.8
31.2
41.6
52.0
Time [days]

(p)

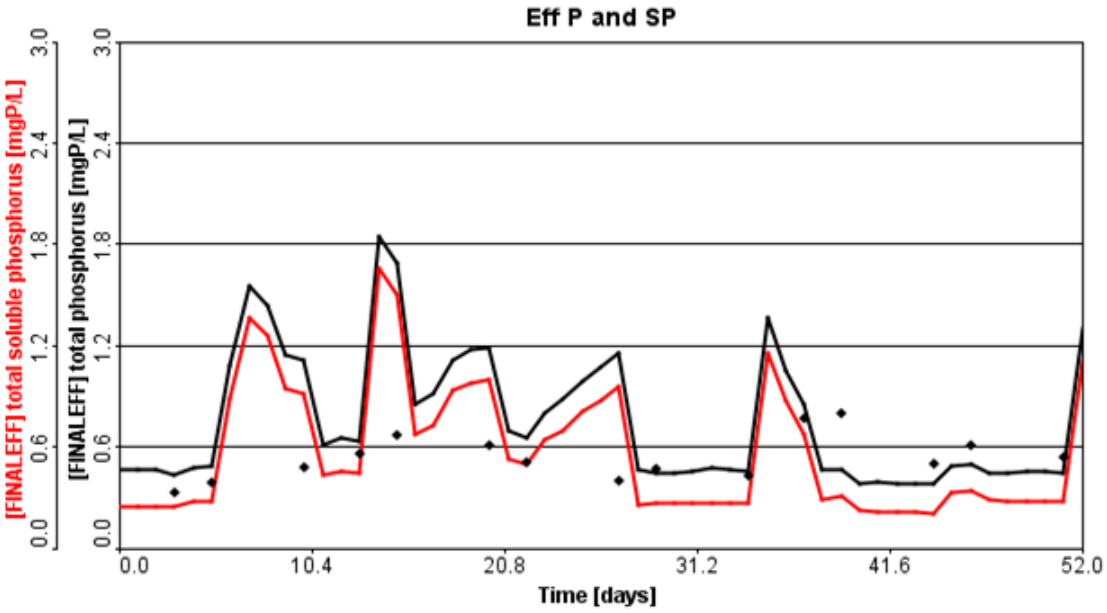
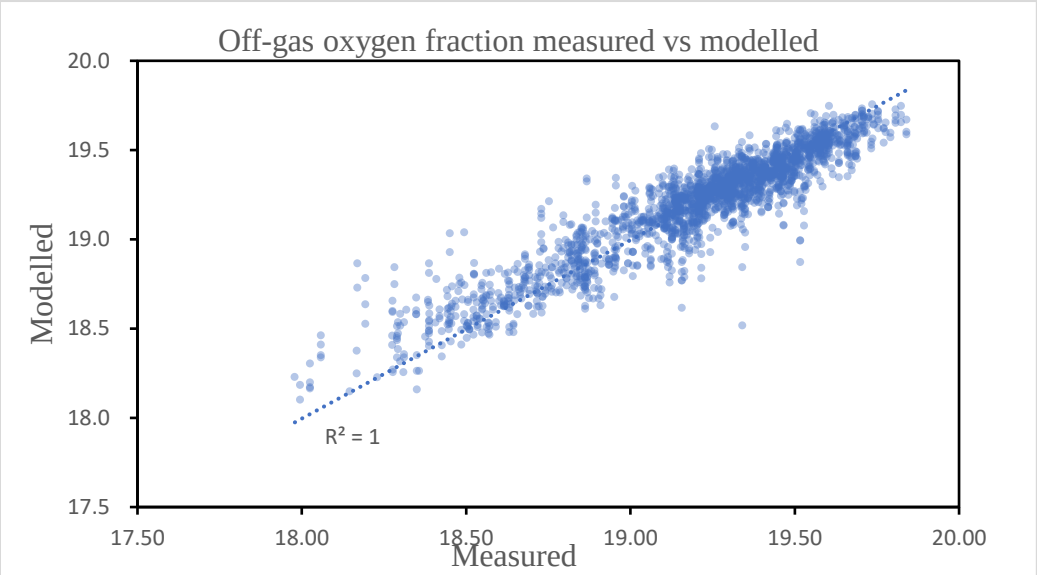


Figure SD4 GPS-X model export modeled operational parameters verses measured, for extracting sCOD data (a) Influent TSS and VSS (b) Influent flow rate (c) Influent BOD5 and COD (d) Influent TKN and NH4-N (e) Influent TP and SP (f) Post-primary TSS (g) post-primary BOD5 (h) MLSS and MLVSS (i) Return activate sludge TSS and VSS (j) dissolved oxygen at different position (k) Effluent TSS and VSS. (l) effluent BOD5 and COD (m) effluent TKN and NH4-N. (n) Effluent TN, NO2-N, and NO3-N. (o) Effluent pH and alkalinity (p) Effluent TP and SP

Appendix E ML modelled alpha and O2 fraction vs measured SD5 (a) (b)

(a)



(b)

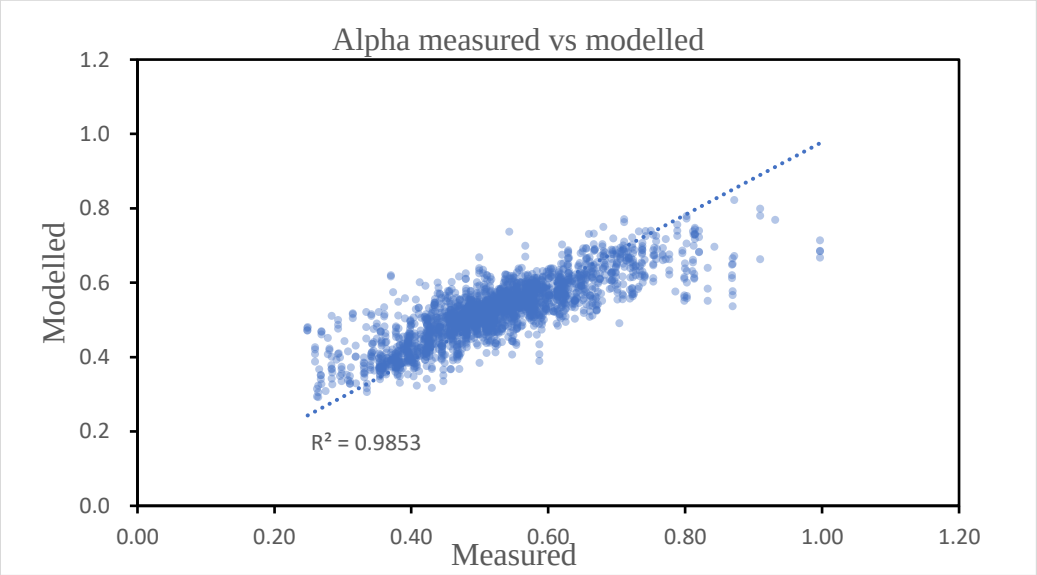


Figure SD5 ML modelled off-gas oxygen fraction and alpha vs measured coefficient of determination (a) Off-gas oxygen fraction (b) Alpha

Curriculum Vitae

Name:	Yuehe Pan
Post-secondary Education and Degrees:	Western University London, Ontario, Canada 2015-2019 B.E. Chemical and Biochemical Engineering Western University London, Ontario, Canada 2019-2021 MESc. Civil and Environmental Engineering
Honours and Awards:	The Western Scholarship of Excellence 2015
Related Work Experience	Research Assistant Guangzhou Institute of Energy Conversion, Chinese Academy of Sciences 2017 Technician Wuhan Engelbart Ecological Technology Inc. 2018

Publications:

Badia, A., Pan, Y. and Dagnew, M. 2022 Nitrite accumulation in denitrifying systems: Impact of COD:N on nitrite versus nitrate denitrification rates, microbial composition and kinetics. (Manuscript submitted for publication)

Pan, Yuehe, and Martha Dagnew. 2022. A new approach to estimate dynamic alpha factor in aeration systems using automatic machine learning models. (Manuscript submitted for publication)